TALLINN UNIVERSITY OF TECHNOLOGY

Department of Computer Science

TUT Centre for Digital forensics and Cyber security

TC70LT

Kevin Lwakatare IVCM122944

CONTRIBUTIONS OF UNDERSTANDING AND DEFENDING AGAINST SOCIAL ENGINEERING ATTACKS

Master thesis

Anton Vedeshin

Visiting lecturer, Tallinn University of technology

Tallinn 2016

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Kevin Lwakatare

24.05.2016

Acknowledgement

First and foremost, I would like to express my gratitude to my supervisor Anton Vedeshin for his guidance in formulating the topic of this masters' thesis. In addition, I would like to thank him for his useful comments, remarks and engagement. His support has introduced me to an interesting world of research with great possibilities.

Second, I would like to thank my opponent for agreeing to review my thesis and also in providing me with useful comments that helped to improve this work greatly.

Thirdly. I would like to express my appreciation to Tallinn University of technology and Tartu University for giving me the opportunity of participating in the graduate program that has now fueled my genuine passion in the area of cyber security. University staffs especially lecturers and my fellow students have been amazing throughout the program and I am grateful for that. Particularly, I would like to express my warm appreciation to Tallinn University of Technology for the tuition fee scholarship for my studies.

Lastly, I would like to thank my family- Dad, Mom, Karen, Muta, Kelly, Mwemezi, Jennifer and Lucy- for their support throughout my studies. In particular, I must acknowledge my young sister Lucy, whom without her love, support and encouragement, I would not have finished this thesis. I want to also thanks all others that have helped and provided me with support in completing studies at Tallin including friends and other family members.

Abstract

With the development of IT technology, in modern society, IT security has become an important reliance to the information security and privacy. Although there are a number of security approaches, such as firewall and intrusion detection system, which can be used to protect the machines from being attacked, there is a lack of widely accepted mechanism to prevent machine users from fraud. Social engineering is the attack meaning smooth communicating with victim to reveal valuable information in order to bypass the secure perimeter in front of the information-related resources. To the best of author knowledge, risk cannot be eliminated thoroughly, but can be reduced and controlled. In this thesis, a novel taxonomy of social engineering attacks was proposed in order to understand the concept of the social engineering and gain insight of the representative social engineering attacks through applying the taxonomy to them. Furthermore, a multi-layer social engineering defense model is proposed to deal with the threats brought by the social engineering attacks. In each layer, different mechanisms are proposed respectively to facilitate the defense against to various social engineering techniques in order to effectively protect information-related resources and guarantee IT security.

Keywords: Social Engineering, Psychological Vulnerabilities, Information System, IT Security, taxonomy, detection model, defense model.

This thesis is written in English and is 66 pages long, including 6 chapters, 12 figures and 4 tables.

Annotatsioon

Seoses IT tehnoloogia arenguga kaasaegses ühiskonnas on IT turvalisus muutunud oluliseks usaldusväärseks abiks infoturbele ja privaatsusele. Kuigi turvalisusele on mitu lähenemist, nagu näiteks tulemüür ja sissetungimise tabamise süsteem rünnakute eest masinatele, puudub laialdaselt aktsepteeritud mehhanism ennetamaks arvutikasutajaid pettuse eest. Tehnosotsiaalne sahkerdamine on rünnak, mille käigus paljastatakse väärtuslik informatsioon läbi ohvriga suhtlemise. Selle informatsiooni abil mõõdetakse turvasüsteemidest erinevate ressursside ees. Tuginedes autori teadmistele riski ei ole võimalik ohtu täielikult kõrvaldada, kuid on võimalik seda vähendada ja kontrollida. Käesolevas magistritöös pakutakse välja tehnosotsiaalse sahkerdamise rünnakute uudne taksonoomia, eesmärgiga mõista tehnosotsiaalse sahkerdamise mõistet ja saada ülevaade tüüpilistest tehnosotsiaalse sahkerdamise rünnakutest neile taksonoomia rakendamise kaudu. Lisaks eelnevale esitatakse mitmekihiline tehnosotsiaalse sahkerdamise kaitsemudel tegelemaks tehnosotsiaalse sahkerdamise rünnakute poolt kaasnevate ohtudega. Igas kihis pakutakse välja erinevad mehhanismid et lihtsustada kaitset erinevate tehnosotsiaalse sahkerdamise tehnikate vastu, eesmärgiga kaitsta tõhusalt infoga seotud ressursse ja tagada IT turvalisus.

Märksõnad: Tehnosotsiaalne sahkerdamine, psühholoogilised haavatavused, infosüsteem, IT turvalisus, taksonoomia, tabamise mudel, kaitsemudel.

Table of abbreviations and terms

SE	Social engineering
SET	Social engineering toolkits
PC	Personal computer
IDS	Intrusion detection system
XSS	Cross-site scripting
Phreaker	Hackers pioneered the art of human hacking

Table of contents

List of figures

List of tables

1. Introduction

Information security and privacy are very important to personal assets, corporate properties, and even state secrets, however, various hacking threats are still prevalent [1] [2][3] (N. Perlroth)[4]. In modern society, people use various digital equipment, such as cell phones, laptops, tablet pads and personal computer, that are connected to the Internet in order to communicate and share information with each other. Hence, with the development of the IT technology, modern information security and privacy have been harmoniously combined with IT security (Hossein Bidgoli) [5] (Ji-Xuan Feng) [6]. IT security includes not only protecting organization's systems from being attacked but also preventing system users from being tricked into leaking valuable information amongst other things.

1.1. Research motivation

Due to the intelligence of blackhat community (Honeywall project) [7], today, there are many hacking techniques, such as buffer overflow, SQL injection, and cross-site scripting (XSS) that, can be used to attack computer systems in order to access sensitive information (R.C. Joshi) [8]. Most attacks exploit system's vulnerabilities, which are often addressed through timely system updates and the use of security devices such as firewall and intrusion detection system (IDS). On the other hand, some hackers pioneered the art of human hacking (also called phreakers in some earlier articles (K. Mitnick) [32]) known as social engineering (SE) attacks to deceive the victims in order to get the valuable information, such as account names, ID numbers and even passwords, which can be used to bypass the access control and avoid intrusion detection. SE attacks are much more difficult for system administrators to defend against. A study from [9] shows that 48% of large companies and 32% of small companies across the globe were victims of 25 or even more SE attacks between 2009 to 2011. Additionally, the SANS institute reported that network attacks cost U.S. companies $266 million every year and 80% of them are SE attacks [10].

A simple example of SE could be like this: a phreaker launches smooth talking to a victim and get a pair of legitimate account and password. Afterwards, the phreaker intrudes an inside system through authorized login and obtain secret information in terms of certain motives. This type of intrusion will be invisible to the secure perimeter. A more sophisticated example could be like case [11]. Thus, SE attacks have become a very important network threats to various organizations and enterprises.

1.2. Research questions

In order to defend against the SE attacks, some research questions have to be taken into account. The research questions are proposed from a meta-how question that gets broken down into several deduced sub-how question, which can be described as follows:

- Meta-how: How to make contribution to the study of social engineering?
  - Sub-how1: How to classify different social engineering attacks?
    - What are the attacking phases of social engineering attacks?
    - What are the terminologies of social engineering attacks?
    - What is the classification scheme of social engineering attacks?
  - Sub-how2: How to defense social engineering attacks?
    - What is the decent security model can be used to defend against SE attacks?
    - Where should be set security mechanism to protect from being attacked by social engineering?
    - What security mechanism should be design?
  - Sub-how3: How to measure the security level to social engineering attacks?
    - What is the metric of measurement?
    - What are the criteria to design awareness questions?

The sub-how questions will be respectively assigned to the contribution of the chapter 3, 4, and 5. The respectively assigned sub-how questions go through a "splitting function" inside the respective chapters to again deduced what-questions. Therefore, besides of the state of the art, the main contributions of this thesis are the answers to these research questions.

1.3. Research methodology

This master thesis applies a research methodology based on the action-design research (ADR) method (Maung K. Sein) [12], which is a research method for generating prescriptive design knowledge through building and evaluating ensemble IT artifacts in an organizational setting.

1.3.1. Problem Formulation

This master thesis first reviewed the related work about SE security projects and research, including the SE taxonomies and conceptual models, the SE defense models and the measuring approaches. Thereafter, the problem space of the SE is presented and the increased security risks are identified. So, in order to approach the objectives for studying SE attacks, the main work of this thesis can be revealed from the problem space.

1.3.2. Proposal Design

According the outcome of the problem formulation, the thesis focuses on three main proposals. First, a novel taxonomy of SE attacks is proposed, which can be used to classify and analyze the SE attacks, and even unveils the SE attacking space in order to provide some designing directions for security researchers to defend against social engineering based attacks. Second, based on the proposed taxonomy, some security approaches and models for defending against SE attacks can be improved. Therefore, a new SE defense model is proposed, which includes three layers: prevent, detect, and control. Each layer consists of several techniques and mechanisms to implement the defending functionality. This multi-layer SE defense model can be extended through supplementing new techniques. Third, the thesis presents some suggestive measuring approaches for the security level against the SE attacks. These measuring approaches including the surveys, questionnaire and even interview to test the employees' awareness and even some developed applications which can perform automated SE attacks to test the security level and train the employees' awareness against SE attacks.

1.3.3. Formalization of Learning

The outcome of the taxonomy of social engineering attacks can be used to generally classify any type of social engineering attacks. The taxonomy is a basic framework that can be extended through adding new class of concepts of social engineering attacks. The proposed social engineering defense model is a formulation that can be used to defend general social engineering attacks. The defense model can also be extended through adding new security mechanisms for future novel social engineering attacks.

1.4. Research objective

The objective of this thesis work is to understand, measure and defend information systems against SE attacks through investigating the presented SE taxonomies, and related conceptual, attacking and defending models. The contributions of this thesis can be summarized as follows:

1. The related work of SE attacks is reviewed, including the taxonomies, a variety of models, measuring approaches and even some security standards in order to gain insight of the study field.
2. A novel taxonomy of social engineering attacks is proposed, which we apply to multiple types of social engineering attacks to make classification.
3. A multi-layer security model is proposed for effective defending against social engineering attacks in order to guarantee IT security and protect information-related resources.
4. Some measuring approaches are presented for the security level against the social engineering attacks.

1.5. Thesis organization

The remainder of this thesis is organized as follows: chapter 2 reviews related work; chapter 3 proposes a novel taxonomy of social engineering attacks and applies it to multiple representative SE attacks; chapter 4 proposes a multi-layer social engineering attacks defense model; chapter 5 suggest several measuring approaches to evaluate the security level under using the SE defense model; chapter 6 makes a conclusion and presents some future work.

2. Related work

In this chapter, the state of the art about social engineering will be presented. The materials will be organized into several subsections, i.e. SE taxonomies, measuring approaches, and SE related models, in order to gain insight of the study field and help us to discover the areas that can be improved as well.

2.1. Social engineering taxonomies

Today we can learn more about social engineering attacks concepts, techniques with interesting real cases from various materials that are readily available such as books (Tipton) [13] (Hadnagy) [14] (Mann) [15], Particularly, we can know more about the study field by studying its dedicated taxonomies which constitute to theoretical study of classifications. To the best of author's knowledge, very little taxonomies that ware specifically designed for social engineering attacks existed a decade ago. However, in that period there existed a number of network attacks taxonomies (Lough) [16] (Hansman) [17] and secure computing [18]. Thereafter, some succeeding taxonomies of network attacks began to consider the classifications about social engineering. For example, Simmons et al. (Simmons) [19] proposed a taxonomy called AVOIDIT, which classifies cyber-attacks into six categories: attack vector, operational impact, defense, informational impact, and attack target. The attack vector is a vulnerability or path used to compromise a system, such as misconfiguration, buffer overflow, insufficient authentication validation, etc., and one of the subcategory of the attack vector is social engineering. Another taxonomy proposed by Van Heerden et al (Heerden) [20] consists of twelve classes and each class containing multiple subclasses. The social engineering is one of the subclasses of the class Attack Mechanism. Hence, both of AVOIDIT and Van Heerden's taxonomy simply treated the social engineering as one of the attacking methods but did not unveil the technique details about social engineering attacks.

Actually, there are some more details that can be studied in the social engineering attacks. In recent years, several novel taxonomies focusing on social engineering attacks have been proposed, which can help us to get out of the mess.

In 2014, Katharina Krombholz et al. proposed a novel taxonomy (Krombholz) [21] aimed to classify social engineering attacks. This taxonomy proposed three main categories for dissecting social engineering, and they are: channel, operator and type. The channel means the medium where the SE attacks conduct. The channel consists of e-mail, instant messaging, telephone, VoIP, social network, cloud and website. The Operator indicates the actor who launches the SE attacks, which can be human or software. The type refers to the approach that the SE attacks carry out. The taxonomy concludes four approaches: physical, technical, social and socio-technical. Furthermore, the author summarized seven representative SE attack vectors (or scenarios): phishing, dumpster diving, shoulder surfing, reverse social engineering, waterholing, advanced persistent threat, and baiting (all these vectors will be described in chapter 3), nevertheless the author mentioned the fact that the individual SE attack scenario have not been technically exhausted. In order to verify the taxonomy, the author applied it to these representative attack scenarios, which proves the taxonomy works well in analyzing these typical SE attack vectors. Indeed, it is a scenario-driven taxonomy, which draws out the attacking characteristics from the actual attacking scenario and then categories the characteristics into taxonomy. However, this taxonomy is designed mainly from the attack point of view, lacks of the characteristics of defending methods against SE attacks.

Another novel taxonomy of SE attacks was proposed by the paper (Heartfield) [22] in 2015. Just in contrast to the Krombholz's work, this proposal designed the taxonomy from a defense perspective. It adopts three distinct control stages, which includes orchestration, exploitation, and execution, defined by CESG [23] as the basic categories of the taxonomy. For each stage, it poses questions that can help to develop the technical protection mechanisms. The answers to these questions compose the corresponding categories, which consequently establish the whole taxonomy. The Orchestration consists of target type (target of choice or opportunity), attacking mode (manual or automated), and attacking approach (software, hardware without software or hardware with software). The Exploitation includes the deception vector (cosmetic, behavior, or hybrid) and the manipulation interface (user interface or programmatic interface). The Execution is comprised of execution steps (single or multiple) and attack persistence (one-off or continual). Hence, this taxonomy is defense stage driven, and in each stage, it depicts several mutual-exclusive sub-categories whose characteristics should be considered in order to develop the technical protection mechanisms. The taxonomy is not exhaustive and can be expanded based on the three main categories. The taxonomy was evaluated by applying to 30 different attacks observed in the wild, which is aimed to help in developing the technical protection mechanisms.

In addition, the paper (F. Mouton) [24] proposed an ontological model to define the social engineering domain. The set of the categories provided by this social engineering ontology can be considered as the taxonomy of social engineering. This model defines a social engineering attack, using either direct communication or indirect communication, has one social engineer, one target, one or more compliance principles, one or more techniques, one medium, and one goal. These six entities are defined as classes in the model.

