Social Engineering Attack Framework Literature Review

Michael Pepper
Department of Computer Science
University of Cape Town
Cape Town, South Africa
pppmic005@myuct.ac.za

ABSTRACT

In modern society, the protection of personal information is of increasing importance [16]. Security measures are constantly evolving in order to prevent malicious attacks and the elicitation of sensitive information, however the human element remains a vulnerability in the system. Social Engineers attempt to exploit this vulnerability by compromising the emotional state of an individual, leading to the elicitation of sensitive information. Various attack frameworks have been proposed to model how these attacks occur as well as classification measures through which the attacks can be categorised. This paper critically analyses various attack frameworks and attack classifications with the aim of identifying the aspects of different attacks that the SEADMv2 detection model should be able to identify within the SEPTT project.

The attack framework proposed by Mouton [16] is found to be the most descriptive as it breaks down an attack into concretely defined phases with predefined objectives. This framework would enable real-life scenarios to be generated, which could be used to test the coverage of the SEADMv2. Two classifications for types of attacks are outlined, namely the manner of communication in the attack and the type of interaction between the parties involved. Generating attack scenarios that conformed to the subcategories within these classifications would ensure that the SEADMv2 be tested against a wide range of attack types, and hence its applicability to the SEPTT project assessed.

CCS Concepts

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

Keywords

Attack Classification; Attack Framework; Attack Scenario; Psychological Vulnerability; Social Engineering; Social Engineering Attack Detection Model

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

1. INTRODUCTION

The field of information security is a fast growing discipline, with the protection of personal information being of vital importance [16]. 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 [16], [17]. This is achieved through the use of psychological ploys which compromise the user's emotional state, hence allowing an exploit to take place [2], [13], [16]. 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 within users, social engineers can elicit responses and perform information gathering that would not be possible had the user been in a more stable state of mind [17], [2]. This ultimately leads to the attacker achieving a predetermined objective, often unbeknownst to the victim. Social Engineering is hence closely related to social psychology [15].

The Social Engineering Prevention Training Tool (SEPTT) project aims at implementing the Social Engineering Attack Detection Model (SEADMv2) proposed by Mouton et al. [14], in order to determine whether it is effective at successfully identifying social engineering attacks within any environment. The results of the implementation will either verify the model's coverage and prediction capabilities, or indicate that the underlying social engineering attack model that the framework was built upon is insufficient at modelling real-world attacks. Should the model successfully identify social engineering attacks, it could be used as a tool to prevent such attacks in various environments, hence reducing user exploitation.

The purpose of this paper is to perform a critical analysis of the available literature on social engineering attacks, in order to determine the aspects of attacks which the SEADMv2 should detect within the SEPTT implementation. Firstly, the differing phases within social engineering attacks will be identified by analysing attack taxonomies. Secondly, differ-

ent attack classifications will be analysed, highlighting similarities between attack implementations. These two sections will outline how attacks are performed and different types of real-world attacks, hence enabling accurate scenarios and examples to be generated. These scenarios will enable the SEADMv2 framework to be tested and its applicability for the SEPTT project assessed.

The remainder of this literature review is structured as follows: Section 2 analyses the different phases of social engineering attacks with reference to Mitnick's attack cycle [11]. Section 3 analyses the phases of an improved attack framework proposed by Mouton et al. [16]. Section 4 will outline different ways social engineering attacks can be classified. Section 5 discusses overall findings and section 6 concludes this literature review with a summary of the paper.

2. MITNICK BASED ATTACK PHASES

A Social Engineering (SE) attack can be defined as the use of techniques to exploit human psychological vulnerabilities in order to gather information and bypass information security systems [11]. These attacks are highly successful as often individuals do not perceive themselves as potential victims of such attacks and hence are not aware of the types of techniques used [14]. This ignorance can be attributed to their lack of knowledge of the potential gains an attacker can receive from the information they possess. Individuals may have the mindset that the information in their possession is not of any value to anyone, so why should they attempt to protect it [14]? Furthermore, some individuals feel they would be able to detect potential social engineering attacks however the social engineer is skilled at exploiting human vulnerabilities via psychological triggers in order to foil human judgement and attain information [17]. This section will detail the most common phases within SE attacks. Mitnick's attack cycle [11] will form the base structure of the analysis as its phases are common amongst most taxonomies.

