2015 International Conference on Cyberworlds

Social Engineering Attack Detection Model: SEADMv2

Francois Mouton*†, Louise Leenen* and H.S. Venter†
*Defence Peace Safety & Security, Council for Industrial and Scientific Research
Pretoria, South Africa
E-mail: moutonf@gmail.com, lleenen@csir.co.za
†University of Pretoria, Department of Computer Science
Pretoria, South Africa
E-mail: hventer@cs.up.ac.za

***Abstract*—Information security is a fast-growing discipline, and therefore the effectiveness of security measures to protect sensitive information needs to be increased. Since people are generally susceptible to manipulation, humans often prove to be the weak link in the security chain. A social engineering attack targets this weakness by using various manipulation techniques to elicit individuals to perform sensitive requests. The field of social engineering is still in its infancy as far as formal definitions, attack frameworks, examples of attacks and detection models are concerned. This paper therefore proposes a revised version of the Social Engineering Attack Detection Model. The previous model was designed with a call centre environment in mind and is only able to cater for social engineering attacks that use bidirectional communication. Previous research discovered that social engineering attacks can be classified into three different categories, namely attacks that utilise bidirectional communication, unidirectional communication or indirect communication. The proposed (and revised) Social Engineering Attack Detection Model addresses this problem by extending the model to cater for social engineering attacks that use bidirectional communication, unidirectional communication or indirect communication. The revised Social Engineering Attack Detection Model is further verified using published generalised social engineering attack examples from each of the three categories mentioned.**

Keywords*—*Bidirectional Communication, Indirect Communication, Social Engineering, Social Engineering Attack Examples, Social Engineering Attack Detection Model, Unidirectional Communication.

I. Introduction

Information security is a fast-growing discipline. The protection of information is of vital importance to organisations and governments, and therefore the development of measures to counter illegal access to information is an area that receives increasing attention. Organisations and governments have a vested interest in securing sensitive information and thus in securing the trust of clients and citizens. Technology on its own is not a sufficient safeguard against information theft. Staff members — often the weak link in an information security system — can be influenced or manipulated to divulge sensitive information that allows unauthorised individuals to gain access to protected systems.

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 a number of different models of a social engineering attack exist [1], [2], [3], [4], [5], [6], [7]. The authors of this paper considered different definitions of social engineering and social engineering attack taxonomies in a previous paper, *Towards an Ontological Model Defining the Social Engineering Domain* [1], and formulated a definition for both social engineering and a social engineering attack. In addition, they also proposed an ontological model for a social engineering attack and defined social engineering as "the science of using social interaction as a means to persuade an individual or an organisation to comply with a specific request from an attacker where either the social interaction, the persuasion or the request involves a computer-related entity" [1].

Furthermore, a social engineering attack framework was developed and has been verified using real-life social engineering examples [2]. This framework, which depicts the logical flow of a social engineering attack [2], was further utilised to develop social engineering attack examples that are used in this paper to verify the Social Engineering Attack Detection Model (SEADM).

As clearly stated by various authors [8], [9], [10], [11], the human element is the 'glitch' or vulnerable element within security systems. Unfortunately it is the basic 'good' characteristics of human nature that make people vulnerable to the techniques used by social engineers, as the latter exploit various psychological vulnerabilities to manipulate the individual to disclose the requested information [8], [11].

Individuals make themselves even more vulnerable to social engineering attacks by not expecting ever to be a victim of such an attack, and many will never even know that they have actually been a victim. Most people are not aware of and do not fully comprehend the extent to which these techniques to obtain information can be (ab)used. Moreover, they are not aware of the potential it holds for personal, economic and social harm, as well as for losses for the individual and the institution. Individuals may believe that the information in their possession is not of particular value to anyone else and cannot be used for a malicious act, and therefore they will be quite willing to disclose information freely. However, the crafty social engineer is dedicated to researching and gathering information from various sources. Combined, the abuse of such acquired information can have dire consequences.

The authors earlier proposed a social engineering attack

978-1-4673-9403-1/15 $31.00 © 2015 IEEE
DOI 10.1109/CW.2015.52

detect model that was specifically aimed at a call centre environment [12]. Although the model worked well, it catered only for social engineering attacks that utilised bidirectional communication. After further research, the authors realised that social engineering attacks can be divided into three categories based on the type of communication, namely bidirectional communication, unidirectional communication and indirect communication. Hence they realised the need for a model to cater for all three categories and are hereby proposing a revised version of the social engineering attack detection model (SEADM).