2.2. Measuring approaches

It is important to study the methods of measuring the security level to social engineering, because one cannot protect what cannot measure. At present, the great percent of information security depends on the individuals rather than the technical security measures [25] (Jagatic) [26]. Hence, unlike the traditional technical vulnerabilities, social engineering attacks focus on human being's psychological vulnerabilities.

The individuals in charge of information security in an organization are all the employees. The paper (Manjak) [27] measures the value of the information security awareness efforts in the organization through applying various social engineering tactics targeting the employees that an information security awareness campaign is designed to counter. Although employees' awareness can be improved by security education, there are several obstacles for education employees, such as poor awareness program design, lack of executive level support, budgetary constraints, etc. However, the security awareness is not training. It actually relies on the employees' consciousness and attitude. Employees will not be motivated to improve the security awareness if they see no reason to concern the information security. Hence, in order to address the employees' ignorance, the first task is to convince staff that they have personal stock in the effort to secure the organization's information assets. In addition, the behaviors that would put the organization's assets at risk have to be identified. Through several information collecting methods, such as surveys, questionnaires and interviews, the core message can be identified for avoiding risk behaviors.

Using the standard security policy is an effective way to help the organizations to train the employee and control security risk. At present, there are a number of standard security standards, such as the "Big Five": ISO/IEC 27001, BS 7799, COBIT, PCI DSS, ITIL&ISO 2000. However, few of them include policies covering social engineering attacks. ISO/IEC

27032[1], extended from ISO/IEC 27001, is a completely new international standard published by ISO that covers the baseline security practices for all stakeholders in cyberspace. In particular, it provides technical guidance for addressing social engineering attacks. Thus, the organization concerning information security can choose ISO/IEC 27032 to implement the cyber security framework to prevent social engineering attacks. However, this new security standard still remains to be seen how it will turn out in practice and how widely it will be accepted. Using the security standards makes an ease of security measurement. ISO27004[2] defines a measurement method with the following steps: 1. Complete list of the controls implemented in accordance with Annex A of ISO27001 standard; 2. Method for measurement of attributes associated with controls; 3. Base measure for the control attributes; 4.Generation of the indicator.

Marcus Nohlberg et al. proposed a new interview protocol (Nohlberg) [28] to measure the readiness of an organization using security policy to deal with automated social engineering attacks. The interview protocol consists of a questionnaire which covers 15 areas of the conceptual model of SE (Nohlberg) [29], and the assessment of the response to the questions was reported based on a well-established behavior/attitude/knowledge triad matrix. Each item of the triad has a 3-level grading scale: none (0), informal (+1), formal (+2). Thus, the best possible score for an organization are six points in each of the 15 areas, which is equivalent to 90 points in total.

Indeed, the behavior/attitude/knowledge triad was inspired by the paper (Kruger) [30], which proposed a prototype for assessing information security awareness. This paper presented the methodology used to develop the measuring tool through answering two questions: what to measure and how to measure. It proposed a tree structure of problem that is developed based on three dimensions: attitude, knowledge, and behavior, to answer the first question. The three dimensions were technically borrowed from the field of social psychology which proposes that learned predispositions to respond in a favorable or unfavorable manner to a particular object have three components: affect, behavior and cognition. It also proposed a questionnaire based scoring model to answer the second question. In the scoring model the awareness also has a 3-level grading scale: good, average, and poor, which have the score of measurement 80-100, 60-79, 59-less, respectively. Although it was a generic assessing prototype and was not aimed to measure the security awareness program to social

[1] http://www.iso27001security.com/html/27032.html

[2] http://www.iso27001security.com/html/27004.html

engineering attacks, the proposed methodology is valuable to help other security researchers to develop novel measuring tools.

The work mentioned above mainly focused on discussing the measuring methodology, but consider little or did not unveil the detail of how to set the questionnaire and what kind of questions should be asked. The paper (Cheng) [31] presented that security metrics can be considered as a standard (or system) used for quantitatively measuring an organization's security posture, so in order to get an accurate assessment, simple but meaningful metrics are necessary. Although it proposed security metrics for general cyber situational awareness, it still can inspire the security researcher to design effective metrics for the awareness program against social engineering attacks.

2.3. Social engineering related models

In this subsection, SE related models, such as conceptual models, detection models and protection models, are presented in order to study how social engineering attacks can be prevented.

2.3.1. Conceptual models

A taxonomy is also a conceptual model. However, in this subsection, several dedicated conceptual models of social engineering will be presented.

Mitnich [32], proposed a conceptual model that describe social engineering attack cycle (SEAC) from the perspective of attackers, however, the SEAC model is explained too briefly and lacks detailed explanations. Based on that, a new model which describes the cycle of deception from the perspective of attacker, defender and victim is proposed by Nohlberg and his (Nohlberg) [29]. The model has five steps in each cycle. That is, if an attacker is not able to meet the requirement in one step, his attack will fail. Similarly, if one of the steps in the defense cycle can stop the attacker, the attack will fail as well. Otherwise, the attacker will be successful and even is going to be able to do it again. This model can be used to build defenses or to map and describe an attack. The paper (Mouton) [33] proposed another social engineering attack framework combining the previous SE ontological model (Mouton) [24] and extended Mitnick's social engineering attack cycle through specifying attack steps. It provides full details in every attacking step and make it can map historical SE attacks to a standardized format.

A system archetype is a good way to conceptualize a warfare framework of social engineering through describing the relations between the system, the countermeasures and the intruder. The paper (Gonzalez) [34] uses system archetypes as idealized patterns to describe the main modes of social engineering attacks. From both the attack and defense perspectives, the presented system archetypes unveil the two feedback loops, called controlling balancing (B) and reinforcing (R) loops, whose four basic combinations can be used to describe the intended consequence (IC) of the social engineering attack and the unintended consequence (UC) as organizational defense. The UC is the result of the organizational reaction to the SE attacks. However, SE attackers also have solution loop (SOL) to deal with the reaction of the organizational reaction, and always seek the ways to outsmart the single-loop defense lines. So, the paper suggests designing organizational security controls that can provide multi-layer feedback against the combined action of SE attacker's IC and SOL.

The system archetype approach is good at conceptualizing the SE to a high level of abstraction. However, the power of its analysis still remains questionable in terms of clarifying the techniques in detail. The paper (Tetri) [35] proposed a conceptualization of SE consisting different dimensions of SE can be used to exam the techniques of SE. Through reviewing the techniques used in actualizing the attacks, the paper extracted three different dimensions of SE techniques: persuasion, fabrication, and data gathering. After that, it proposed an abstract SE framework: intruder-techniques-dupe. The authors emphasized that in real scenario the SE attacker would use multidimensional approaches to attack organization, which can prove, in particular case, the information security policy is the weakest link rather than the human element.

In particular, Sherly Abraham et al. developed a framework (Abraham) [36] that shows the steps social engineering malware executes to be successful. Indeed, this paper reveals some malware using social engineering channels to be activated, which includes psychological and technical ploys. The psychological techniques include some persuasive tactics as well, such as using the victim's curiosity, empathy, excitement, fear and greed. The authors claimed that, although it is important for organizations to build comprehensive information security program, the SE malware cannot be mitigated by organizations alone, the shared responsibility of governments, ISPs, end users, and international bodies is needed to combat SE malware.

2.3.2. Detection models

The target of social engineering attacks is often the employees who have limited knowledge about the information technology infrastructure. The paper (Bezuidenhout) [37] proposes a social engineering attack detection model (SEADM), which can be used by the workers to detect social engineering attacks from the requesters in a call centre environment. The authors claimed that the social engineers often use psychological vulnerabilities to influence the victim's emotional state and cognitive abilities in order to get objective information. Hence, in order to enhance the individual's awareness to the social engineering requests, the paper proposed an automated self-evaluation electronic questionnaire. If the individual is detected to emotional, the call or the email request will be elevated to another individual. However, this strategy could initiate the work responsibility shift and even promote further frustration with all individuals involved.

In the previous SEADM, the determining of one's emotional state is subjective, and it is impossible to make instantaneous decision whilst working under pressure. Thus, the paper (Mouton) [38] improved the SEADM by proposing and incorporating a cognitive functioning psychological measure in order to determine the emotional state and decision-making ability. This paper applied three psychological measures from Psychology Experiment Building Language (PEBL): Wisconsin Card Sorting Test, Eriksen's Flanker Test and the Dot Judgment Test. In order to incorporate the psychological measures into the SEADM, the paper suggests the individual to apply two shorten versions of the three measures, one at the initial state and the other at the end state of the SEADM. Due to each measure only spends up to third seconds, which will not compromise the efficiency in the call environment.

The previous two papers related to SEADM both focus on the call centre environment using bidirectional communication. Hence, the paper (Mouton) [39] therefore proposes a revised version of social engineering attack detection model, namely SEADMv2, through extending the model to be able to cater for SE attacks that use bidirectional communication, unidirectional communication or indirect communication. Thus, the SEADMv2 is more capable and can be applied to map much more general SE attacks.

Inspired by the generic network intrusion detection system (IDS), which uses the signature-based approach to detect the malicious network traffic, the paper (Bhakta) [39] present a novel approach based on a pre-defined Topic Blacklist (TBL) to detecting social engineering attacks by verifying whether the discussion topics of each line of text generated by the potential attacker matches the topic in the TBL. The experimental results show that this approach has high detection accuracy and low false positive rates.

2.3.3. Protection models

The paper (Gragg) [41] firstly defined seven psychological vulnerabilities: strong affect, overloading, reciprocation, deceptive relationships, diffusion of responsibility and moral duty, authority, integrity and consistency, and then it defined a multi-level defense that will address these psychological triggers. The multi-layer defense includes security policy, security awareness training for employees, resistance training for key personnel, ongoing reminders, social engineering land mines (SELM), and incident response.

The study (Mataracioglu) [42] proposed a qualitative method called security lifecycle model against social engineering attacks (SLM-SEA). Although this approach still mainly focuses on enhancing the individuals' awareness to prevent social engineering, it proposed a comprehensive model consisting of user training, testing, measuring, and result feedback, which fits the Plan-Do-Check-Act (PDCA) cycle using in all ISMS [43] processes. Hence, the PDCA cycle may work in every management system.

2.4. Conclusion

Therefore, through reviewing the state of the art of social engineering, a solid basic knowledge of the study field can be obtained. Although, the present SE taxonomies can be improved into a much more complete one, which combined with the proposed conceptual models and frameworks of social engineering provide us a formalized knowledge in this study field. They are also the basement of the measuring approaches.

Due to the targets of the SE attacks are the individuals involved in the information security, the target of measuring approaches is the awareness programs, the applied security standards and the SE security models which educate and train the employees of the organization. On type of measuring approach is to launch the real SE attacks to the target organization, but this method has law and ethics problem. Hence, an ethical methodology of measuring the value of the information security awareness efforts is questionnaire based interview to the employees of the organization. Beside, the evaluation value depends on the good metrics that can make the questionnaire simple but meaningful.

Nevertheless, currently most of the SE attacking defense approaches focuses on enhancing the victim's awareness from the psychological perspective. There is a lack of a comprehensive and effective social engineering defense model that can provide multi-layer defense including security policy, individual education, even automated protection mechanism, and even incident response, etc, in order to mitigate the SE attacks to organizations.

3. Taxonomy of social engineering

Taxonomy is the theoretical study of classification. A taxonomy of the social engineering attacks can help researchers to explore the problem space and evaluate the applicability and scope of proposed solutions for a variety of current and future threats. In this chapter, a novel taxonomy of social engineering attacks is proposed and it is applied to several typical SE attacks as case study in order to validate its effectiveness in classifying real instances.

3.1. Classification scheme

As we all know, a general network intrusion can be divided into five steps: reconnaissance, scan, exploit, gain access and maintain access. For the social engineering attacks, specially, Kevin Mitnick's model (Mitnich) [32] proposed another five steps: research, developing trust, exploiting trust and utilize information. Similarly, Gartner proposed a social engineering attack cycle (Gartner) [44]: information gathering, relationship development, exploitation and execution. Furthermore, some other proposals (Heartfield) [22] (Nohlberg) [29] dissected the social engineering attack into several phases in order to discover the common characteristics. Admittedly, it is a good way to analyze the SE attack in order to abstract the categories to form the classification scheme. However, these proposed taxonomies and conceptual models put too great emphasis on describing the step-by-step approaches while giving little support to the characteristics in each attacking step, which cannot greatly help gaining an insight of the social engineering attacks, and is also not be able to help developing effective defense strategies due to the weak investigation capability.

The objective is to provide a taxonomy of social engineering attacks that can be used as both a holistic study and a foundation for developing a SE defense model. Therefore, the novel taxonomy of social engineering attacks including three phases: orchestration, exploitation, and compromise, as the main categories of the taxonomy. However, in each phase, enough subclasses are provided in terms of the characteristics of the phase in order to fully analyze and classify the real instances. Figure 1 presents an overview of the novel proposed taxonomy of social engineering attacks.