2.1 Information Gathering

Initially, the social engineer gathers as much information about the target as possible [16]. This information gathering can take many forms and aims at acquiring information and resources necessary to successfully perform the 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 [16]. Techniques such as gathering Facebook pictures of the targets friends and identifying the language and tone used between the target and those friends are two techniques that could be used in this phase [1]. Such information would assist in masquerading as one of the targets friends in order to exploit their relationship and attain valuable information from that individual. The first taxonomy proposed by Harley [6] identifies other techniques that can be used in this phase, such as password stealing, dumpster diving, leftover, hoax virus alerts and other chain letters, spam and direct psychological manipulation. All of these techniques aid the attacker in attaining the information required to establish a relationship with the intended target. In Tetri & Vuorinen [19], this stage is referred to as *data gathering* and is one of the three dimensions in that model.

2.2 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 [16]. 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 [16]. 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. This stage is present in the taxonomies proposed by Mitnick [11], Laribee [9] and Tetri & Vuorinen [19].

2.3 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 by manipulating the targets emotional state by preying on the seven psychological vulnerabilities [5]. They are: strong affect, overloading, reciprocation, deceptive relationship, diffusion of responsibility and moral duty, authority, integrity and consistency [18], [10], [20], [3]. By exploiting these psychological vulnerabilities, the target's emotional state is altered and they become more likely to comply with the attackers requests for information [16].

The attack model documented by Laribee [9] groups the psychological techniques into manipulation, deception, persuasion and influence. Harley [6] defines a set of vulnerabilities consisting of gullibility, curiosity, courtesy, greed, diffidence, thoughtlessness and apathy. While the underlying psychological principles within these models may differ, the goal remains the same - to influence the individuals emotional state in order to solicit information from them. This phase is much the same in the Tetri & Vuorinen [19] however it is classified as *persuasion of the individual*.

2.4 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 [11]. 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.

3. IMPROVED ATTACK FRAMEWORK

This section will outline a proposed improved attack framework, based on Kevin Mitnick's attack cycle above [11]. The framework proposed by Mouton et al. expands on Mitnick's work by adding detail to the phases within the cycle, and defining the phases more concretely so as to make the cycle less open to interpretation [16].

3.1 Attack Formulation

As outlined in section 2, Mitnick's model suggests attackers gather as much information about the target as possible. This is true however it assumes that the target has already

been identified and the goal of the attack established. Mouton et al. suggest a prerequisite step in which the goal of the attack is established, and the target selected based on their ability of assisting in reaching the attack goal [16]. The target can be an individual or a group.

3.2 Information Gathering

The information gathering phase follows much the same process as in Mitnick's attack cycle above, with the aim of improving the chances of establishing a relationship with the target. This model elaborates and places more emphasis on the sources of information. Firstly potential information sources are identified, whether publicly or privately available. Once identified, the information gathering takes place and the information attained is assessed for relevance. This continues until sufficient information is attained.

3.3 Preparation

The authors of the model proposed an intermediary phase between *information gathering* and *development of rapport and trust*, mainly to allow for the attained information to be consolidated and the attack vector developed [16]. The consolidated information allows for pre-texting of the scenario that will force the target into the required psychological state and is hence of great importance. The result of this step is an attack vector that contains all the elements of a social engineering attack [13]. This vector contains the medium through which communication will take place and the compliance principles to be used [16].

3.4 Develop Relationship

Similar to Mitnick's model, this phase focuses on establishing the relationship with the target by means of the attack vector and the information gathered. This phase is divided into two stages, namely establishment of communication between attacker and target, and rapport building between the two parties. The communication is established using the medium identified during the preparation phase. Once established, the relationship can be developed using the techniques outlined in the model above.

3.5 Exploit Relationship

Now that a trusting relationship exists between the attacker and target, the exploitation can commence. This is achieved in the same fashion as Mitnick's model, through the exploitation of psychological vulnerabilities. For this exploitation to be successful, the target need be in an emotional state where exploitation is possible [16]. Getting the target to this state is referred to as "Priming the target", and the state need be congruent to that required by the attack vector. Once in this state, information extraction from the target should be possible. The attacker commences by probing the target for the required information.