The problem at hand is to successfully detect social engineering attacks while working in a stressful environment where decisions must be made instantaneously. It is for this reason that a practical model that can be easily implemented and used by all levels of employees is necessary and proposed in this paper. This model should be used in combination with training on various social engineering techniques, the psychological vulnerabilities they may elicit, and institutional policies and procedures.

Section II provides a background on the previous social engineering model and discusses the authors' previous work. Section III proposes the revised version of the SEADM. Section IV maps social engineering attack examples to demonstrate the use of the SEADM. Section V concludes the paper.

II. Previous Social Engineering Attack Detection Model

Many models and taxonomies have been proposed for social engineering attacks [1], [2], [3], [4], [5], [6], [7]. The authors' ontological model depicts that a social engineering attack "employs either direct communication or indirect communication, and has a social engineer, a target, a medium, a goal, one or more compliance principles and one or more techniques" [1]. The ontological model clearly splits social engineering into three categories, namely bidirectional communication, unidirectional communication and indirect communication.

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

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

The previous SEADM was designed to cater specifically for social engineering attacks in a call centre environment [12], [13]. This research was the first attempt to develop a detection model for social engineering attacks, and at the time of publishing this article there was still only limited research available in this field. Most of the research in this domain still centres around the training of users [14], [8], [15]. The steps in the revised SEADM have been generalised to cater for all three communication categories, whereas the previous model was only able to deal with bidirectional communication. The following section proposes the revised version of the SEADM and then shows how the revised SEADM is able to thwart social engineering attacks from all three categories of social engineering.

III. Revised Social Engineering Attack Detection Model

As was indicated above, a model is needed as a guideline to detect social engineering attacks. The authors propose a revised SEADM as depicted in Figure 1. This model makes use of a decision tree and breaks down the process into more manageable components to aid decision making.

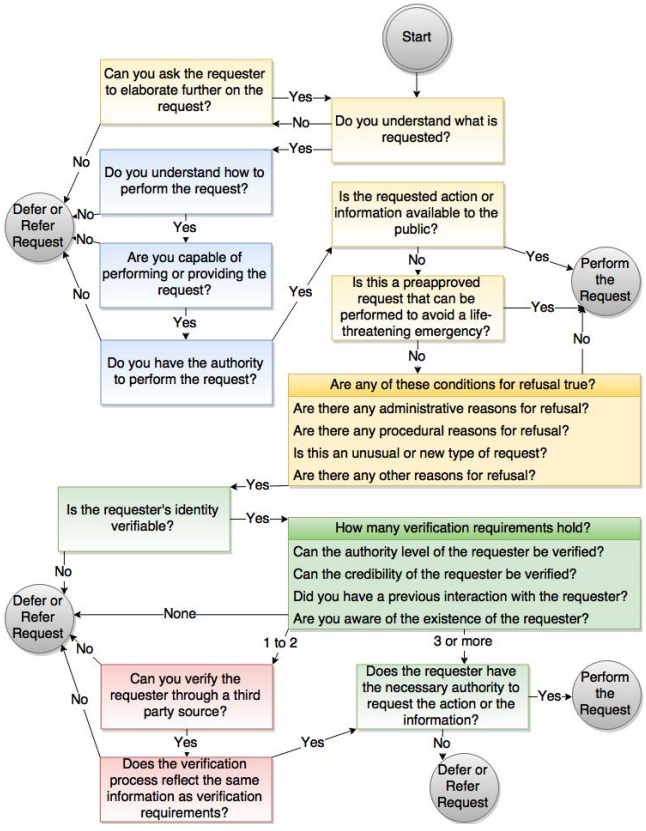

Fig. 1. Social Engineering Attack Detection Model

The model also depicts the flow of action and how any type of request should be handled by a 'receiver'. Throughout this discussion this term is understood as the person dealing with the request, while the term 'requester' is defined as the person or object who requests the specific action or information from the receiver. The model should be used as a guideline to aid

in decision making and it is an improvement on the initial SEADM due to its ability to cater for both typical requests and inherent requests. This generalisation allows the revised SEADM to cater for the both the unidirectional communication and indirect communication categories of social engineering.

An example of a typical request is where the requester, in this case a person, requests the receiver to perform a task/favour for him/her. This request can range from the requester requesting information about an organisation to the requester requesting that the receiver performs a password reset for an individual's Internet banking logon.

An example of an inherent request is where the receiver receives a request, in this case an object that contains either a request or a process which needs to be completed by the receiver. This type of request can range from a parking ticket detailing how to pay the ticket on the pamphlet to a receiver finding a storage medium device and wanting to return the device to its rightful owner. In the case of the parking ticket, the receiver is inherently requested to pay the ticket using the information on the pamphlet. In the case of the storage device the situation is a little bit more complicated. The receiver, upon finding the device, is inherently requested to return it to its rightful owner.