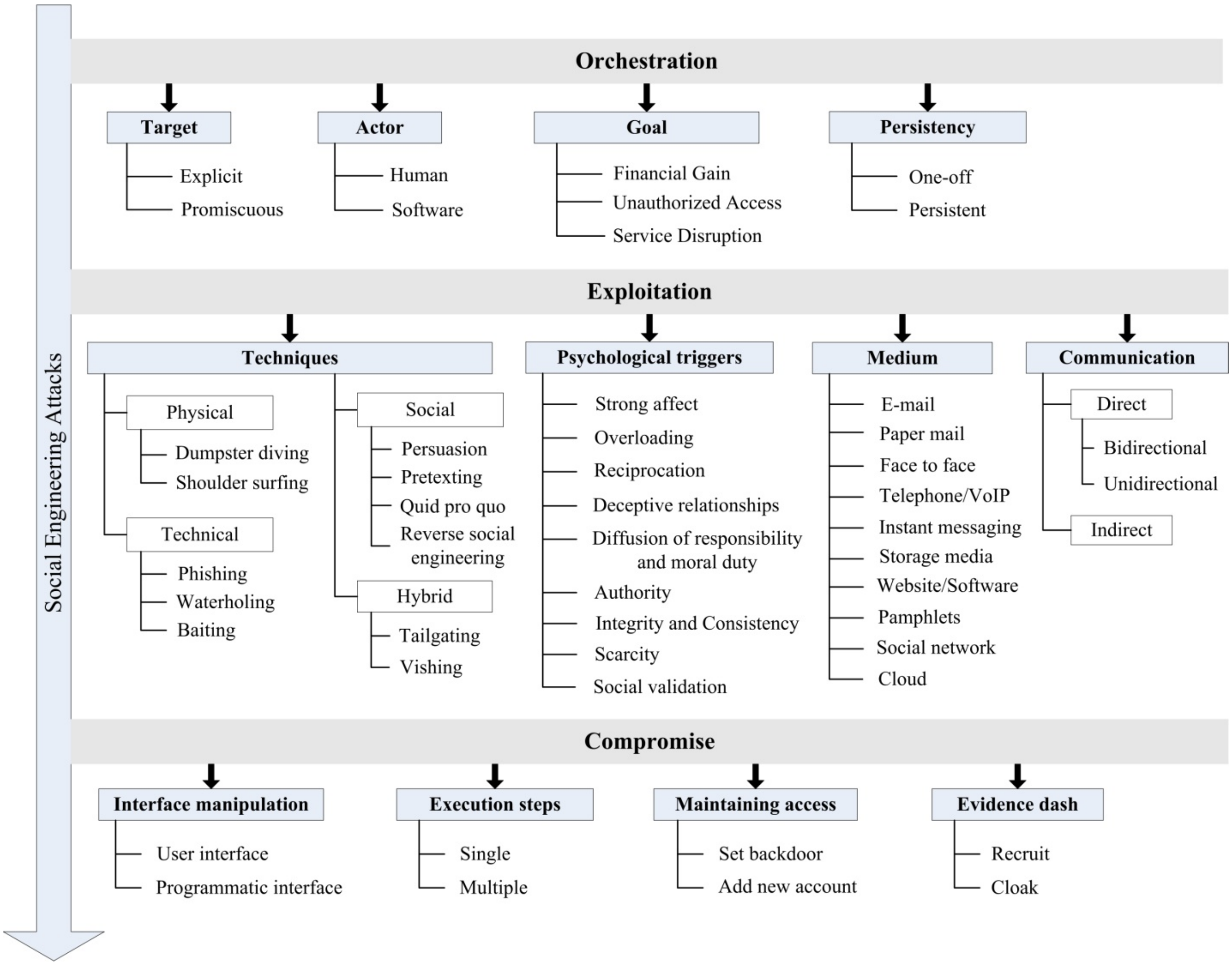


Figure 1. An overview of the novel taxonomy of social engineering attacks

Nevertheless, some principles need to be taken into account in order to define a high quality taxonomy. These requirements were listed in a comprehensive list, which can be briefly concluded into a number of terms: acceptable, comprehensible, completeness/exhaustive, determinism, mutually exclusive, repeatable, consistency, will defined, unambiguous and useful. The detail definitions of these terms can refer to in paper (Hansman) [17]. The taxonomy also complies these principles, however, note that it is hard to say the presented classification scheme is completeness or exhaustive, because it is infeasible to excavate all the classes that can cover the increasing number of characteristics in the study field. So, the taxonomy is a framework that can be extended by adding new classes to meet the requirement of completeness.

In the next three subsections, the three phases with their subclasses will be described in detail.

3.1.1. Orchestration

The orchestration phase includes the arrangement of the attacking target, actor, goal and persistency.

Target (TG)

This class refers to the methodology applied to select the attacking target in the orchestration phase. An **explicit (TG1)** attack aims to the target of choice, which means an adversary has a specific object of attack. So, the adversary will do a lot of explicit research and reconnaissance about the specific target in order to collect the dedicated information for attacking the object. In contrast, a **promiscuous (TG2)** attack aims to the target of opportunity. In this case, the adversary aims to a group of general objects that have the opportunity to be attacked, rather than focuses on attacking a specific one.

Actor (AC)

This class presents the entity originating the attack. A **human (AC1)** can directly conduct an attack, though the concurrent number of attack is limited due to the low capacity of individual operation. Thus, in this case the adversary has to manually conduct the attack step-by-step. **Software (AC2)** has a higher capacity compared to the human, so it can automatically conduct multiple attacks in the same time. Hence, the software is able to attack a number of targets simultaneously in a short time.

Goal (GL)

This class indicates the purpose of the social engineering attack. Admittedly, there are a number of SE attack goals. Here, three common and widely accepted goals are listed. The **financial gain (GL1)** refers to financial or other gain such as stealing cash or credit cards, manipulation of the stock market, and even corporate espionage attempting to gain a business competitive advantage. The **unauthorized access (GL2)** means gain access to the sensitive information by breaching the access control, such as access the sensitive information through using an inside legitimate pair of account and password or exploiting vulnerability to bypass the secure perimeter from outside. Instead of gaining some value, the **service disruption (GL3)** is aimed to break the functionality in order to make the service stop working due to some motives such as bragging rights and political struggle.

Persistency (PS)

The persistency describes the intentions of the attacker, which result in one-off or persistent deception. The **one-off (PS1)** attack, just as its name implies, launches one time and will not continually execute. The spamming is typical of one-off SE attack. The adversary sets spoofed webpage to gather information, which once succeeded, redirects the user to legitimate websites and disappear. The **persistent (PS2)** attack refers to long-term, mostly Internet-based attacks for financial gain. This attack causes advanced persistent threat which will not expire upon one time successful exploitation but will periodically conduct in order to gain further profit.

3.1.2. Exploitation

The exploitation phase presents the factors involved in bypassing the secure perimeter, which includes the techniques, the psychological triggers, the medium and the communication.

Techniques (TC)

The techniques represent the approaches used to exploit the vulnerabilities in social engineering attacks. It includes four subclasses: physical, technical, social and hybrid.

1) Physical (TC-PH)

The physical approaches refer to those where the adversary performs some forms of physical activities for gathering information (see Figure 2). The **dumpster diving (TC-PH1)** represents the action of digging through trash at corporations in search sensitive data. The **shoulder surfing (TC-PH2)** indicates the observation techniques, such as looking over someone's shoulder, in order to get security information.

Figure 2. Physical approaches: dumpster diving and shoulder surfing

2) Technical (TC-TC)

The technical approaches refer to the technical actions mainly carried out over the Internet to gather sensitive information.

The **phishing (TC-TC1)** is the attempt to acquire sensitive information, such as username, passwords, credit card details, etc., or to make someone to act in a desired way by masquerading as a trustworthy entity in an electronic communication. The general phishing will attack a group of targets of opportunity. However, the **spear-phishing (TC-TC1S)** focus on attacking some specific individuals or cooperators, thus it requires the adversary to gather information on the intended target in advance. So, the spear-phishing needs more effort but also has a higher success rate than the general phishing attacks.

The **waterholing (TC-TC2)** refers to the adversary compromise websites which are often browsed or are likely to be of interest to the targets of choice, and infect them with malware, and then waits the target victims getting infected at the waterhole.

The **baiting (TC-TC3)** is like the real-world Trojan Horse that exploits the victims' greedy and curiosity to access the malware infected temptation, which could be physical media or software and online item. The baiting attack is very similar to the phishing attack, while the baiting is more like a gift or a good left in somewhere can be found by the victims.

3) Social (TC-SC)

The social approaches rely on socio-psychological triggers to manipulate the victims in order to get sensitive information.

The **persuasion (TC-SC1)** is aimed to get a victim to comply with an inappropriate request to make them perform some illicit action based on some psychological triggers, such as purported authority. One representative persuasion is **diversion theft (TC-SC1D)**, which is also known as the “Corner Game” or “Round the Corner Game”. It is a “Con” exercised by professional thieves, normally against a transport or courier company. The objective of diversion theft is to persuade the persons responsible for a legitimate delivery that the consignment is requested elsewhere – hence, “round the corner”.

The **pretexting (TC-SC2)** refers to the art of creating and using a fabricated scenario (the pretext) that can be used to increase the chance the victim divulge sensitive information or perform actions that would be unlikely in ordinary circumstances. In comparison with the persuasion, pertexting stands for deceiving the dupe though using some of the techniques, such as impersonation, name-dropping and using false ID, etc.

The **quid pro quo (TC-SC3),** meaning “something for something” or “this for that” in Latin, refers to the social engineering attacks promise a benefit in exchange for information. This benefit usually assumes the form of a service, whereas baiting frequently takes the form of a good.

The **reverse social engineering (TC-SC4)** is a type of attacks calling back when the victim needs help from someone who claimed can solve the problem, which relies on the established trust between the attacker and the victim, so that the attacker will be allowed to gain the privileged information.

4) Hybrid (TC-HB)

This hybrid approaches refer to the exploiting techniques consisting of multiple different single approaches described above. One typical hybrid approach is the tailgating.

The **tailgating (TC-HB1)**, also known as “piggybacking”, refers to that type of attacks involve the adversary who lacks of the proper authentication seek entry to a restricted area through following a person who has legitimate access (see Figure 3).

Figure 3. One type of hybrid approach: tailgating

These restricted areas, i.e. organizations and corporations, are often secured by unattended and electronic access control such as RFID based entrance guard card. For instance, the adversary impersonates a delivery driver and waits outside a building. When an employee appears to gain security's approval and open the door, the adversary will hold the door open, or the employee may hold the door open for the trailing adversary following common courtesy, or the attacker may even ask the employee to hold the door open while the legitimate employee may fail to ask for identification for any of several reasons, such as accept a fabricated assertion that the attacker has forgotten or lost the appropriate identity token.

The **vishing (TC-HB2)**, known as phone phishing, is the act of using the telephone in an attempt to scam the dupe into surrendering private information that will be used for identity theft. The scammer usually pretends to be a legitimate business, and fools the victim into thinking he or she will profit.

Psychological Triggers (PT)

As we all know, social engineering is a social exercise, the attackers usually exploit the victims' psychological triggers, or psychological vulnerabilities, to get the desired sensitive information. Hence, it is necessary to make sense of the psychological triggers that take effect during a social engineering attack.

Strong affect (PT1) : it is a trigger using a heightened emotion as a powerful distraction, such feeling a strong sense of surprise, anticipation or even anger, to interfere with the victim’s ability to evaluate and think logically when arguments are being presented. This psychological trigger could be fear, excitement, panic, curious and greedy, etc., which are always exploited by the persuasion technique.

Overloading (PT2) : it refers to the victim has too much information to process but does not have enough time to evaluate it. Hence, this is a trigger reducing the victim’s ability to process and scrutinize the argument so that the target is more willing to accept arguments that should have been challenged

Reciprocation (PT3) : this trigger relies on the social interaction rule: if someone give us something or promise us something, we should return the favor. The reasoning follows that people are more willing to comply with a request if the requester has treated them favorably in the past. So, this trigger can be exploited by some techniques, such as quid pro quo and reverse social engineering.

Deceptive Relationships (PT4) : this trigger indicates that the attacker builds the fabricated relationship with the target in order to increase the chance the dupe divulge private information to the attacker. The reason is people are more willing to comply with requests from friends or people they like and perform activity under a legitimated and trustworthy relationship. So, one way of doing this is sharing information through discussing a common enemy. Another example is the attacker appears to the target as if they are very much alike, e.g., have the same interests or desire the same things out of life. This trigger can be exploited by the pretexting techniques.

Diffusion of Responsibility and Moral Duty (PT5) : this trigger means victims are more willing to accept requests or perform actions when they feel it is none of their business or they will not be held solely responsible for their actions. Hence, this psychological trigger can also be exploited by the pretexting techniques.

Authority (PT6) : it indicates that people are easily to response the requests given by the people with more authority than they have. This trigger can be exploited by the persuasion and pretexting techniques.

Integrity and consistency (PT7) : this trigger refers to people have a tendency to follow commitments and comply the request consistent with them, even though commitments may not very wise at the first place.

Social Validation (PT8): this trigger means victims are more easily to comply to requests if they are seen as the socially correct thing to do. The tailgating techniques can exploits this trigger through using victim's common courtesy to open the door for the adversary.

Scarcity (PT9): this trigger presents people are more likely to comply with a request that is scarce or decreasing in availability. The reason hiding behind is people subconsciously approve the fact that objects are valued because of their rarity.

Medium (MD)

The medium refers to the channel through which the social engineering attacks perform the exploitation.

Pamphlets (MD8): the social engineers often use pamphlets to perform social techniques such as the reverse social engineering to the targets of opportunity.

Storage media (MD6): it is a physical medium that can be used to perform baiting attacks through exploiting the victims' greedy and curiosity.

Paper mail (MD2): the social engineers sending paper mail to victims can perform social exploiting techniques such as reverse social engineering attacks. Furthermore, the victims' paper mails including sensitive personal information are also the resource to the dumpster diving attack.

Email (MD1): it is the most common medium used by the technical and social exploiting techniques, such as phishing and reverse social engineering attacks.

Face-to-face (MD3): the face-to-face conversation is the direct way aimed to get sensitive information from the victims through exploiting psychological triggers.

Telephone/VoIP (MD4): it is commonly used by the social exploiting techniques in order to make the victims leak sensitive information through exploiting psychological triggers as well.

Instant messaging (MD5): it is a type of online chat that offers real-time text transmission over the Internet, which is increasingly popular among social engineers to perform phishing and reverse engineering and even identity theft by exploiting trustworthy relationship.

Website/Software (MD7): it is mainly used to facilitate the waterholing attacks. In addition, it can also fabricate malicious webpage as the victims' intended webpage to perform phishing attacks.

Social network (MD9): social engineers can use it to fabricate fake identities and complex lure information in order to build trustworthy relationship with the victims.

Cloud (MD10): the social engineers can use the cloud service in collaboration scenario to make the victim hand sensitive information over in a share directory.

Communication (CM)

This class presents the communication strategies to facilitate the social engineering techniques. The communication strategy consists of two subclasses: direct communication and indirect communication, and furthermore, the direct communication can be divided into two subclasses: bidirectional communication and unidirectional communication. The **bidirectional communication (CM1)** indicates the attacker and the victim both participate in the conversation, such as the face-to-face, telephone/VoIP, instant messaging, and even the email send by the attacker to the target and the victim replies to the attacker. The **unidirectional communication (CM2)** means the conversation is one-way only from the attacker to the victim, such as the phishing attacks. Another case in point is that the attacker sends a message through paper mail without a return address so that the victim cannot reply to the attacker. The **indirect communication (CM3)** occurs when there is no actual interaction between the attacker and the victim but the communication occurs by some third-party medium. A representative social engineering attack belong to this category is the baiting, which could be a malware-infected storage medium found by the victim and thereafter the victim infects the malware through plugging the storage medium into personal computer due to some curiosity and greedy.

3.1.3. Compromise

The compromise phase describes the way to gain access, maintain access and hide the illicit access. So, this phase consists of interface manipulation, execution steps, maintaining access and evidence dash.

Interface manipulation (IM)

This class refers to the interface, through which the victim target can be compromised by the attack vector. The **user interface (IM1)** manipulation is aimed to abuse the functionality provided by the user interface of the target system, which can be hardware and software user interfaces. For example, the result of tailgating is gaining access of the target corporate settings through the hardware user interface, i.e. the entrance of the building. The **programmatic interface (IM2)** manipulation refers to exploit the program flaw or the vulnerability of a target system through program modification.

Execution steps (ES)

This class indicates one social engineering attack could consist of one or multiple steps. The **single step (ES1)** means the attack vector only needs to carry out one individual action to achieve the result. The **multiple-step (ES2)** attack must consist of more than one steps to facilitate the execution in order to finish the compromise.

Maintaining access (MA)

This class presents the approaches that the adversaries used to maintain access to the compromised system. To **set backdoor (MA1)** means to leave an easier way back into the compromised system later. By using this method, even the vulnerability is patched later, the adversary can still gain access to the victim system for future use. To **add a new account (MA2)** refers to create a legitimate account against to the access control in order to freely pass in and out the victim system in the future, e.g. to fabricate an entrance guard card or key so that the adversary can enter the victim organization or house when there is no one else in there.