3.6 Debrief

Lastly, the model proposes a final stage in which the target is returned to a desired "stable" emotional state [16]. In doing this, the aim is to make the target feel reassured that they were not under attack, and in a normal state of mind. Should this be the case, they will not reflect on the situation and hence may not identify that they were a victim of a social engineering attack. This is important for the success of the attack as no counter measures will be put in place to prevent the use of the obtained information. This phase is also present in the Laribee model whereby the relationship between attacker and target is preserved after the attack [9].

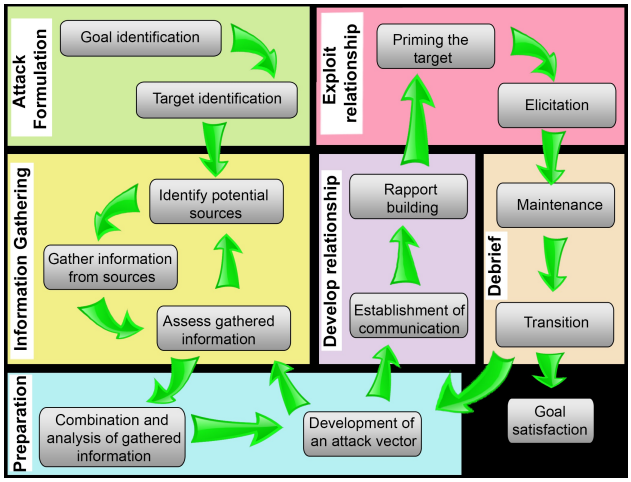

Figure 1: Mouton et al. [16] Improved Attack Framework

4. ATTACK CLASSIFICATIONS

This section will outline the different ways SE attacks can be classified according to the manner in which the communication takes place during the exploit, and the interaction between attacker and target.

4.1 Communication

A common classification criterion amongst the literature was how the communication between the parties took place during the attack. Ivaturi & Janczewski [7] classified SE attacks as being either *person-person* (direct communication involving a human) or *person-person via media* where some medium is involved in the communication. This notion of classifying attacks based on the communication within them is furthered by Mouton [13] whereby attacks are divided into direct and indirect. 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 in Ivaturi & Janczewski [7]. 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 has no way to communicate back with the attacker. Examples of the mediums used include emails and one-way text messages.

	Mitnick [11]	Laribee [9]	Harley [6]	Tetri & Vuorinen [7]	Mouton [16]
Attack Formulation	-	-	-	-	Yes
Information Gathering	Yes	-	Yes	Yes	Yes
Preparation	-	-	-	-	Yes
Develop Relationship	Yes	Yes	-	Yes	Yes
Exploit Relationship	Yes	Yes	Yes	Yes	Yes
Debrief	-	-	-	-	Yes
Extra Phases Identified	*Information Utilisation*	-	-	-	-

Table 1: Attack Phases Identified In Different Models

4.2 Interaction

Mohd et al. [12] classifies social engineering attacks based on whether they were *human-based* or *technical-based*. Human-based attacks can be likened to the *person-person* attacks defined in the Ivaturi & Janczewski model above, and deal with the use of persuasion techniques during a physical interaction. Note that the classifying factor is the manner of interaction, as this model does not deal with the communication itself and its directionality. The technical-based attacks that the model identifies can be likened to *person-person via media* whereby email, software and websites are the mediums through which the communication and hence exploitation take place. Again the classifying factor is the manner of interaction and is not limited by whether the communication is direct or indirect as in Mouton [13].

Khrombholz's approach [8] can be viewed as a concatenation of the above two approaches as an SE taxonomy is proposed whereby attacks are classified according to three categories: Channel, Operator and Type. Channels (referred to above as medium) include email, instant messaging, websites etc.. Operator refers to either humans or software and identifies the originator of the attack and is more closely related to the model proposed by Mohd et al. [12] . Lastly, the type of attack is categorized into four types: Physical, Technical, Social and Socio-technical.

5. DISCUSSION

Upon analysis, it is obvious that the model that defines the structure of SE attacks in the most detail is that proposed by Mouton [16] in Section 3. This can be attributed to the decomposition of each aspect of an SE attack, resulting in distinct phases and goals for those phases. Table 1 illustrates this as the overarching phases identified by the various models considered in this paper can be seen, with Mouton's clearly being the most descriptive and low level. This concrete framework of the stages within an SE attack enable a greater understanding of the amount of work that goes into successfully implementing an attack, as well as how each phase aids in the execution of the next.