The model provides for four different types of states — the request, receiver, requester and third party — which provide a brief idea of what can be expected to be performed in each state. The request states, indicated in yellow, directly deals with information about the request itself. The receiver states, indicated in blue, directly deals with the person handling the request and whether this person (the receiver) understands and is allowed to perform the request. The requester states, indicated in green, directly deals with the requester and whether any information about the requester can be verified. The third party states, indicated in red, directly depict the involvement of a third party in the model and whether the information about the requester can be externally verified.

This paper addresses each of these states individually as shown in Figure 1, before the full model is discussed with examples.

A. Do you understand what is requested?

The first step of the model tests whether the receiver is able to understand the request. One must fully comprehend the entire request before one can accurately determine whether it involves a social engineering attack or not. In the case where the receiver does not comprehend the request in full or where additional information is required to comprehend the request, the 'no' option is selected so that more information about the request can be asked. The 'no' option can be selected repeatedly if the additional information provided is not yet enough for the receiver to comprehend the request fully. When the receiver fully comprehends what the request is, the 'yes' option is selected.

B. Can you ask the requester to elaborate further on the request?

This step is provided to allow the receiver to request additional information about the request from the requester. The receiver now attempts to ask for more information about the request to better understand the request. The step can also be repeated if additional information is required; however, when no additional information about the request can be obtained, the 'no' option is selected. The 'no' option leads the receiver to either defer or refer the request. Every time more information is obtained, the 'yes' option is selected and a test is once again performed to determine whether the receiver understands what is requested.

C. Do you understand how to perform the request?

It is very important to determine whether the receiver is knowledgeable enough to perform the request in full. This step differs from the first step, because the receiver has to measure whether he or she has the required knowledgeable to perform the request and not whether the request itself is understood. The receiver performs this measurement based on the own capability to understand the procedure required to perform the request. If the receiver has the necessary knowledge to perform the request, the 'yes' option is chosen and the receiver proceeds to the following step. In the case where the receiver does not understand how to perform the request, the 'no' option is chosen.

D. Are you capable of performing or providing the request?

Once the receiver has verified that he or she understands how to perform the request, it needs to be determined whether the receiver is capable of performing the request. This step measures whether the receiver has the means or capability to perform the request, in other words whether he or she is able to determine which option to choose. If the 'no' option is taken, the request is deferred or referred to another individual. Otherwise, the 'yes' option is selected and the next step involves the receiver's authority to perform the request.

E. Do you have the authority to perform the request?

The receiver needs to measure whether he has the required authority to perform the request. Even though at this step the receiver may already have determined that he or she understands how to perform the request and is capable of performing it, it must still be determined whether the receiver has the authority to perform the request. This step therefore measures whether the receiver, as part of his or her duty, has the authority to provide the requester with the requested action or information. If the receiver has the necessary authority, the 'yes' option is taken, otherwise the 'no' option is selected. This is the last step for the receiver to measure him- or herself — the next two steps focus on the request itself.

F. Is the requested action or information available to the public?

This is the first step that may lead to a request being performed, as it is here determined whether the requested action or information is already publicly available. In the case where the request involves an action, it must be determined whether this requested action is available to the public or whether the requester should possess any level of authority to request the specific action. In the case where the requested action is available to everyone, the 'yes' option is taken and

the request is performed. If any type of authority is required for the specific action, the 'no' option is chosen and the receiver proceeds to the next step. The procedure is similar for requested information, since it needs to be measured whether the information is public or private. The receiver should have a clear understanding of what information is readily accessible to the public. For example, information in the public domain for an institution could include contact details and working hours, which could be available on the institution's website and may thus be provided to a requester.

G. Is this a preapproved request that can be performed to avoid a life-threatening emergency?

The receiver needs to assess the urgency of the request and the consequences for the requester if the request is not performed. If the request does not constitute a life-threatening emergency, the 'no' option must be taken. An example of a life-threatening emergency is where an individual's medical insurance number is required because he or she was injured in an accident. This is a very difficult step to measure, because a skilled social engineer can use a life-threatening emergency to get the receiver to perform an unauthorised request. It should be noted that a request must be performed when an individual's life could be at risk upon denial of the request and could potentially cause harm to the individual. For this specific step, the receiver must be able to clearly determine what constitutes a life-threatening emergency and which requests may be performed during a life-threatening emergency. If the list of requests that may be performed is continually revised, the social engineer will be limited and thus it can be deemed safe to provide the limited set of requests during a life-threatening emergency. In the case where it is known that an organisation does not deal with life-threatening emergencies, this step can be omitted from the model.