Evidence dash (ED)

This class indication the adversary covers the tracks and hides the intrusion after achieving the objectives. The **recruit (ED1)** refers to use the trustworthy to recruit the victim to work for the attacker or even as an ambassador to find new victims. The **cloak (ED2)** is the actions performed after compromising a victim in order to camouflage the illegal activities.

3.2. Applying the taxonomy to real social engineering attacks

In this subsection first a case study is presented to explain how to use the proposed taxonomy to classify and analyze social engineering attack. Thereafter, more real social engineering attacks will be classified in order to unveil the attacking space by using the novel taxonomy.

3.2.1. Case study

The Amazon's customer service backdoor [11] is taken into account as an instant to apply the taxonomy. The case described a story where the social engineering attackers impersonating the victim try to exploit the psychology trigger of the Amazon's customer service in order to get private information of the victim's credit card for financial gain. So, it can use the novel taxonomy to analyze this SE attack by classifying each attacking step.

In the orchestration phase, each action's classification should be TG1, AC1, GL1, and PS2. The reasons can be shown as follows:

Target: The attacks focus on gaining the information of the specific victim's credit card, so the target was classified as *explicit*.

Actor: The attacks were launched by online-chatting and telephone conversation to contact with the custom service, it therefore was classified *human* actor.

Goal: The attacker is aimed to use the victim's credit card to buy something, so it is obvious the goal is *financial gain*.

Persistency: The story shows the social engineering attacks were continually happened, including three times within several months. So, it was classified as *persistent* attack.

In the exploitation phase, the classifications are: TC-SC2, PT4, MD4, MD5, and CM1. The classifications can be presented in detail as follows:

Techniques: the attacker impersonated the victim in order to deceive the custom service, so it was classified as *pretexting*.

Psychological trigger: Even though the attacker tries many times to exploit the custom service through using the impersonation, the service stuffs were not compromised by the adversary. So, it was classified as deceptive relationships.

Medium: The story shows the conversations between the attacker and the custom service stuff were established through online chatting application and Telephone service. Therefore, the mediums were *instant messaging* and *telephone/VoIP*.

Communication: In this case, both the attacker and the custom service stuff participated in the conversation, so the communication is *bidirectional*.

In the compromise phase, these classifications can be obtained: IM1 and ES1. The classifications can be presented in detail as follows:

Interface manipulation: the attacker just use the normal system interface of the bank that the victim's credit card belongs to. So, it was classified as *user interface*.

Execution steps: after the attacker get the information of the credit card, he only needs on step to compromise the victim's account in the back e-system. Therefore, it was classified as *single-step*.

Maintaining access: in this case, the story did not show the attacker set any backdoor to maintain the access.

Evidence dash: the story also did not mention any activity the attacker performed in order to hide his track.

So, it has presented how to apply the proposed taxonomy to classify and analyze a real social engineering attack. The taxonomy can fully classify a SE attack. However, note that it is possible that some SE attacks lack of the values in some classes. For example, the case mentioned above does not have value in maintaining access and evidence dash. It is normal because some SE attacks are not completely performed or were not fully described so that some relative classes cannot be used to classify them. In contract, some class may have multiple values, such as the medium using in this case. Because a SE attack could consists of using different mediums to perform their exploiting. In the next subsection, several other SE attacks are selected as the supplement to the case mentioned above to show richer social engineering classifications.

3.2.2. Classification and analysis

In this subsection, the proposed taxonomy is applied to classify several representative social engineering attacks provided by the paper (Heartfield) [22] as well as the case described in the last subsection, which were presented in Table 1. Note that some attacks in paper (Heartfield) [22] are not typically social engineering attack. For example, some man-in-the-middle attack, such as WiFi Evil Twin phishing and HTTPS man-in-the-middle adware, which are general cyber attacks based on network programming techniques.

The classifications of these representative instances show that nowadays most of SE attacks are automated and aimed to the targets of opportunity. For example, the malware and malicious website based phishing attacks. The reason is the fact that the new technology, i.e. IT technology, based SE attacks and can hide the attacker's real identity. Hence, the technical exploitation approaches are much more popular than others among present SE attacks.

Though the attacking techniques became more automated, the main human being's psychological trigger used for exploiting by the SE attacks is still the deceptive relationships. Many traditional SE attacks launched by social engineers often rely on the pretexting in order to fabricate the deceptive relationships. However, the present programmatic automated SE attacks can use the trustworthy that is established by the application entities, such as webpage, software, digital file, etc., which the victims widely accept, to embed the malicious code to directly infect the victims. The main attacking goals are financial gain and unauthorized access.

Table 1 Applying the taxonomy to multiple SE attacks

SE attacks	**Orchestration**	**Exploitation**	**Compromise**
Amazon's custom service backdoor [11]	TG1, AC1, GL1, PS2	TC-SC2, PT4, MD4, MD5, CM1	IM1, ES1
Drive-by download attack (Cova et al.) [45]	TG1/TG2, AC2, GL2, PS1	TC-TC2, PT4, MD7, CM3	IM2, ES1
Malver tisements in social media (Li et al.) [46]	TG2, AC2, GL2, PS2	TC-TC1, PT4, MD9, CM3	IM1, ES1, ED2
Fake mobile applications (Felt and Wagner) [47]	TG2, AC2, GL2, PS2	TC-TC1, PT4, MD7, CM3	IM2, ES2
Instant Message Phishing—Automated (Mannan and Oorschot) [48]	TG2, AC2, GL2, PS2	TC-TC1, PT4, MD5, CM2	IM1, ES1
Multimedia Masquerading (Ford et al.) [49]	TG2, AC2, GL2, PS1	TC-TC1, PT4, MD9, CM3	IM2, ES1, ED1
NFC Phishing (Madlmayr et al) [50]	TG2, AC2, GL2, PS1	TC-TC1, PT6, CM3	IM2, ES2

SE attacks	**Orchestration**	**Exploitation**	**Compromise**
Peripheral Masquerading by USB (Jacobs) [51]	TG1/TG2, AC2, GL2, PS2	TC-TC3, PT1, MD6, CM3	IM2, ES2, MA1, ED2
Ransomware (Gazet) [52]	TG2, AC2, GL1, PS2	TC-TC3, PT6, MD7, CM3	IM2, ES2, MA1, ED1
SMS Worm --Selfmite (Ducklin) [53]	TG2, AC2, GL2, PS1	TC-TC1, PT1, MD5, CM2	IM2, ES2, ED1

Furthermore, with the development of IT technology, some new attacking mediums have been appeared, such as the NFC tag in the NFC phishing attack. The newer technology the attacker uses can lead to more difficult to defend the SE attack.

The table also presents the representative malware-based SE attacks can automatically perform the camouflage activities after compromising the victims, such as setting backdoors, hiding the tracks, and even spreading the malware to other victims.

In addition, theoretically, according to the class combination of this classification scheme, there are 1140480 types of SE attacks in the attacking space. However, in practice, the number of types of SE attacks is much less. Because some class values have specific relationship to other class values. For example, the medium “face to face” indicates that the communication way must be “bidirectional”.

3.3. Conclusion

In this chapter, a novel taxonomy of social engineering attacks is proposed. For the research questions, the classification scheme is defined, each terminology is described, and the three attacking phases are used to organize the classes into a formal conceptual model. The taxonomy is applied into a real SE attack in detail in order to present how to use the classification scheme to analyze the attack. Then a number of typical SE attacks is classified by applying the taxonomy. The classification result shows that this taxonomy can fully classify the real SE attacks, where the security researchers can gain an insight of the study field. Furthermore, the new taxonomy as a conceptual model can be extended through adding new classes or new values in the existing classes. For instance, in the class medium, the NFC tag can be added as a new value. It is hoped that this work can be used for further study of SE attacks and even inspire the idea to defend the SE attacks.

The problem of this chapter is the factor that the taxonomy has not been properly formalized. The taxonomy can be formalized in an ontology (supportive OWL[3] or NeOn[4]) or in concept maps[5]. Thereafter, it is needed more time budget to deduce a set of pattern specifications. A properly formalized taxonomy can deduce a set of pattern specification, which can be used to better diagnose with more discernment social engineering attacks. In the next step, a translation of the findings into an attack-process tree based on the process-tree formalizations should be worked out as well.

[3] https://www.w3.org/2001/sw/wiki/OWL

[4] http://neon-toolkit.org/wiki/Main_Page.html

[5] http://cmap.ihmc.us/docs/conceptmap.php

4. Social engineering defense model

As stated, the taxonomy is used to gain insight of social engineering attacks. However, the goal of this thesis is to propose defense solutions to protect information system resource from being attacked by social engineering. In this chapter, therefore, according to the knowledge of the proposed taxonomy, a multi-layer social engineering defense model is designed, which includes three security levels: prevent, detect and control.

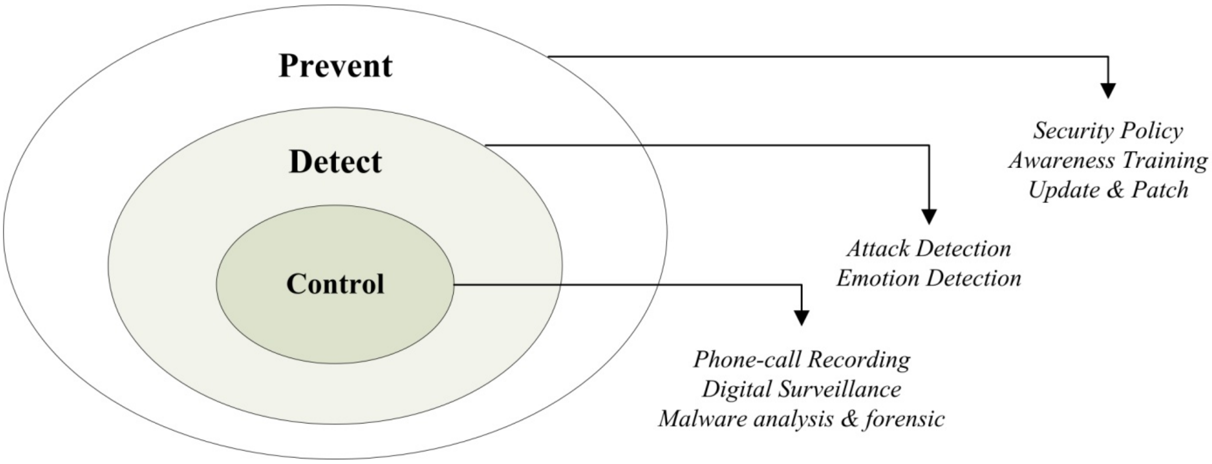


Figure 4. Social engineering defense model

In each layer, several strategies and mechanisms are proposed to defend the social engineering attacks and protect the potential victims. Each layer will be described in detail in the next subsections.

4.1. Prevent

The first defense layer is aimed to remove the vulnerabilities, which can be exploited by the social engineering attacks.

4.1.1. Security policy

Well defined and documented security policy is the foundation for defending SE attacks. Organizations often use information security management system (ISMS) to provide a framework for information security risk management. ISMS consists of sets of security policies to define, construct, develop and maintain the computer system (including hardware and software resources) based security within companies. There were several security standards for IT Governance which leads to information security, and the big five of ISMS standards are ISO27001, BS7799, PCIDSS, ITIL and COBIT. These policies dictate the way where computer resources can be used.

However, most security standards and policies are defined to address general information security risks, including viruses, worms, hackers, phishers, and social engineers, threaten organizations. Hence, these general security policies are ineffective owing to a failure to acknowledge all that is actually required to cope with social engineering attacks. For defending social engineering attacks, the set of policies provided by the security standards should cover not only the computer-based risks but also the human-based risks.

Therefore, as stated in chapter 2, organizations are recommended to apply (if the budget is available) the ISO/IEC 27032 which is a new international standard published by ISO that covers the baseline security practices for all stakeholders in cyberspace, but particularly, provides technical guidance for addressing social engineering attacks. However, this novel security standard still needs to be validated how it will turn out in practice and how widely it will be accepted. But it will function much better as a supporting standard for ISO27001 implementation than as an independent framework.

In addition, an entire life cycle of security standards includes research, getting policies down in writing, getting management buy-in, getting them approved, getting them disseminated across the enterprise, keeping users aware of them, getting them enforced, tracking them and ensuring that they are kept current, getting rid of old policies, and other similar tasks. Unless an organization recognizes the various functions involved in the policy development task, it runs the risk of developing policies that are poorly thought out, incomplete, redundant, not fully supported by users or management, superfluous, or irrelevant.

4.1.2. Awareness training

Once the foundation of a security policy has been established and approved, all employees should be trained in security awareness. Thus, though the organizations apply appropriate security standards, they still need to train the employees' awareness to defend the SE attacks. Indeed, an effective security policy life cycle should include the functions and corresponding responsibilities, such as security awareness tasks and even policy compliance oversight. Therefore, the security policy development for SE attacks beyond simple policy writing and implementation. It also requires much more activities than staffing a newly created policy, e.g. making employees aware of it, and ensuring that they comply with its provisions.

This task can be done by defining the awareness needs of various audience groups within the organization (executives, line managers, users, etc.); determining the most effective awareness methods for each audience group (i.e., briefings, messages, courses); and developing and disseminating awareness materials (presentations, posters, mailings, etc.) regarding the need for adherence to the policy. The awareness function also includes efforts to integrate up-to-date policy compliance and enforcement feedback as well as current threat information to make awareness information as topical and realistic as possible. The final task is measuring the awareness of employees with the policy and adjusting awareness efforts based on the results of measurement activities (that will be described in Chapter 5).

However, the conventional human being involved awareness training methods, such as educational courses, routine remind, interview, awareness quiz and survey, etc., are labor intensive, repeated and even perhaps tedious. At present, there are some automated tools can be used to train and promote user awareness by simulating real world social engineering attacks. For example, the King Phisher[6] is an open-source tool for automatically training the users' awareness to prevent phishing attacks (see Figure 5). It can be used to run campaigns ranging from simple awareness training to more complicated scenarios in which user aware content is served for harvesting credentials.

Nevertheless, the previous mentioned methods (even though there are several automated tools can be applied to address the repeated tasks) are passive solutions that enforce the employees to be aware of the sensitive information protected by the security policy. Indeed, awareness training does not simply require the employees to keep secret of the sensitive information, but desire them to know how to identify confidential information and understand their responsibility to protect it. Thus, a positive method is to combine the employee's profit, which could be the bonus, reward or merit pay, with the sensitive information security. Thereafter, all employees will actively improve their awareness because the information security has associated to their own financial benefit.