This rigid understanding of the events and actions that culminate in an SE attack makes this model the most well suited for generating real-world SE attack scenarios which could be used to test the SEADMv2's coverage and accuracy within the SEPTT project. By generating scenarios according to the phases identified in Section 3, the most in-depth and realistic scenarios can be generated, without neglecting any aspect of the attacks by relying on assumptions. Testing the SEADMv2 against these scenarios would highlight the lacking areas within its detection framework and hence the vulnerabilities of the SEPTT project.

Mouton's model [16] achieves this highly segregated definition of SE attacks by being built on the major phases outlined in Mitnick's attack cycle [11], and adding on further phases that are crucial in the overall success of the attack. Mouton's model also breaks down large phases into their constituent sub-tasks and outlines the relationships between these sub-tasks, resulting in his model being the most thorough of the available literature. One could assess the SEADMv2 with reference to the other models dealt with in this paper, however as the attacks that comply with said models are not representative of real-world attacks, the detection framework would not be tested in a useful way.

Section 4 deals with the type of attack independent of its overall structure and proposes that attacks be divided into two main groups according to the communication used in the attack and the manner of interaction between the parties involved. One should generate attack scenarios that fall into the subcategories of both these classifications when testing the SEADMv2, as they both model attacks that would need to be detected in the SEPTT project. By considering attacks from both classifications, the widest possible range of attacks can be simulated and hence the applicability of the SEPTT project for real-world detection can be assessed. These attack scenarios should be generated in accordance with Mouton's model [16].

6. CONCLUSIONS

The protection of personal information is extremely important in modern society. Measures are put in place to ensure the protection of this information however skilled individuals manage to bypass them and attain the information they desire through the exploitation of some weak point in the system. This paper focussed on the exploitation of the individual as the weak point in the system, rather than the technology itself, and identified the stages that culminate in the successful elicitation of personal information. This was achieved by reviewing the available literature on social engineering attack frameworks and analysing their constituent phases. The techniques used within each of these phases was noted with reference to how they aid in achieving the final goal. The underlying psychological vulnerabilities identified in the different models was noted as well as how the social engineer manipulates a target's emotional state to enable information extraction.

This was achieved by assessing Mitnick's attack cycle [11] as it forms the basis of most attack models, and identifying the differences between it and similar models. An improved framework proposed by Mouton [16] was analysed, identifying the subtasks within each phase and how they interact to achieve the overall goal. The different classifications for

social engineering attacks were analysed, outlining the subcategories within each, and the types of attacks that would fall under each category.

It was noted that Mouton's model [16] was the most accurate depiction of how real-world attacks are performed, due to its concrete definition of each phase and the tasks performed within those phases. It was also noted that in order to test the SEADMv2 in a meaningful way, attack scenarios from each category in Section 4 should be generated, using this model as a structural guideline.

In conclusion, the available literature on attack frameworks and attack classifications were critically analysed and the aspects of social engineering attacks that the SEPTT project should be able to detect were identified.

7. 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] Harley, D. Re-floating the titanic: Dealing with social engineering attacks. *London: EICAR* (1998), 13.

[7] 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).

[8] Krombholz, K., Hobel, H., Huber, M., and Weippl, E. Advanced social engineering attacks. *Journal of Information Security and applications 22* (2015), 113–122.

[9] Laribee, L. *Development of methodical social engineering taxonomy project*. PhD thesis, Monterey, California. Naval Postgraduate School, 2006.

[10] 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.

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

[12] Mohd Foozy, F., Ahmad, R., Abdollah, M. F., Yusof, R., and Mas' ud, M. Generic taxonomy of social engineering attack.

[13] 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.

[14] 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.

[15] Mouton, F., Malan, M. M., Kimppa, K. K., and Venter, H. S. Necessity for ethics in social engineering research. *Computers & Security 55* (2015), 114–127.

[16] 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.

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

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

[19] Tetri, P., and Vuorinen, J. Dissecting social engineering. *Behaviour & Information Technology 32*, 10 (2013), 1014–1023.

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