H. Are any of these conditions for refusal true?

In this step, it is determined whether there are any conditions that can constitute a refusal to perform the request. Only four conditions are provided in the model, therefore they do not constitute an exhaustive list. These conditions can be developed to cater for specific organisations, depending on where the model is implemented. The four conditions that are provided are of a very generalised nature to cater for the most probable conditions without mentioning the specific requirement. If any condition here constitutes a sufficient reason for refusal, the 'yes' option is taken; otherwise, the request is performed. The following subsections elaborate on the conditions used in the refusal component of the model.

1) Are there any administrative reasons for refusal?: This condition refers to all the possible administrative reasons that may constitute a sufficient reason for refusal. For example, the requester may be required to provide information if there is an administrative process in place that requires a specific level of authority for requesting the receiver to perform the request. The term 'administrative reason' is used to refer to a wide variety of reasons and keeps the model more generic.

2) Are there any procedural reasons for refusal?: This condition refers to all the possible procedural reasons that may constitute a sufficient reason for refusal. For example, the receiver may have stumbled on a portable storage medium and wants to return it to its rightful owner. However, there is a procedural policy in place that forbids the receiver to plug in any storage medium into a workstation at the organisation and this will then constitute a valid reason to refuse the request and take the 'yes' option. The term 'procedural reason' is used to refer to a wide variety of reasons and it keeps the model more generic.

3) Is this an unusual or new type of request?: One also needs to consider the case where the type of request has not yet been defined in an organisation or where the receiver has never before dealt with such a type of request. In this case, the unusual or novel nature of the request will be sufficient reason to refuse it and take the 'yes' option to further ascertain the requester's identity. For example, the receiver may receive a request to perform a password reset for a colleague; however, the request is received via e-mail and not telephonically as it is usually done. In this case, it is sufficient reason for the receiver to refuse the request.

4) Are there any other reasons for refusal?: This condition is open-ended as the receiver may feel uneasy with the request or there may be a reason why the receiver intuitively does not want to perform the request. Although it is added as an additional safeguard to the model, this step is not seen as redundant, as it provides the receiver with an additional means to check out information about the requester. Because the receiver will have the ability to further verify the requester in the steps that follow, it may put the receiver at ease to perform the request.

I. Is the requester's identity verifiable?

The receiver now needs to verify the identity of the requester to be able to make an informed and rational decision about whether the request should be performed. If at this stage of the model the requester's identity cannot be verified, the 'no' option is taken and the request is either deferred or referred. In the case where the receiver can perform steps to identify the requester and the associated roles, authority or additional details, the 'yes' option is chosen to determine the number of levels on which the requester can be verified.

J. How many verification requirements hold?

Depending on the extent to which the requester's identity can be verified, a different set of states will be examined to determine whether the request should be performed. Important to remember is that the social engineer may be portraying him- or herself as an authority figure in the institution; a computer technician, or any other persona that may elicit compliance. As people we are inclined to make quick assumptions regarding others and their status, sometimes even based on trivialities such as clothing. If someone is dressed in the proper attire, uses the appropriate institutional jargon or uses an important individual's name, it does not necessarily indicate that such individual is trustworthy. The same holds for physical objects, as individuals are inclined to help other people. If an individual finds a storage medium lying around and the storage medium is marked to be important, the individual will likely feel the urge to either return the storage medium to its rightful owner or curiosity may drive the individual to examine the contents

on the storage medium. If the receiver feels unsure at any time, a super user should be contacted to obtain the authority to provide or decline the request.

The following verification requirements should be taken into account and used as a basis for a decision on whether to perform or not perform the request: authority, credibility, previous interaction, and knowledge of the person's existence. The model has only four verification requirements and hence does not constitute an exhaustive list — additional requirements can be added to cater for specialised environments. The current list of requirements is very broad and only includes the most common verification requirements.

For example, some of the techniques that can aid in the verification of an individual's identity, specifically in a call centre environment, are the following: Caller Identification; Calling back the requester on a predetermined phone number; Requesting a secure email address; Requesting a secure password; Requesting face-to-face interaction with the individual where proper identification can be provided; Having another employee to vouch for the requester; Contacting the requester's immediate supervisor to verify the former's identity; Using an employee directory [9]. This illustrates how the verification requirements can be changed based on the environment and organisation where the model is implemented.