[6] https://n0where.net/phishing-campaign-toolkit-king-phisher/

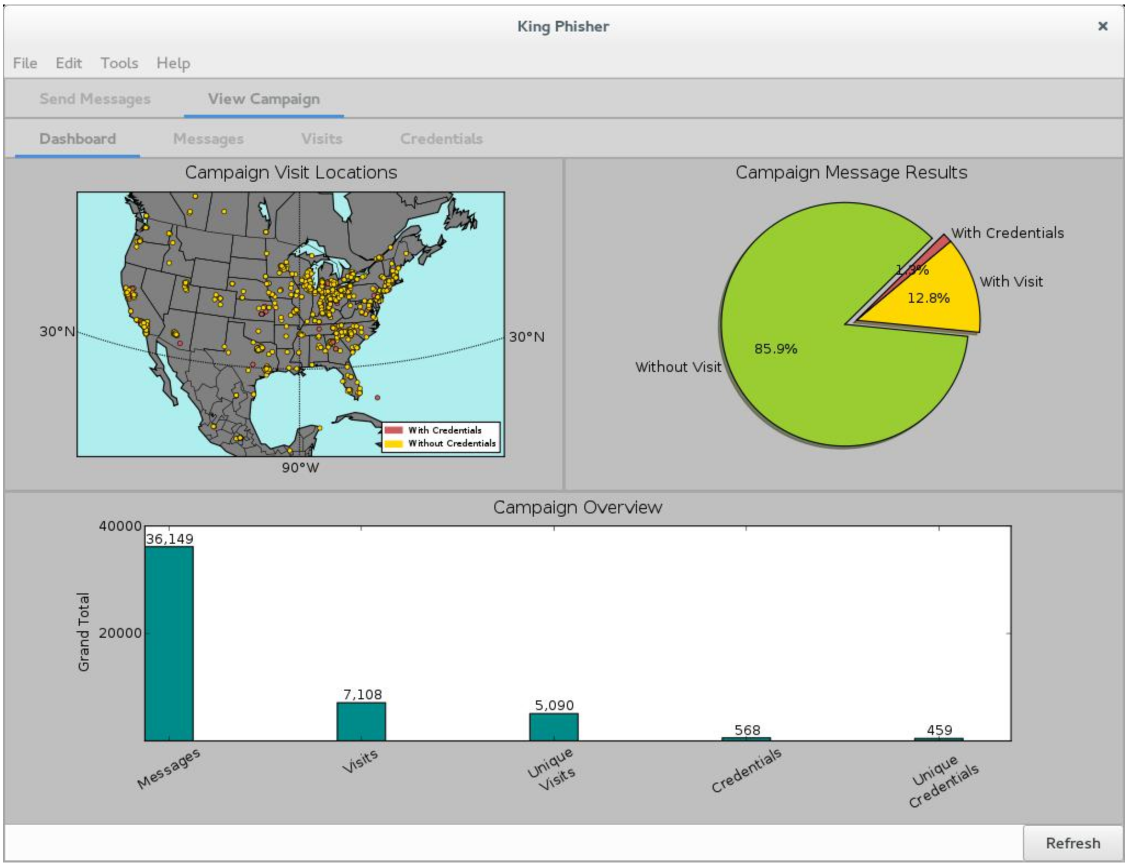


Figure 5. King Phisher dashboard

Furthermore, some resistance based awareness training should be taken into account for the key personnel, such as secretaries, receptionists, system administrators, business assistants, customer service staffs, etc., whose job is to help others especially the general public and those whose job includes escalated rights.

4.1.3. Update & patch

This mechanism is aimed to timely repair the physical, technical and even psychological vulnerabilities.

If the corporation has a good financial position, it is suggested to update the office facilities (see Figure 6). For example, in order to prevent the dumpster diving, the organization should equip the paper shredder to avoid the sensitive information leaving over in the trash. Furthermore, using the fingerprinting based authentication approach replace the password typing based access control to avoid shoulder surfing.

Figure 6. Equipment update for preventing physical SE approaches: paper shredder and fingerprint identification

Another case in point is to hire the security guards at the entrance of building and some restricted areas as the supplement of the electronic access control. All persons entering the building are required to swipe the ID card. The one no ID entry has to register his information and pass the security check by the security guards.

Furthermore, the security policy should update timely in case the organization's equipments are upgraded or the security policy is no longer effective for the current security purpose. It is aimed to address the process of ensuring the currency and integrity of the policy. This includes tracking drivers for changes (i.e., changes in technology, processes, people, organization, business focus, etc.) that may affect the policy; recommending and coordinating policy modifications resulting from these changes; and documenting policy changes and recording change activities. This function also ensures the continued availability of the policy to all parties affected by it, as well as maintaining the integrity of the policy through effective version control.

4.2. Detect

Though the organizations can apply security policy to form the first line of defense, it is not enough to defend against SE attacks. There are still some security concerns that have to be taken into account. First, the organization cannot completely ensure that all the employees are objectively aware of the confidential information. Second, even if the employees have well trained and more aware of sensitive information's, there is still a risk of leaking sensitive information subjectively. Though the information security can associate to the employee's financial benefit to reduce this risk, there still has the hidden danger when the employee is extorted or even bribed or by a higher financial benefit from the attacker. Third, the outdated policy and the violation of the security policy are hidden risks as well. Thus, if it is assumed that the social engineering attack has broken the first defense layer, the second defense layer needs to detect these attacks and make alerts in time.

As stated, social engineering is the exploitation of the natural human tendency to trust. Hence, the SE attack detection is aimed to detect the attack that exploits human vulnerabilities. In this section, two approaches are described to detect the SE attack: attack detection and emotion detection. Attack detection is used to detect the attack pattern spread through the digital media, such as email, instant messaging, website, etc., which can be analyzed by security program. However, there are some conventional human-based SE attacks just using sophisticated conversation by face to face and telephone/VoIP, which is difficult to be analyzed by automated program. Hence, another approach is proposed, emotion detection, which is used to detect the employee's emotion state in order to determine when to make an alert. These two detection approaches will be described in the next two subsections.

4.2.1. Attack detection

As stated, there are two effective ways to exploit psychological weakness of the target: phishing and dialog-based attack. The phishing is often based on the unidirectional communication media and dialog-based attacks are often based on the bidirectional communication media. Hence, both of them have the probability to be detected by automated program.

In this subsection, a novel authentication and topic blacklist based SE detection model is proposed (see Figure 7). In this SE detection model, there are two authentications: relationship authentication and resource access authentication. The relationship authentication is used to check whether the request sender is trusted to the receiver. The resource access authentication is aimed to determine if the request sender has the privilege to access the intended resource.

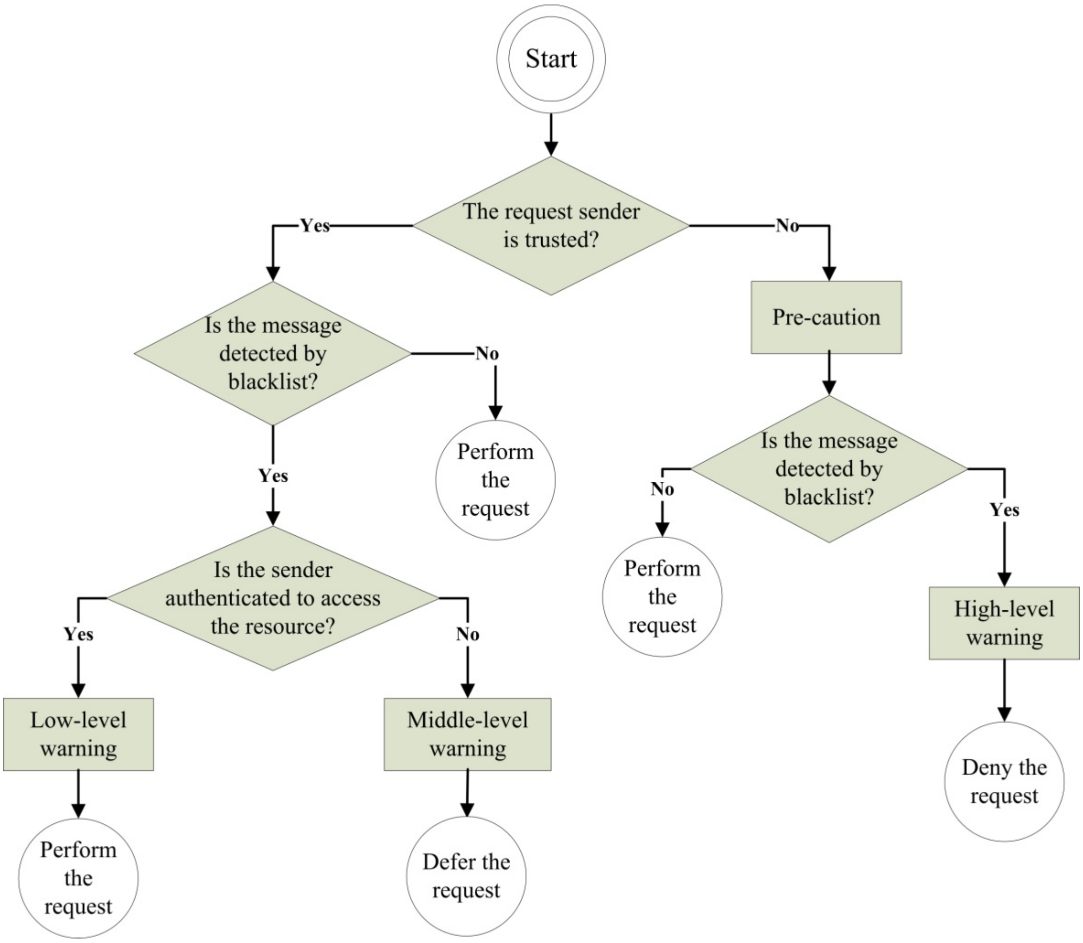


Figure 7. An overview of SE detection model

On the other hand, the sensitivity of the resource itself is another important concern that the model has to take into account. Hence, the topic blacklist (TBL) is proposed to check if the sender requests sensitive information or not. The TBL is a list of statement topics, which describe a sensitive operation associating to a sensitive data. So, each topic consists of two elements: an action and a resource. The action describes an operation that the request receiver may perform, while the resource is the sensitive information system resource to which access is restricted or at least should be authenticated. The TBL can be generated according to the existing security policy document associated with the information system, or based on common security requirements. The form of each topic is a two-tuple, such as {“send”, “money”}, {“tell”, “account number”} and {“input”, “password”}, etc.

Furthermore, four warning levels are proposed: pre-caution, low-level warning, middle-level warning and high-level warning. Besides, three corresponding responses are suggested as well: perform the request, defer the request and deny the request.

So, it is assumed that the SE attacks are launched by email, instant message, SMS text, social media chat, etc. The requested message is already provided in the text form. This SE detection model will first check whether the sender is trusted. If the sender is not trusted with the receiver, the model will make a pre-caution to the receiver. Then, if the request message does not hit the topic blacklist, the receiver will perform the request. Otherwise, if the request message hits the blacklist, the model will make a high-level warning and deny the request.

On the other side, the requested message sender is trusted. So, when the requested message does not include any topic inside the blacklist, the receiver performs the request. Then, if the requested message hit the TBL, the model will first check whether the sender is authenticated to access the resource. If the sender is authenticated, the model will make a low-level warning, and performs the request. The model makes the low-level warning is aimed to remind the receiver that the sender could be an SE attacker who has got all authentication information from a victim then bypassed these two authentications. Because the technique to completely determine whether the sender is a true legal user or a SE disguiser is beyond the scope of this thesis, so it can simply consider the sender passing the two authentications is legal user but the model generates a low-level warning as a reminder. However, if the request sender is not authenticated to access his intended resource, the model will make a middle-level warning and then defer the request, which means the sender needs to perform further authentication and his request will put into a waiting list. This approach can deter the SE attack but still keep the possibility to answer the legal user when he passes the authentication.

4.2.2. Emotion detection

As mentioned before, some SE attack is hard to detect by the proposed detection model. Hence, the employee's emotion detection is proposed as a supplement. To our best knowledge, change in emotional state will have an influence over the individual's cognitive functioning. In other words, the employee's emotional state can affect his awareness of the sensitive information. However, it is not an easy task to determine one's emotional state, and it is even an impossible task for an individual to adjust his or her own emotional state, because individuals have their own perception of emotional state and even some individuals are unable to perform this kind of task in a rational way when their emotions are irrationally challenged. Hence, it is desired to propose an emotion detection model for automatically performing this task.

Before the emotion detection model is proposed, some basic concepts of emotional state should be comprehended. First, one's emotional state is something that can stay constant for a long time unless the individual experience great discomforting incident, such as economic crisis, health issue, loved ones die, etc., which have intense effect on the cognitive function. Second, however, an individual's emotional state can be impacted in a short time when the individual is under attack by the SE attacker. The experiencing severe stress will have an influence over the individual's cognitive function (Mathews) [54]. Thereafter, the emotion detection model can be proposed as Figure 8 shows.

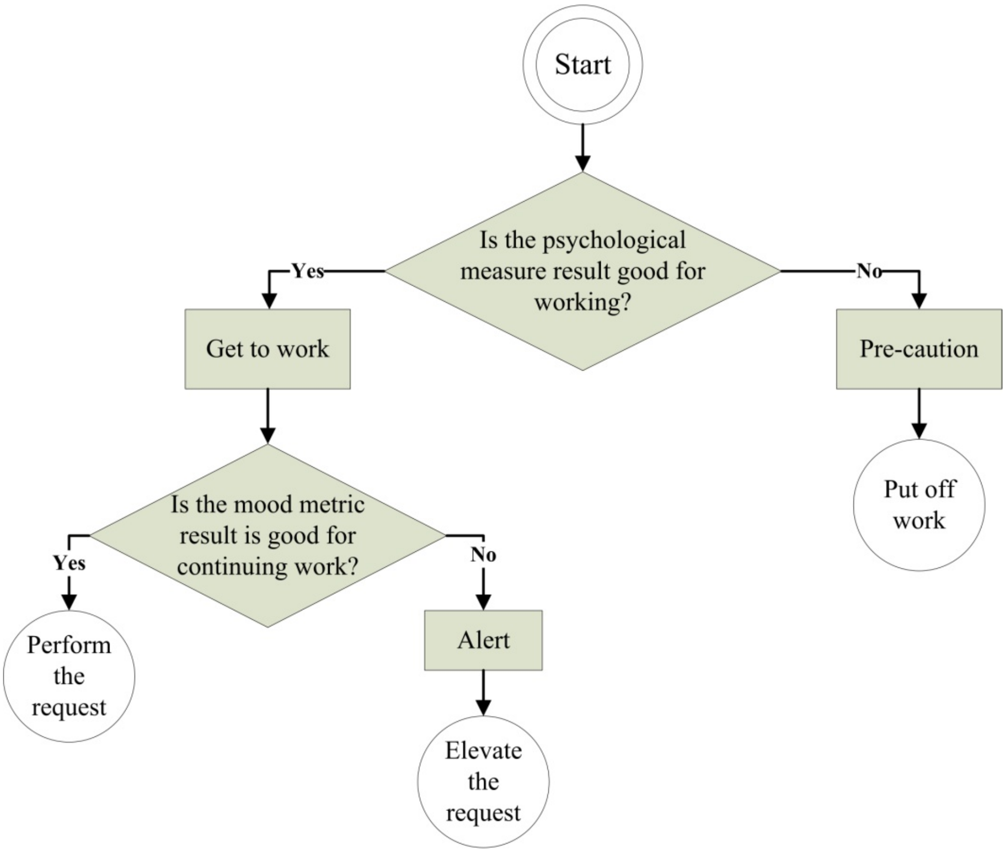
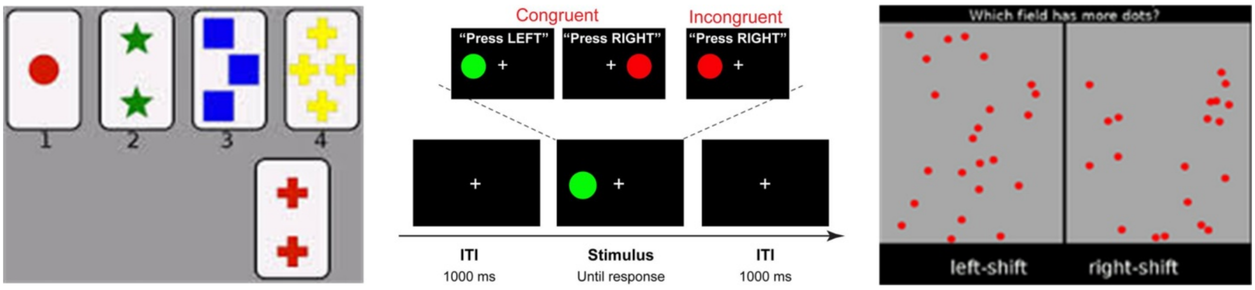

Start
Is the psychological measure result good for working?
Yes
No
Get to work
Pre-caution
Put off work
Is the mood metric result is good for continuing work?
Yes
No
Perform the request
Alert
Elevate the request

Figure 8 An overview of the emotion detection model

The model itself is not complex, which has two types of warning: pre-caution and alert, as well as three suggestive responses. However, the emotional state measurement approaches are the most essential components. Two types of emotional state measurement approaches are proposed: long-term psychological measure and instant mood metric.

