1 INTRODUCTION

This research project aims to understand the characterization of phishing email properties considering persuasion techniques. This chapter provides an introductory information regarding the research area. The objectives are to understand the challenges and introduce ways to overcome those challenges. Section 1.1 discusses the problem statement related to phishing in general and phishing email in particular, section 1.2 presents the research goal of the thesis, section 1.3 presents the outline of the research questions, section 1.4 describes the outline of the research method used to answer the research questions and section 1.5 presents the structure of the thesis.

1.1 PROBLEM STATEMENT

With the enormous growing trends of information technology in modern generation, the evolution of digital era has become more mature in the sense of effectiveness and easiness for societies. Trusted entities such as financial institutions may offer their products and services to the public through the Internet. It has greatly impacted our society in different ways, for example the way we communicate with each other. In the recent days, we are no longer need to use computer to send an email, we can just use our smartphone, which we carry every day in our pockets, with internet connectivity to send an email. Consequently, human society has been enthusiastically eager to utilize technology means such as emails, websites, online payment system, social networks to achieve their tasks efficiently, affordable and more relevant. However, the advancement in information and communication technology has been a double-edged sword. It becomes easier to get personal information about someone in the cyber world. Cyber criminals see this opportunity as a way to manipulate consumers and exploit their confidential information such as usernames, passwords, bank account information, credit card or social security numbers. Personalized information about someone such as email address, phone number, birth date, relationships or work place might be obtained from the internet such as online social network sites. As a result, cyber criminals can compose an attack in a personalized way to persuade intended victim to grant their malicious requests.

One of the well known cyber crimes is called phishing. One of the objectives of phishing attack is to gain a financial benefit by masquerading legitimate institutions [**?**]. Evidently, phishing techniques is often associated with bogus emails and websites [**?**][**?**]. Moreover,

studies suggest that the essence of this malicious behavior is the art of pretext social engineering and deception [**?**][**?**][**?**][**?**][**?**]. Social engineering involves the techniques used in order to manipulate people into responding to perform an action or disclosing sensitive information [**?**]. Social engineers often attempt to persuade potential victims to engage in some short of emotion such as excitement and fear as well as interpersonal relationship such as trust and commitments [**?**]. Such deception and social engineering attacks might be delivered through phone calls, text messages, private messages or emails as a medium to distract recipients' decisions.

To make phishing email efficient, its context might require the intended victim to urgently act upon it, for example an email informs about suspending an account if recipient does not respond or perform an action within a limited time. In order to gain trust from recipient in a phishing email, persuasion techniques are used to get a positive response from intended victim. Several countermeasures have been studied to detect phishing emails, such as the technique of looking for bit string that are previously determined as spams and phishing emails [**?**]. The common characteristics of bogus email sent by the phishers would be misleading hyperlinks and misleading header information [**?** **?**]. Furthermore, one of the non technical approaches to defense against phishing attacks is to make people aware of the threats. Security awareness in regards of phishing attacks might be achieved by aiding an education to the public to ignore links within an email, even though the source of the email appear to be legitimate [**?**].

The success of phishing attack through distributed emails is determined by the response of the unsuspecting recipients. User decisions to click a link or open an attachment in an email might be influenced by how strong a phisher can persuade the intended victim. However, current literatures so far show an absence of real world analysis in characterizing phishing email properties by considering persuasion techniques, which it might be important for user decision. Thus, the characterization of phishing email properties by considering persuasion techniques in the real word analysis could fill the void to show to what extent do they involve in phishing emails.

1.2 RESEARCH GOAL

The main goal of this research is to characterize phishing email properties considering persuasive principles by finding the association between generic properties and persuasive principles. These generic properties consist of phishing email structural properties or features based on the literature survey findings. Each of these properties and each of the persuasive principle will be introduced as a variable in our methodology. We will look for frequency and relationship in-

volving these variables. This relationship can be used to show new characteristics of phishing email properties considering the persuasive elements within its content. The characterization of phishing email considering persuasive principle also can be used to generate a new method in an automatic way of detecting phishing email as one of the primary delivery techniques of phishing attacks.

1.3 RESEARCH QUESTIONS

To be able to meet the goal, we formulated two main research questions as follows:

- RQ1: What are the characteristics of reported phishing emails?
- RQ2: To what extent the persuasive principles are used in phishing emails?

Several aspects of phishing email characteristics and hypotheses related to the research questions will be addressed in detail in **??**.

1.4 RESEARCH METHODOLOGY

This subsection describes the outline of our research methodology of our analysis, more details will be addressed in **??**. Literature surveys and experimental exploration will be conducted in order to have a basic understanding concerning phishing in general. Furthermore, the data for the our analysis will be collected from security organization based in Netherlands which we can not disclose its identity. We are looking for reported phishing emails in 2013. We will be enumerating the outline our methodology into several steps:

1. Data collection
2. Data categorization
3. Variables and concepts
5. Data classification
6. Analysis

Data categorization will be determined by several aspects; languages (Dutch, English, others) and whether the email is indeed a phish, spam or legitimate. We predict there will be duplicated phishing emails so that we will only consider only unique email in our dataset. Structural properties of phishing emails will be reviewed in our literature survey and we will establish our variables based on these properties. Furthermore, the data classification phase will be indicated by our coding of the categorization result into these variables. The data will be analyzed from mainly three different viewpoints; properties characteristics, persuasion principles characteristics and their relationships. Lastly, we will look at how the findings answer our research questions.

This research project is structured as follows:

Chapter 2 describes background and literature reviews about phishing in general. The subsections include; a general understanding of what is phishing in term of history and definition, an overview of its damage in term of money, an exploration of its modus operandi based on phishing stages or phases, a basic understanding of bad neighborhood and phishing, general phishing countermeasures and lastly the human factor in phishing.

Experimental exploration with methods and results will be conducted in chapter 3. It consists of an experiment on the existence of bad neighborhood and finding differences of phishing website vs its original website. The methods and the results of the individual experiment will be presented in the respective sections.

In chapter 4 presents the rationale of our main research questions and hypotheses. It includes what aspect to be considered to answer the characteristics of phishing emails in the dataset and the motivation of our hypotheses to support our research questions.

In chapter 5, we will discuss our main data analysis and results. It includes the details of research methodology that we conducted as well as the results of our analysis.

Lastly, in chapter 6 we will present our discussion and conclusion of the research project, how the research questions are answered along with the recommendations, limitations and how these limitations could become the basis of further research.