It is not always possible to verify all of the requirements, so the model has three different outcomes for this specific step. If none of the verification requirements hold, the 'none' option is selected and the request is either deferred or referred. In the case where only one to two requirements hold, the path to further verify the information from a third party source is selected. Lastly, if three or more requirements hold, the receiver proceeds to the final step of the model where it is measured whether the requester has the necessary authority to request either the action or the information. The strictness of this step can also be changed depending on the environment or organisation where this model is implemented. The strictness is highly dependent on the verification requirements that are utilised. Each of the verification qualities is addressed in the next subsection.

1) Can the authority level of the requester be verified?: Authority is an integral part of any institution, with an almost conditioned response from employees to adhere to an authoritative requester's wishes and demands, combined with a fear of punishment should the receiver appear to undermine the requester [16]. For these reasons, impersonating an authoritative individual is a very effective technique used by social engineers to obtain privileged information or actions. The institution needs to provide an environment in which the employee feels comfortable and is indeed expected to question the authority figure's identity before disclosing sensitive information. The employee also needs to know — with the help of a clear institutional policy — on what authorisation level a particular person of authority is rated, in regard to what privileged information or action can be provided.

The same situation applies in our daily lives, where individuals adhere to rules and regulations put in place by an authoritative figure. Examples of such official procedures are road rules and regulations, and individuals tend to follow these rules and regulations because they create a safe environment for everyone to travel in. Theft is also frowned upon. When a storage medium is found by an individual, the individual will rather want to return this storage medium to its rightful owner, otherwise, it might be considered as stealing the device.

2) Can the credibility of the requester be verified?: The employee needs to judge the level of credibility of the requester. However, this is a challenging task, as establishing credibility is the first aim that the social engineer tries to achieve, and what the attack will be based on. If the requester knows the jargon used by a particular institution, people easily assume that he or she is an employee at their particular institution. The requester could, for example, be an ex-employee who is (still) quite knowledgeable about the jargon and procedures. Such an aggrieved ex-employee may try to seek revenge with the goal of obtaining particular sensitive information. The credibility of the requester is measured on the basis of how well or bad he or she responds to a predefined of set of questions used to determine the credibility of a requester.

Similarly, for objects and items, the credibility of an item has to be measured in terms of whether it conforms to the institution's guidelines. One must examine whether it is the same brand used by the institution or whether the institution's lanyard is attached to it. These are all minor techniques that can be used to ensure that the object or item indeed belongs to or is part of the institution.

3) Did you have a previous interaction with the requester?: If the individual has had previous interaction with the requester, especially a long-standing history of interaction, the decision whether information can be provided will be an easier task. However, limited interactions with the requester, especially by telephone and email alone, should be taken into account in conjunction with other verification techniques, to be able to make an informed and safe decision regarding the request. This test can only be performed on individuals and not on items or objects that have an inherent request.

4) Are you aware of the existence of the requester?: This refers to the knowledge that the requester exists in the institution or the fact that an outside collaborating partner on a project can support the verification of the requester. However, this should also be used in conjunction with the other verification techniques, as the requester could be a social engineer portraying him- or herself to be some well-known individual in order to get the receiver to perform the request. This awareness test can only be performed on individuals and not when the requester is an object that has an inherent request.

K. Can you verify the requester through a third party source?

When the requester did not adhere to enough verification requirements and only some of the verification requirements held, this step is reached. It tests whether it is possible to verify the information obtained from the requester through an external source. An example of such a case is where the requester provides information and claims to be part of a specific organisation, as well as to have been requested by his or her organisation to ask the receiver to perform a request. If it is possible for the receiver to contact another individual at this organisation to verify this information, the 'yes' option is chosen, otherwise the 'no' option is taken and the requested is either deferred or referred.

L. Does the verification process reflect the same information as the verification requirements?

In this step, the receiver will verify all the verification requirements as obtained from the third party source. Continuing with the previous example, the receiver will attempt to contact the organisation and verify the information received from the requester. If the information matches what has been received from the requester, the 'yes' option is taken, otherwise, the request is either deferred or referred.

M. Does the requester have the necessary authority to request the action or the information?

With the aid of the previous steps the receiver has acquired the necessary knowledge regarding the requester's identity and authority level, together with whether the receiver may perform the request. The receiver can now determine whether the requester has a level of authority on the same level or higher than the level of sensitivity of the request. If the requester has the authorisation on the same authority level as or higher than needed for the particular request, the request is performed. However, if the requester does not have the necessary authorisation, the 'no' option is chosen and the request is deferred or referred.