The long-term psychological measure should not be lengthy but should be effective to determine the level of cognitive function of an individual before the individual start working. There are several well developed psychological measures that can be found in the

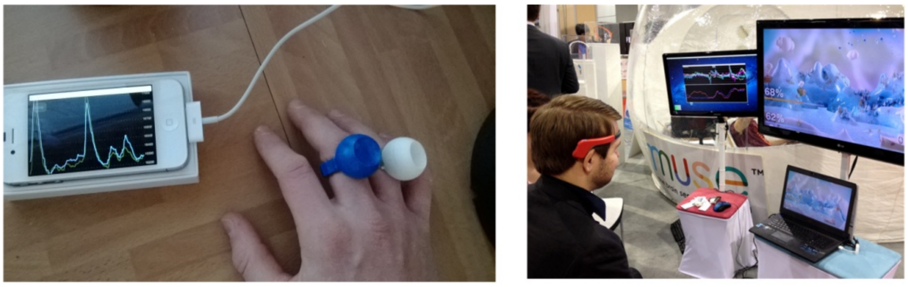
Psychological Experiment Building Language (PEBL), which is an open source project that allows easy creation of computer based psychological measures[7]. In this thesis three representative psychological measures are applied: Wisconsin Card Sorting Test (Monchi) [55], Eriksen's Flanker Test (Eriksen) [56], and Dot Judgment Task (Cicchetti) [57] (see Figure 9).



1. Wisconsin Card Sorting Test **2. Eriksen's Flanker Test** **3. Dot Judgment Task**

Figure 9 Long-term psychological measures

On the other hand, the instant mood metric is aimed to monitor the individual's emotional state during working period through measuring some physical signs such as pulse wave and brain wave. At present, there are a number of wearable instruments that can monitor the individual's physical signs in order to determine the mood situation and emotional state, such as Moodmetric ring[8] and Muse brain sensing headband[9] (see Figure 10).



1. Mood metric ring **2. Brain sensing headband**

Figure 10 Wearable devices for instant mood metric

All the emotional state measurement approaches are summarized in Table 2.

Table 2 The representative emotional state measurement approaches

Measurement approach	Usage in the model	Expense

[7] http://pebl.sourceforge.net/
[8] http://urbanwearables.technology/moodmetric-smart-jewelry-emotional-guidance/
[9] http://www.choosemuse.com/what-does-muse-measure/

Measurement approach	Usage in the model	Expense
Wisconsin Card Sorting Test	Long-term psychological measure	Free
Eriksen’s Flanker Test	Long-term psychological measure	Free
Dot Judgment Task	Long-term psychological measure	Free
Mood metric ring	Instant mood metric	Budget
Brain sensing headband	Instant mood metric	Budget

So, it is assumed that the employee works as a call centre agent. Before start working, the employee has to take a long-term psychological measure to determine if the emotional state is suitable for working. If the result indicates that the individual’s emotional state is not good for providing service, the model will generate a pre-caution and the employee should put off work. In contrast, the employee can wear the mood monitoring device and get to work. When the employee receives a request and the mood metric result is good enough at that time, the employee can perform the request. Otherwise, the emotion detection model will make an alert and the request will be elevated to other employee who has stable mood at that time to deal with.

4.3. Control

The third defense layer is aimed to record the malicious behavior and perform the data capture for further investigation, so that even the social engineer has compromised the victim system, his activity is still under control by the security mechanism.

4.3.1. Phone-call recording

This mechanism should be used to capture the social technique based activity happened in the call center environment, such as the e-bank system where the call center agents will directly communicate with the social engineer. Indeed, the phone-call recording mechanism has been applied widely in many bank call centers. For example, the user will be asked at the beginning of the conversation with the call center agent that this talk will be recorded. Therefore, it can be imagined that even the social engineer can disguise perfectly bypass the detection layer, his malicious behavior has been captured. These data can be used to track the social engineer and even applied as the evidence of crime.

4.3.2. Digital surveillance

The digital surveillance is aimed to capture the physical malicious behavior, such as dumpster diving, shoulder surfing, and even tailgating. The surveillance camera is the widely used device to facilitate the monitoring task. It can be seen that many public places, e.g. super markets, hospitals, banks, etc., using digital surveillance. Definitely, many enterprises also apply the surveillance camera as a security approach. Note that the camera should be equipped not only at the entrance but also should be fixed inside of the enterprises in order to monitor the potential insider social engineering behavior. The captured data by the camera can be used to track the social engineer and even applied as the evidence of crime as well.

4.3.3. Malware analysis and forensic

This mechanism is designed to capture the malicious activity launched by the technical SE attacks, i.e. computer-based automated malware.

It is recommend using honeypot system (Spitzner) [58] to facilitate the data capture task. Honeypot is a information system used to be probed, attacked and compromised in order to capture the malicious behavior for further investigation. Figure 11 shows a representative generation II honeynet system, which is a network consisting of multiple honeypots following certain network topology.

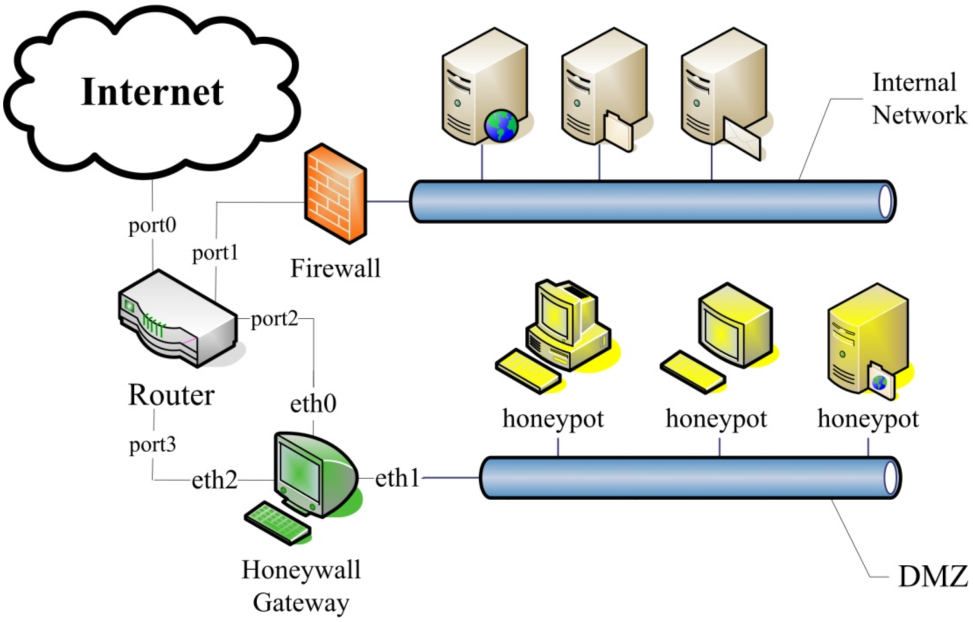


Figure 11. Gen II honeynet

In a typical corporate network, there often will be two subnets separated by a router. The internal network is the network of an organization which is protected by a firewall, and the Demilitarized Zone (DMZ) is the subnet where the honeypots are located. The honeypots can be running on physical machines or virtual machines. Usually, the DMZ is a copy of the internal network. The internal network always stores security information. Thus, a firewall stands in front of the internal network. The firewall allows the output traffic from the internal network freely, but filters some input traffic which are not permitted. For instance, some firewall rules can be set up so that the internal network cannot be accessed from outside but can be accessed by the user from DMZ.

There is a gateway called Honeywall which is used to monitor the honeypots, observe and record the adversaries' behavior when they compromise the honeypots. Honeywall is a gateway device that separates the honeypots from the rest of the world. Any traffic going to or from the honeypots must go through the Honeywall. This gateway is traditionally a layer 2 bridging/switching device, meaning the device should be invisible to anyone interacting with the honeypots. From the diagram, it can be observed that the Honeywall has 3 interfaces. The first 2 interfaces (eth0 and eth1) are what segregate the honeypots from everything else, and they are bridged interfaces that have no IP stack. The 3rd interface (eth2, which is optional) has an IP stack allowing for remote administration. Note that Honeywall is only one type of containment device gateway, so it can be substituted by future advanced devices.

With the development of virtualization technology, many virtual honeypots has been proposed and widely used for research and production purposes. Currently, there are several virtual honeypots focusing on capturing and analyzing automated malware, such as malicious webpage, infected file, etc. Cuckoo sandbox[10] is one of the virtual honeypots aimed to capture and analyze the automated malware. Therefore, in the Gen II honeynet architecture, the virtual machine can install the cuckoo sandbox running as a virtual honeypot to perform the malware capture and analysis task.

For example, a Windows virtual honeypot installed cuckoo sandbox can deal with PDF document that contain malware samples. Assume that the virtual machine was installed Adobe Acrobat Reader and had Internet connection, which are the requirements that can make the malware analysis running smoothly. The steps of how to apply the cuckoo sandbox to analysis the PDF file as a malware document are described in Appendix I.

[10] https://www.cuckoosandbox.org/

4.4. Conclusion

In this chapter, a multi-layer social engineering defense model is proposed, which is the decent security model to the research question one of this chapter. To the research question two and three of this chapter, the answers can be summaries as follows. The first layer, prevent layer, focus on preprocess the potential vulnerabilities that can be exploited by the SE attackers. The second layer, detect layer, is used to detect the SE attack that has broken the first layer. The third layer, control layer, is aimed to control every malicious behavior for further investigation or using as crime evidence, even if the social engineer compromised the information system resource. In each layer, several mechanisms are proposed to facilitate the defense purpose against different social engineering techniques in order to effectively protect the information-related resources and guarantee IT security. Besides, the proposed model is extendable, which means in the future more advanced mechanisms can be integrated into this SE defense model to improve its effectiveness.

The drawback of this chapter is the factor that the content of this chapter could be written in a far more detailed way with a pattern catalog of pattern specifications deduced from a nice taxonomy in chapter 3. In the future, therefore, a detailed social engineering defense model will be proposed in terms of pattern specifications deduced from a formalized taxonomy.

5. Measuring approaches

In this chapter, several suggestive measuring approaches will be provided without the complete implement. Despite this thesis does not emphasis on the study of measuring approaches. Two types of measuring approaches are considered: human-based measuring approach and computer-based measuring approach, which will be described in the next subsections.

5.1. human-based measuring approach

The human-based measuring approach focuses on measuring the employees' awareness to the social engineering attacks. The conventional measuring approaches are survey, questionnaire and even individual interview. In this thesis it is suggested to apply the score based measuring approach. A scoring scheme is proposed as Table 3 shows.

Table 3 Score grading for human-based measuring approach

Awareness Level	Measurement Score Grade
Negative	0-19
Weak-negative	20-39
Average	40-59
Weak-positive	60-79
Positive	80-100

The scoring scheme has five grades, which can be used to any survey, questionnaire and even individual interview that are used to test the employees' awareness consisting of knowledge, attitude and behavior. Definitely, there could be some other dimensions of employees' awareness can be used to test as well. The scoring scheme is helpful to quantify the human-based measuring approach for ease of generating statistics and making report.

Another essential concern is how to design the questions that will be used in the survey, questionnaire and interview. However, it is a complicated issue to design simple but effective measuring questions. The work of setting the effective questions to test the employees' awareness can be designed according to the behavior/attitude/knowledge triad matrix [29].

The first task is to determine what to measure. To this end, a value tree similar to the one in Fig.12 can be constructed.

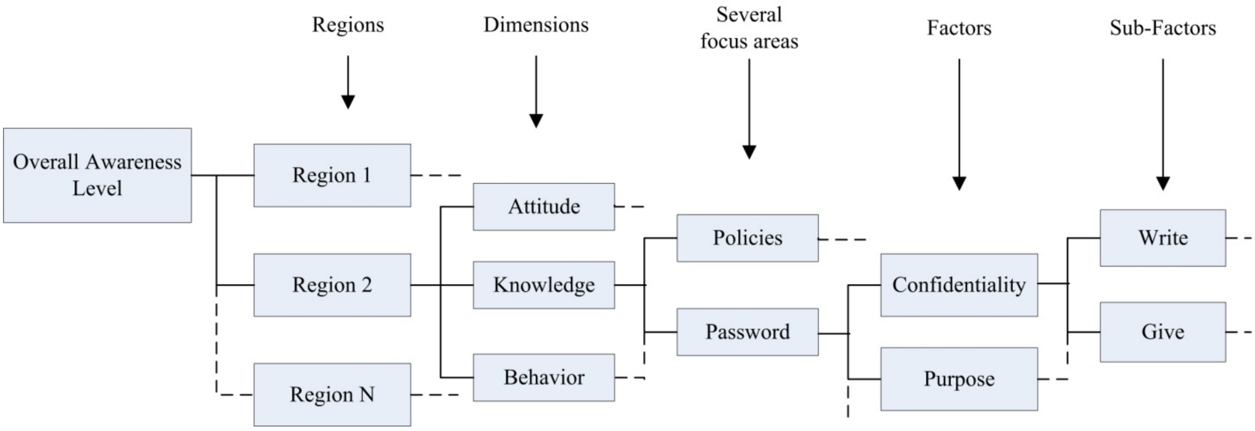


Figure 12 Tree structure of problem

From the tree, the social engineering related aspects are identified that can be measured to cover the knowledge, attitude and behavior dimensions with the associated focus areas in each dimension. Then, a questionnaire containing several questions (some questions are aimed to measure more than one aspect), to capture the information required is developed and tested at the region's head office as well as at one of the operational sites in the region. Different tests are performed and include tests using open-ended questions, multiple-choice questions, one-on-one contact with respondents and the use of e-mail facilities.