N. Defer or refer request

This is the negative result state of the model. In this state the request can either be deferred or handled at a later stage. Deferring the request can also lead to the request never being performed and can be considered as the request having been halted. In the case where the receiver is part of an organisation, there is the option to refer the request to a more authoritative person in the same organisation. This will allow someone else who may have better judgement on whether to perform or halt the request to actually deal with the request.

O. Perform the request

This is the positive result state of the model. In this state the receiver is allowed to perform the single request from the requester.

IV. Mapping Social Engineering Examples to SEADM

Three example scenarios that are provided in this section are used to test the model. In previous work, social engineering was divided into three distinct categories based on the type of communication utilised (see section II). The three categories are respectively bidirectional communication, unidirectional communication and indirect communication. An example from each of these categories is used to illustrate how the model works.

In the first scenario, from the bidirectional communication category, the social engineer attempts to obtain sensitive information that is accessible only to the employees of the organisation. In the second scenario, from the unidirectional communication category, the social engineer attempts to obtain financial gain by sending out paper mail in which the letter requests a group of individuals to make a small deposit into a bank account owned by the attacker. In the third scenario, from the indirect communication category, the social engineer attempts to gain unauthorised access to a workstation in an organisation by using a storage medium device.

A. Scenario One

In this scenario a social engineer attempts to obtain the sensitive information of an organisation to which only the employees of the organisation have access. The information is not available to members of the public. This attack is performed using bidirectional communication. The social engineer sends an e-mail to an employee at the targeted organisation. This e-mail would therefore come from a different e-mail address than the company's own e-mail address. The social engineer uses the friendship and liking compliance principle to persuade the receiver to perform the requested action of providing the information to the social engineer. The attacker may use a pretext such as that he is coming to the organisation for an interview and requires information about the company policies to ensure that he is well prepared for the interview. The rest of this section maps the example to the model.

Do you understand what is requested?: The e-mail from the social engineer should clearly state what information is required and make the request easily understandable to the receiver. When the receiver understands the request, the 'yes' option is selected.

Do you understand how to perform the request?: The social engineer would have made certain that the targeted employee fully understands the request, is capable of performing the request and has the authority to perform the request. This will allow the current step, and the following two steps to take the 'yes' option.

Are you capable of performing or providing the request?: As indicated earlier, the 'yes' option is chosen.

Do you have the authority to perform the request?: As indicated earlier, the 'yes' option is taken.

Is the requested action or information available to the public?: In the scenario it states that the information is not available to the public and thus the 'no' option is chosen.

Is this a preapproved request that can be performed to avoid a life-threatening emergency?: This is not a life-threatening request and thus the 'no' option is selected.

Are any of these conditions for refusal true?: Seeing that the requested information is privileged and accessible to employees only, there is an administrative reason for refusal and thus the 'yes' option is selected.

Is the requester's identity verifiable?: In this case, bidirectional communication is utilised; thus it allows for the receiver to communicate back via e-mail and ask more questions to verify the requester. Hence the 'yes' option is taken.

How many verification requirements hold?: In this case, the friendship and liking compliance principle is utilised and the social engineer builds up a trust relationship, via e-mail, with the receiver. The pretext utilised during this attack is that sensitive information about the company policies is requested because the attacker, as part of the pretext, will be attending an

interview at the targeted employee's organisation. The receiver is able to verify all four of the verification requirements, i.e. the attacker's authority, credibility, previous interaction and knowledge of existence, as there have been several e-mails before the request. This allows the receiver to choose the 'three or more' option and proceed to the following step.

Does the requester have the necessary authority to request the action or the information?: At this step this attack will fail, even though the social engineer has built up a trust relationship with the receiver, because the attacker is not part of the organisation and thus does not have the necessary authority. The authority level is verified in the previous step, and it has been verified that the authority level of the requester is that of an individual who does not work at the organisation. This will force the receiver to defer or refer the request. The 'no' must be selected here and hence the social engineering attack is thwarted.

B. Scenario Two

In this scenario, a social engineer attempts to obtain financial gain by sending out paper mail. In the letter, a group of individuals are requested to make a small deposit into a bank account owned by the attacker. In this example, the attacker will develop a phishing letter that masks the attacker as a charity organisation requesting donations. The phishing letter contains the contact details, the logo and the purpose of the charity to improve the authenticity of the letter. This attack uses unidirectional communication and thus the receiver is not able to communicate with the attacker. The rest of this section maps the example to the model.

Do you understand what is requested?: The letter from the social engineer should clearly state that a receiver is requested to make a donation to the specific charity. The letter will include all the required details because this receiver cannot communicate with the social engineer. The 'yes' option is taken.

Do you understand how to perform the request?: The social engineer would have ensured that the targeted individual fully understands the request, is capable of performing the request and has the authority to perform the request. This will cause the receiver to select the 'yes' option in this step, as well as in the following two steps.

Are you capable of performing or providing the request?: As indicated before, the 'yes' option is taken.

Do you have the authority to perform the request?: As was the case earlier, the 'yes' option is chosen.

Is the requested action or information available to the public?: The requested action is to make a deposit into the bank account of the requester. This request is directed at the receiver and not at the public. The action of the specific receiver making a deposit is only available to the specific receiver, thus the 'no' option is taken.

Is this a preapproved request that can be performed to avoid a life-threatening emergency?: This is not a life-threatening request and thus the 'no' option is selected.

Are any of these conditions for refusal true?: This request can be seen as either unusual or new as the requester would not usually receive this specific type of letter from the charity. It can also be the case that the requester feels uneasy about the request and his or her uneasiness about the request can be seen as a reason to refuse at this point. The 'yes' option is selected because there is sufficient reason to refuse the request without even verifying the identity of the requester.

Is the requester's identity verifiable?: Since unidirectional communication is utilised in this case, the receiver can only verify the identity using the information as provided in the letter. At this point one can defer or refer the request if it does not contain additional information such as the requester's contact details. In the current scenario, the letter actually contains the contact details of the charity organisation and thus the 'yes' option is chosen.

How many verification requirements hold?: The requirement that the receiver should be aware of the existence of the requester will definitely hold, because the social engineer would have chosen a well-known charity. One can also argue that receiver may have had a previous interaction with the charity; however, from the letter alone, the authority and credibility of the requester cannot be verified. In this case the 'one to two' option is selected.

Can you verify the requester through a third party source?: The receiver will now have the ability to verify the information in the letter directly from the charity organisation. The receiver will make a phone call to the charity to verify the information. It is assumed that the charity organisation can be reached to verify the information and thus the 'yes' option is taken.

Does the verification process reflect the same information as the verification requirements?: It is at this step that the receiver will be able to ask the organisation whether such a letter has in fact been sent out. The charity organisation will deny this and thus the verification process will show that the information provided is not the same as the verification requirements. Consequently, the 'no' option will be taken and the social engineering attack will be thwarted.

C. Scenario Three

In this scenario the social engineer attempts to gain unauthorised access to a workstation in an organisation by using a storage medium device. The organisation does not have a company policy in place that disallows employees plugging storage devices into their workstations. The social engineer will leave the device outside the organisation's building to be found by an employee. The device will be infected with a trojan so that when it is plugged into the workstation, it opens a backdoor for the social engineer to connect to the system remotely. As the storage device is left unattended, this attack utilises indirect communication. The rest of this section maps this example to the model.

Do you understand what is requested?: The storage medium device planted by the social engineer should be marked clearly to indicate that it contains important and confidential information. Thus the receiver who finds this device will want to return it to its rightful owner. As it is an inherent request that the receiver should return the device, the request is easily understandable and the 'yes' option is selected.

Do you understand how to perform the request?: The social engineer would have made certain that the storage medium device is deployed at such a location that only individuals who have access to a workstation and who understand how such devices work should find the device. This will cause the receiver to take the 'yes' option in this step as well as in the following step..

Are you capable of performing or providing the request?: As was the case previously, the 'yes' option is selected.

Do you have the authority to perform the request?: In this step, the receiver should ask him- or herself whether he or she has the authority to plug the storage device into a workstation at the organisation. If there is a company policy that disallows or forbids this, then the 'no' option will be selected and the attack be thwarted. However, for the present scenario there are no company policies in place and thus the receiver has the necessary authority. Consequently, the 'yes' option is taken.

Is the requested action or information available to the public?: The inherent requested action is to return the storage device to its rightful owner. This request is directed at the receiver who found the device. Because only the receiver can perform this action, the 'no' option is taken.

Is this a preapproved request that can be performed to avoid a life-threatening emergency?: The scenario does not involve a life-threatening request and thus the 'no' option is chosen.

Are any of these conditions for refusal true?: In this scenario, the storage device has been marked as confidential and important. Hence, the receiver will not be allowed to plug the device into a workstation. This will be considered as a reason for refusal and cause the 'yes' option to be taken. Administrative and procedural reasons are ruled out for this example because there are no company policies that govern storage devices.