The example of questions can be described as follows:

Question to test knowledge:
Internet access on the company's systems is a corporate resource and should be used for business purposes only. 1. True 2. False 3. Do not know

Question to test attitude:
Mobile equipment is usually covered with existing insurance cover and there is no special need to include them in security policies. 1. True 2. False 3. Do not know

Question to test behavior:
I am aware that you should never give your password to somebody else – however, my work is of such a nature that I do give my password from time to time to a colleague (only to those that I trust!). 1. True 2. False

Thereafter, every region can get an awareness score according to the answers. The awareness score divide the full score of this region can calculate the awareness level of the employee in this region. For example, the awareness score is 68, so the awareness level is weak-positive in terms of Table 3

The above mentioned example of questions is only a sample. The further work of proposing effective questions for enhancing the employees' awareness needs other security researchers to address in the future.

5.2. computer-based measuring approach

The computer-based measuring approach is aimed to measure the organizational computer-based security mechanism against social engineering. This approach can include a number of technical penetrating tests over the information system resource that should be protected by the organizational security mechanism. Penetration Testing evaluates the effectiveness of the information security program and identifies weaknesses. Penetration tests are great leverage points for the security group to get management exposure and emphasize the importance towards critical security programs.

This thesis does not focus on analyzing and applying the social engineering toolkits for penetrating tests. This work just lists several SE toolkits that could be taken into account to apply by the interested researchers (see Table 4).

Table 4 A list of SE toolkits for penetrating tests

Tool	Usage
Metasploit[11]	General
SET[12] (Social-Engineer Toolkit)	Social engineering dedicated
SecurityIQ[13]	Phishing dedicated

SET is an open-source Python-driven tool. It is a dedicated social engineering framework aimed at penetration testing around social engineering. Hence it is the most popular toolkits among social engineer. Metasploit is a general penetrating toolkit that can be used to perform a variety of network attacking, such as buffer overflow, DDoS, SQL Inject, etc. Nevertheless, it also integrates the SET. Hence the Metasploit user can enjoy using the SET over Metasploit. SecurityIQ is a dedicated fully customizable Spear Phishing Simulator, which can help the penetrator in charge of security protect their organization by combining a phishing simulator and security awareness learning management platform into a single easy to use solution.

[11] https://www.rapid7.com/resources/videos/phishing-campaigns-in-metasploit-pro.jsp

[12] https://www.trustedsec.com/social-engineer-toolkit/

[13] http://resources.infosecinstitute.com/phishing-and-social-engineering-techniques/

A penetration test's main purpose is to simulate an attacker and the ability to impact the business' ability to generate revenue. All the tools mentioned above can simulate attackers attempting to circumvent security controls and gain unauthorized access to systems. They can be used to impact the organization in a way to identify systemic weaknesses within the overall information security program. They will penetrate an organization and attempt to identify weaknesses and attempt to gain access to sensitive systems, intellectual property, and/or key business systems.

5.3. Conclusion

This chapter suggests two measuring approaches: human-based measuring approach and computer-based measuring approach. The measurement score grade is proposed as the metric. The tree structure of problem can be used to deduce the criteria to set awareness questions. There are two shortcomings in this chapter. First, the methodology of proposing awareness questions is not proposed. Second, the author should develop an automated SE penetration test through any suggestive social engineering toolkits. These two task could be the work in the future.

6. Conclusion and Future Work

Social engineering attack is an open issue and big challenge to the IT security of the modern organizations. In this thesis, the author performed a study work on understanding SE attacks, defending SE attacks and measuring security mechanisms against SE attacks.

First, the author reviewed the state of the art of the social engineering including the conceptual models, taxonomies, security policies, detection models, and measuring approaches. Though there are a number of proposals but some open issues were not addressed very well. For example, there is not a detailed taxonomy of SE attacks which can be used to fully analyze SE attacks. There is also a lack of an extendable SE defense model which can be used to deal with different SE attacks. Hence, they are the problem space where the author can make effort.

Second, a novel taxonomy of social engineering attacks was proposed. It is a detailed taxonomy, however, it is still can be extended based on the current structure for the future categories. Thanks to this novel SE taxonomy, the security researchers can gain insight of different SE attacks due to its fully classification scheme. It can also help the security researchers to predict the future SE attacks in terms of the attacking space that can be formed by the combination of different categories.

Third, a multi-layer social engineering defense model was designed and proposed. This SE defense model consists of three layers: protect, detect and control. Each layer has its own defending aim. The protect-layer is aimed to mitigate the potential social engineering vulnerabilities, though we know risk cannot be eliminated but can only be reduced. The detect-layer is used to detect the malicious behavior that is intended to exploit the SE vulnerabilities. And the control-layer is to guarantee that any malicious will be captured even though the SE attackers compromised the information system resources. Hence, in each layer several mechanisms are proposed to facilitate these defending aims in order to resist different SE attacks. Furthermore, the SE defense model can be extended through adding new advanced mechanisms into the corresponding layer for the future.

Fourth, several suggestive measuring approaches are provided, including human-based measuring approach and computer-based measuring approach. The author prefers using scoring scheme to quantify the measuring approach of psychology awareness. Some candidate social engineering toolkits are listed for penetrating tests.

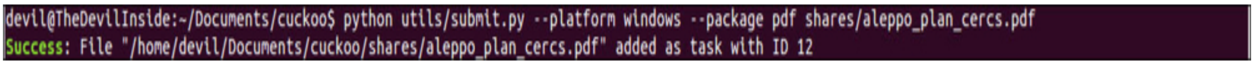
However, the work including the taxonomy and the defense model that can be improved if the author formalized the taxonomy in ontology, and deduce the pattern specifications from the taxonomy in order to properly analyze real social engineering scenarios. In addition, a translation of the findings into an attack-process tree should be worked out to help proposing a detailed social engineering defense model. Furthermore, it is planned to develop some penetrating tests based on the SET to evaluate the proposed SE defense model. That can be conducted by launching a variety of social engineering attacks to the information system resource that are conducted under the proposed SE defense model. Hope this work is useful and can bring inspiration to the interested security researchers and other people in the field.

Appendix I

First, on the host (which is a Linux OS, i.e. Ubuntu), open a new Terminal and type the following command line:

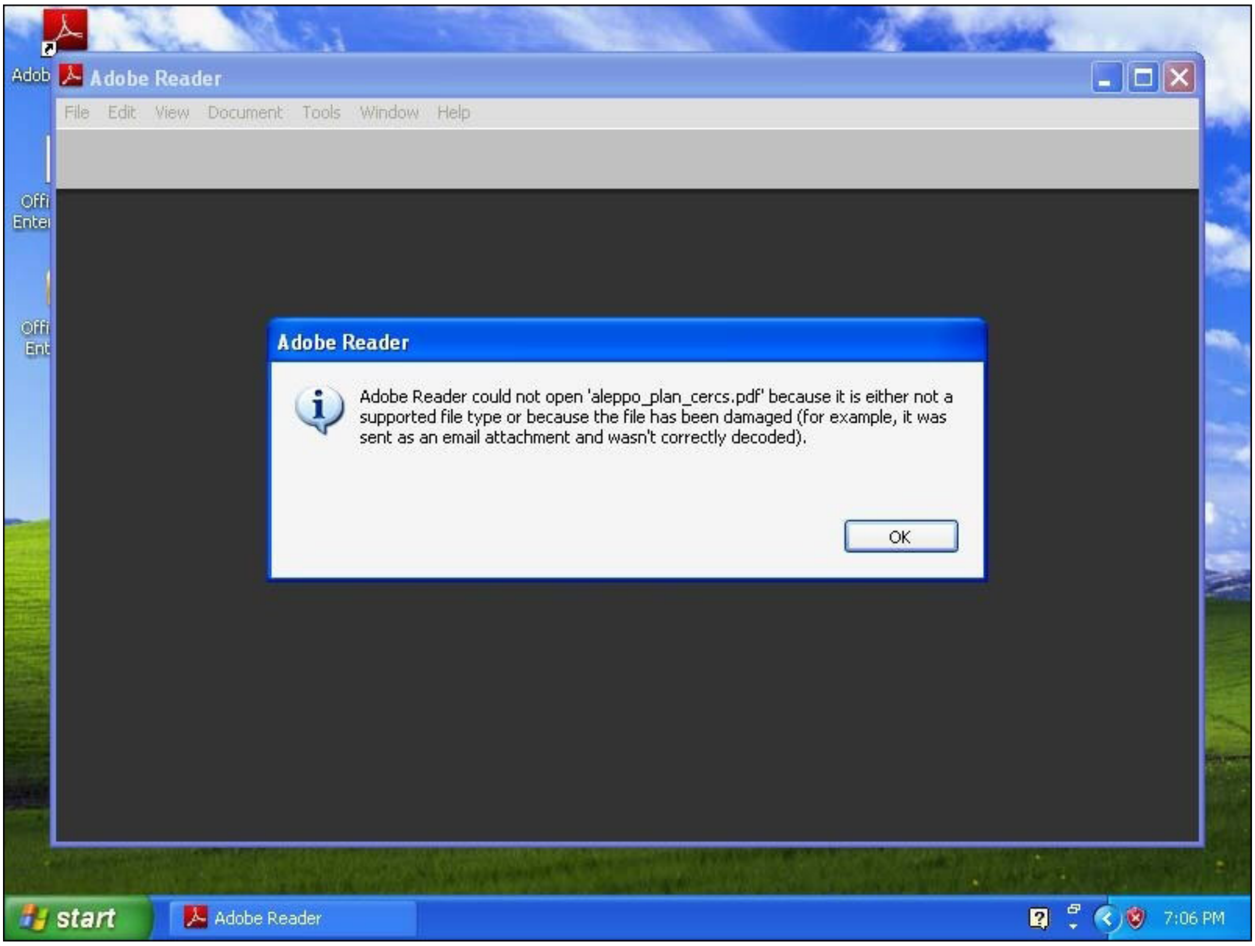
```
$ python uti
ls/submit.py --platform windows --package pdf shares/aleppo_plan_cercs.pdf
```

The Aleppo_plan_cercs.pdf is the malware document. Then if the process is successful, you will get an output as the following screenshot:



```
devil@TheDevilInside:~/Documents/cuckoo$ python utils/submit.py --platform windows --package pdf shares/aleppo_plan_cercs.pdf
Success: File "/home/devil/Documents/cuckoo/shares/aleppo_plan_cercs.pdf" added as task with ID 12
```

Cuckoo will then start taking the latest snapshot of the virtual machine that has been made. Windows will open the PDF document automatically.



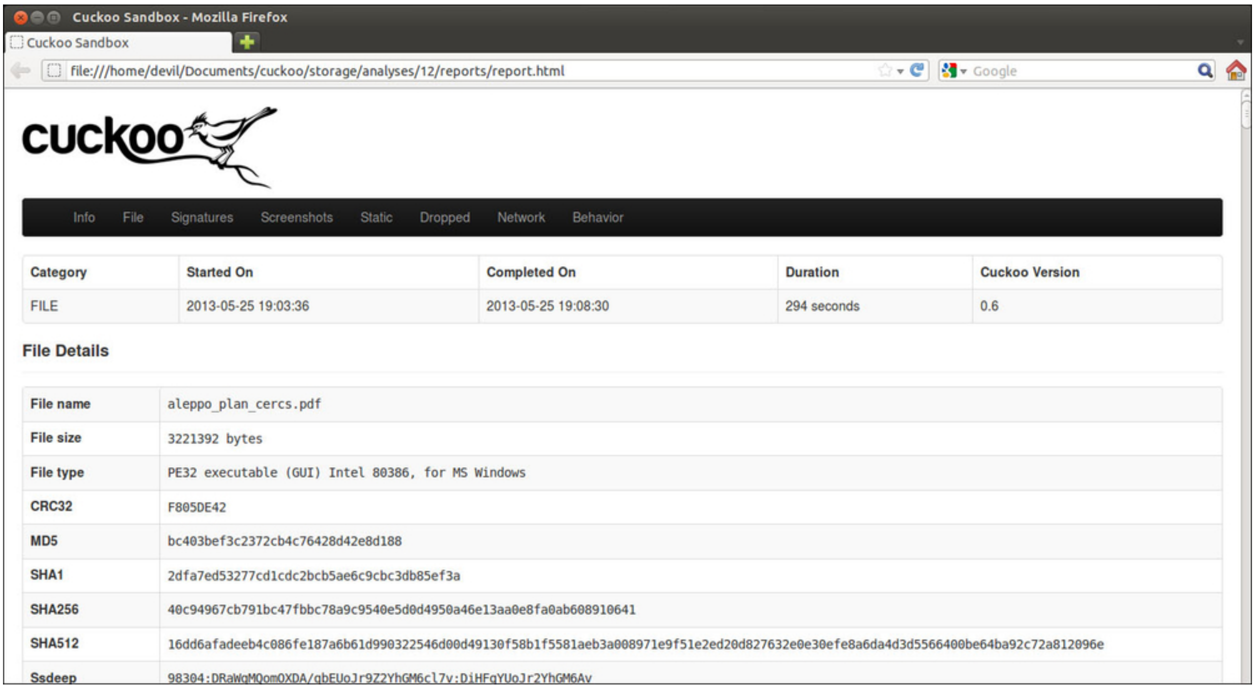
Adobe Reader

Adobe Reader could not open 'aleppo_plan_cercs.pdf' because it is either not a supported file type or because the file has been damaged (for example, it was sent as an email attachment and wasn't correctly decoded).

OK

It seems that the document cannot be opened. You may want to know why. The answer to this may be available at the Cuckoo report. Click on OK in the information window. Wait a moment to make sure that Cuckoo can log all the activities happening. Close Adobe Reader and wait until VM closes automatically.

After the VM has closed and task 12 (this task ID may be different in your OS) is finished, let's see the report.html file which is available at storage/analyses/12. Now, you can open the report.html file in your web browser.

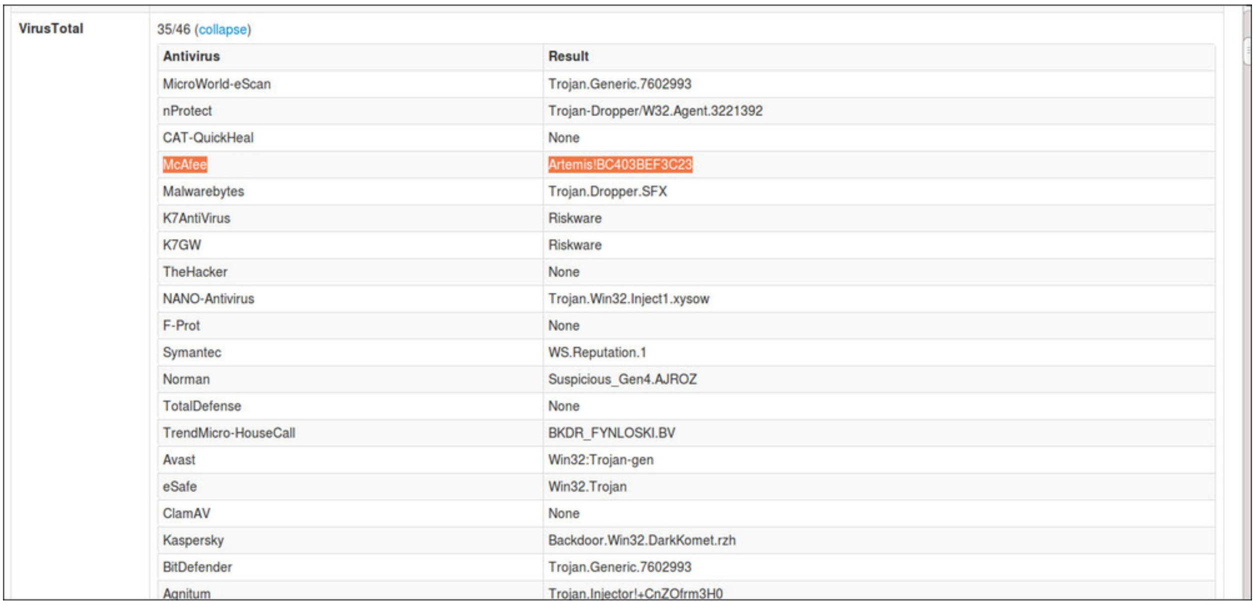

Cuckoo Sandbox - Mozilla Firefox
Cuckoo Sandbox
file:///home/devil/Documents/cuckoo/storage/analyses/12/reports/report.html
Google
cuckoo
Info
File
Signatures
Screenshots
Static
Dropped
Network
Behavior

Category	Started On	Completed On	Duration	Cuckoo Version
FILE	2013-05-25 19:03:36	2013-05-25 19:08:30	294 seconds	0.6

File Details

File name	aleppo_plan_cercs.pdf
File size	3221392 bytes
File type	PE32 executable (GUI) Intel 80386, for MS Windows
CRC32	F805DE42
MD5	bc403bef3c2372cb4c76428d42e8d188
SHA1	2dfa7ed53277cd1cdc2bcb5ae6c9cbc3db85ef3a
SHA256	40c94967cb791bc47fbbc78a9c9540e5d0d4950a46e13aa0e8fa0ab608910641
SHA512	16dd6afadeeb4c086fe187a6b61d990322546d00d49130f58b1f5581aeb3a008971e9f51e2ed20d827632e0e30efe8a6da4d3d5566400be64ba92c72a812096e
Ssdeep	98304:DRaWgMQomOXDA/qbEUoJr9Z2YhGM6cl7v:DiHFqYUoJr2YhGM6Av

And the report in the VirusTotal section will be:



VirusTotal

35/46 (collapse)

Antivirus	Result
MicroWorld-eScan	Trojan.Generic.7602993
nProtect	Trojan-Dropper/W32.Agent.3221392
CAT-QuickHeal	None
McAfee	Artemis!BC403BEF3C23
Malwarebytes	Trojan.Dropper.SFX
K7AntiVirus	Riskware
K7GW	Riskware
TheHacker	None
NANO-Antivirus	Trojan.Win32.Inject1.xysow
F-Prot	None
Symantec	WS.Reputation.1
Norman	Suspicious_Gen4.AJROZ
TotalDefense	None
TrendMicro-HouseCall	BKDR_FYNLOSKI.BV
Avast	Win32:Trojan-gen
eSafe	Win32.Trojan
ClamAV	None
Kaspersky	Backdoor.Win32.DarkKomet.rzh
BitDefender	Trojan.Generic.7602993
Agnitum	Trojan.Injector!+CnZOfrm3H0

From the report of VirusTotal, one can see that the malware PDF is a Trojan. McAfee antivirus called this malware **Artemis!BC403BEF3C23**, while ClamAV seems to not recognize it. Kaspersky calls it by the name **Backdoor.Win32.DarkKomet.rzh**. Whatever the name is, it is concluded that the document may harm your computer by because it contains Trojan inside it.

Reference

1. Social Engineer. What is phishing - paypal phishing examples. available online: http://www.socialengineer.org/wiki/archives/Phishing/Phishing-PayPal.html, last accessed on 2013-07-04.
2. Google hack attack was ultra sophisticated. available online: http://www.wired.com/threatlevel/2010/01/operation-aurora/, last accessed on 2013-07-17.
3. Microsoft hacked: Joins apple, facebook, twitter - InformationWeek. available online: http://www.informationweek.com/security/\Attackacks/microsoft-hacked-joins-apple-facebook-tw/240149323, last accessed on 2013-07-10.
4. N. Perlroth. Chinese hackers infiltrate new york times computers, Jan. 2013. available at https://www.nytimes.com/2013/01/31/technology/chinese-hackers-infiltrate-new-york-times-computers.html, last accessed on: 2013-07-01.
5. Hossein Bidgoli. *Handbook of Information Security, Information Warfare, Social, Legal, and International Issues and Security Foundations (Handbook of Information Security).* John Wiley & Sons, Inc., New York, NY, USA. 2006.
6. Ji-Xuan Feng and Janet Hughes. 2009. Analyzing privacy and security issues in the information age - an ethical perspective. *WSEAS Trans. Info. Sci. and App.* 6, 1 (January 2009), 126-135.
7. Honeywall project. Know Your Enemy: Learning about Security Threats. Addison Wesley, 2004.
8. Joshi, R.C. and Sardana, A. eds., 2011. Honeypots: A New Paradigm to Information Security. CRC Press.
9. Dimensional Research Study about Social Engineering, Available on: http://www.checkpoint.com/press/downloads/social-engineering-survey.pdf.
10. SANS Institute InfoSec Reading Room, Available on: http://www.sans.org/reading-room/whitepapers/engineering/threat-social-engineering-defense-1232.
11. Amazon’s customer service backdoor, Available on: https://medium.com/@espringe/amazon-s-customer-service-backdoor-be375b3428c4#.n2cjkqgv9
12. Maung K. Sein, Ola Henfridsson, Sandeep Purao, Matti Rossi, and Rikard Lindgren, “Action design research,” MIS Quarterly. Vol. 35, Issue 1, No. 2, pp. 37-56, March 2011.
13. Tipton, Harold F., and Micki Krause. Information security management handbook. CRC Press, 2003.
14. Hadnagy, Christopher. Social engineering: The art of human hacking. John Wiley & Sons, 2010
15. Mann, Mr Ian. Hacking the human: social engineering techniques and security countermeasures. Gower Publishing, Ltd., 2012.
16. Daniel L Lough, "A Taxonomy of Computer Attacks with Applications to Wireless Networks," Virginia Tech, Ph.D. Dissertation April 2001.
17. Simon Hansman and Ray Hunt, "A Taxonomy of Network and Computer Attacks," Computers & Security, vol. 24, no. 1, pp. 31-43, February 2005.
18. A. Avizienis, J. C. Laprie, B. Randell and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing," in *IEEE Transactions on Dependable and Secure Computing*, vol. 1, no. 1, pp. 11-33, Jan.-March 2004.

19. Chris Simmons, Charles Ellis, Sajjan Shiva, Dipankar Dasgupta, and Qishi Wu, "AVOIDIT: A Cyber Attack Taxonomy," University of Memphis, Technical Report CS-09-003, August 2009.
20. van Heerden, R. P., Irwin, B., Burke, I. D., & Leenen, L. (2012). A Computer Network Attack Taxonomy and Ontology. International Journal of Cyber Warfare and Terrorism (IJCWT), 2(3), 12-25. doi:10.4018/ijcwt.2012070102
21. Krombholz, Katharina, Heidelinde Hobel, Markus Huber, and Edgar Weippl. "Advanced social engineering attacks." Journal of Information Security and applications 22 (2015): 113-122.
22. Ryan Heartfield, and George Loukas. "A Taxonomy of Attacks and a Survey of Defence Mechanisms for Semantic Social Engineering Attacks." ACM Computing Surveys (CSUR) 48, no. 3 (2015): 37.
23. CESG. 2015. Common Cyber Attacks: Reducing the Impact. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/400106/Common_Cyber_Attacks-Reducing_The_Impact.pdf.
24. F. Mouton, L. Leenen, M. M. Malan, and H. S. Venter, “Towards an ontological model defining the social engineering domain,” in 11th Human Choice and Computers International Conference, Turku, Finland, July 2014, pp. 266–279.
25. Study on the Theft of Proprietary Information, American Society of Industrial Security, Arlington, VA: ASIS, 1996
26. T. N. Jagatic, N. A. Johnson, M. Jakobsson, F. Menczer, “Social Phishing”, Communications of the ACM, Vol. 50, No. 10, Oct. 2007
27. Manjak, Martin. "Social engineering your employees to information security."GIAC Gold Paper for Security Essentials, As part of the Information Security Reading Room, SANS Institute (2006).
28. Nohlberg, Marcus, Stewart Kowalski, and Markus Huber. "Measuring Readiness against Automated Social Engineering." (2008): 20-1.
29. Nohlberg, Marcus, and Stewart Kowalski. "The cycle of deception: a model of social engineering attacks, defenses and victims." In Second International Symposium on Human Aspects of Information Security and Assurance (HAISA 2008), Plymouth, UK, 8-9 July 2008, pp. 1-11. University of Plymouth, 2008.
30. Kruger, Hennie A., and Wayne D. Kearney. "A prototype for assessing information security awareness." computers & security 25, no. 4 (2006): 289-296.
31. Yi Cheng, Julia Deng, Jason Li, Scott A. DeLoach, Anoop Singhal, and Xinming Ou. "Metrics of Security," In Cyber Defense and Situational Awareness, pp. 263-295. Springer International Publishing, 2014.
32. Mitnick, K., The Art of deception. Indianapolis:Wiley Publishing, Inc., ISBN: 076454280X, 2002
33. Mouton, Francois, Mercia M. Malan, Louise Leenen, and Hein S. Venter. "Social engineering attack framework." In Information Security for South Africa (ISSA), 2014, pp. 1-9. IEEE, 2014.
34. Gonzalez, Jose J., Jose M. Sarriegi, and Alazne Gurrutxaga. "A framework for conceptualizing social engineering attacks." In Critical Information Infrastructures Security, pp. 79-90. Springer Berlin Heidelberg, 2006.
35. Tetri, Pekka, and Jukka Vuorinen. "Dissecting social engineering."Behaviour & Information Technology 32, no. 10 (2013): 1014-1023.
36. Abraham, Sherly, and InduShobha Chengalur-Smith. "An overview of social engineering malware: Trends, tactics, and implications." Technology in Society 32, no. 3 (2010): 183-196.

37. M. Bezuidenhout, F. Mouton and H. S. Venter, "Social engineering attack detection model: SEADM," *2010 Information Security for South Africa*, Sandton, Johannesburg, 2010, pp. 1-8.
38. Francois Mouton, Mercia M. Malan, and Hein S. Venter. "Development of Cognitive Functioning Psychological Measures for the SEADM." In HAISA, pp. 40-51. 2012.
39. Francois Mouton, Louise Leenen and H.S. Venter, Social Engineering Attack Detection Model:SEADMv2, 2015 International Conference on Cyberworlds, 2015
40. R. Bhakta and I. G. Harris, "Semantic analysis of dialogs to detect social engineering attacks," *Semantic Computing (ICSC), 2015 IEEE International Conference on*, Anaheim, CA, 2015, pp. 424-427.
41. Gragg, David. "A multi-level defense against social engineering." SANS Reading Room, March 13 (2003).
42. Tolga Mataracioglu, Sevgi Ozkan, and Ray Hackney. "Towards a Security Lifecycle Model against Social Engineering Attacks: SLM-SEA." arXiv preprint arXiv:1507.02458 (2015).
43. International Organization for Standardization, 2005. ISO/IEC 27001 Information Technology - Security Techniques - Information Security Management Systems – Requirements.
44. Allan, Ant, Noakes-Fry, Kristen, Mogull, Rich. “Business Update: How Businesses Can Defend Against Social Engineering Attacks”. Gartner, March 16, 2005.
45. M. Cova, C. Kruegel, and G. Vigna, “Detection and analysis of drive-by-download attacks and malicious JavaScript code,” In Proceedings of the 19th International Conference on World Wide Web. ACM, 281–290.
46. Z. Li, K. Zhang, Y. Xie, F. Yu, and X. Wang, “Knowing your enemy: Understanding and detecting malicious web advertising,” In Proceedings of the ACM Conference on Computer and Communications Security, 2012.
47. A. P. Felt and D. Wagner, “Phishing on Mobile Devices,” In W2SP.
48. M. Mannan and P. C. van Oorschot, “On instant messaging worms, analysis and countermeasures,” In Proceedings of the ACM Workshop on Rapid Malcode, 2–11, 2005.
49. S. Ford, M. Cova, C. Kruegel, and G. Vigna, “Analyzing and detecting malicious flash advertisements,” In Proceedings of the Annual Computer Security Applications Conference (ACSAC’09). IEEE, 363–372, 2009
50. G. Madlmayr, J. Langer, C. Kantner, and J. Scharinger, “NFC devices: Security and privacy,” In Availability, Reliability and Security (ARES’08). IEEE, 642–647, 2008
51. J. R. Jacobs, “Measuring the Effectiveness of the USB Flash Drive as a Vector for Social Engineering Attacks on Commercial and Residential Computer Systems,” Master’s thesis, Embry-Riddle Aeronautical University.
52. A. Gazet, “Comparative analysis of various ransomware virii,” Journal in Computer Virology 6, 1, 77–90, 2010
53. P. Ducklin. 2014. Anatomy of an Android SMS Virus—Watch Out for TextMessages, Even from Your Friends!Retrieved from https://nakedsecurity.sophos.com/2014/06/29/anatomy-of-an-android-sms-virus-watchout-for-text-messages-even-from-your-friends/.
54. Andrew Mathews, "Why worry? The cognitive function of anxiety."Behaviour research and therapy 28, no. 6 (1990): 455-468.
55. Monchi, Oury, Michael Petrides, Valentina Petre, Keith Worsley, and Alain Dagher. "Wisconsin Card Sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging." The Journal of Neuroscience 21, no. 19 (2001): 7733-7741.

56. Eriksen, Charles W. "The flankers task and response competition: A useful tool for investigating a variety of cognitive problems." Visual Cognition 2, no. 2-3 (1995): 101-118.
57. Cicchetti, Domenic V., and Byron P. Rourke, eds. Methodological and biostatistical foundations of clinical neuropsychology and medical and health disciplines. CRC Press, 2004.
58. L. Spitzner, "Honeypots: catching the insider threat," *Computer Security Applications Conference, 2003. Proceedings. 19th Annual*, 2003, pp. 170-179.