Is the requester's identity verifiable?: Since indirect communication is utilised in this case, the only piece of information the receiver has is the physical storage medium device. Due to the confidentiality of the device, the receiver is unable to verify the requester's identity, therefore the request is deferred or referred and the attack is thwarted. In the present scenario, the request will most likely be referred to another individual in the organisation who is allowed to safely, and on a secure workstation, verify the contents of the storage device and potentially contact the rightful owner.

V. Conclusion

The protection of information is extremely important in modern society and even though the security around information is continuously improving, a weak point is still the human being who is susceptible to manipulation techniques. This paper explored social engineering as a domain and social engineering attack detection techniques as a process inside this domain. A previous paper by the authors, *Social Engineering Attack Detection Model: SEADM* [1] was revisited.

The authors found that the previous SEADM catered only for social engineering attacks utilising bidirectional communication. This paper proposed a revised version of SEADM, which caters for all three categories of social engineering attacks. The revised SEADM has been verified using generalised social engineering attack examples. It was shown that the revised SEADM is able to thwart social engineering attacks that make use of either bidirectional communication, unidirectional communication or indirect communication.

The proposed revised SEADM can now be used as a tool to protect oneself against social engineering attacks. Even if the model is not adhered to in respect of every request, it will cause one to think differently about requests — and this is already a huge step in the right direction.

References

[1] F. Mouton, L. Leenen, M. M. Malan, and H. Venter, "Towards an ontological model defining the social engineering domain," in *ICT and Society*, ser. IFIP Advances in Information and Communication Technology, K. Kimppa, D. Whitehouse, T. Kuusela, and J. Phahlamohlaka, Eds. Springer Berlin Heidelberg, 2014, vol. 431, pp. 266–279.

[2] F. Mouton, M. M. Malan, L. Leenen, and H. Venter, "Social engineering attack framework," in *Information Security for South Africa*, Johannesburg, South Africa, Aug 2014, pp. 1–9.

[3] D. Harley, "Re-floating the titanic: Dealing with social engineering attacks," in *European Institute for Computer Antivirus Research*, 1998.

[4] L. Laribee, "Development of methodical social engineering taxonomy project," MSc, Naval Postgraduate School, Monterey, California, June 2006.

[5] K. Ivaturi and L. Janczewski, "A taxonomy for social engineering attacks," in *International Conference on Information Resources Management*, G. Grant, Ed. Centre for Information Technology, Organizations, and People, June 2011.

[6] F. Mohd Foozy, R. Ahmad, M. Abdollah, R. Yusof, and M. Mas'ud, "Generic taxonomy of social engineering attack," in *Malaysian Technical Universities International Conference on Engineering & Technology*, Batu Pahat, Johor, November 2011.

[7] P. Tetri and J. Vuorinen, "Dissecting social engineering," *Behaviour & Information Technology*, vol. 32, no. 10, pp. 1014–1023, 2013.

[8] J. W. Scheeres, "Establishing the human firewall: reducing an individual's vulnerability to social engineering attacks," Master's thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, March 2008.

[9] K. D. Mitnick and W. L. Simon, *The art of intrusion: the real stories behind the exploits of hackers, intruders and deceivers.*, W. Publishing., Ed. Indianapolis: Wiley Publishing, 2005.

[10] J. Debrosse and D. Harley, "Malice through the looking glass: behaviour analysis for the next decade," in *Proceedings of the 19th Virus Bulletin International Conference*, September 2009.

[11] G. L. Orgill, G. W. Romney, M. G. Bailey, and P. M. Orgill, "The urgency for effective user privacy-education to counter social engineering attacks on secure computer systems," in *Proceedings of the 5th Conference on Information Technology Education*, ser. CITC5 '04. New York, NY, USA: ACM, 2004, pp. 177–181. [Online]. Available: http://doi.acm.org/10.1145/1029533.1029577

[12] M. Bezuidenhout, F. Mouton, and H. Venter, "Social engineering attack detection model: Seadm," in *Information Security for South Africa*, Johannesburg, South Africa, August 2010, pp. 1–8.

[13] F. Mouton, M. Malan, and H. Venter, "Development of cognitive functioning psychological measures for the seadm," in *Human Aspects of Information Security & Assurance*, Crete, Greece, June 2012.

[14] D. Gragg, "A multi-level defense against social engineering," SANS Institute InfoSec Reading Room, Tech. Rep., December 2002.

[15] R. Bhakta and I. Harris, "Semantic analysis of dialogs to detect social engineering attacks," in *Semantic Computing (ICSC), 2015 IEEE International Conference on*, Feb 2015, pp. 424–427.

[16] M. Workman, "A test of interventions for security threats from social engineering," *Information Management & Computer Security*, vol. 16, no. 5, pp. 463–483, 2008.