Decision Support Systems 50 (2011) 662–672

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Assessing the severity of phishing attacks: A hybrid data mining approach

Xi Chen [a,*], Indranil Bose [b], Alvin Chung Man Leung [b,c], Chenhui Guo [d]

[a] School of Management, Zhejiang University, China
[b] School of Business, The University of Hong Kong, Hong Kong
[c] McCombs School of Business, The University of Texas at Austin, TX, United States
[d] Eller College of Management, The University of Arizona, AZ, United States

ARTICLE INFO

Available online 19 August 2010

Keywords:
Financial loss
Phishing
Risk
Supervised classification
Text phrase extraction
Variable importance

ABSTRACT

Phishing is an online crime that increasingly plagues firms and their consumers. We assess the severity of phishing attacks in terms of their risk levels and the potential loss in market value suffered by the targeted firms. We analyze 1030 phishing alerts released on a public database as well as financial data related to the targeted firms using a hybrid method that predicts the severity of the attack with up to 89% accuracy using text phrase extraction and supervised classification. Our research identifies some important textual and financial variables that impact the severity of the attacks and potential financial loss.

1. Introduction

Phishing is a major security threat to the online community. It is a kind of identity theft that makes use of social engineering skills and technical subterfuge to entice the unsuspecting online consumer to give away their personal information and financial credentials [5]. A typical phishing attack consists of four phases, namely, preparation, mass broadcast, mature, and account hijack [8]. The tremendous financial impact of phishing is borne by the fact that phishing caused an estimated financial loss of US $3.2 billion affecting 3.6 million people from September 2006 to August 2007 [40]. The number of reported phishing incidents grew exponentially, and increased by 293.7% from 8829 in December 2004 to 34,758 in October 2008 [4,5]. Not only do phishing attacks cause financial loss, but they also shatter the confidence of customers in conducting e-commerce. Managers of some of the US super regional banks have indicated that the deteriorating customer trust is a major concern with respect to phishing [46]. A recent survey found that most customers of European banks only use online banking to check their account balances instead of conducting online transactions due to the fear of getting phished [15]. Another study also reported that the customer fear psychosis has resulted in a 20% decrease in the rate of opening of genuine emails [10].

To make customers aware of latest phishing attacks, some international organizations and government statutory bodies, such as the Anti-phishing Working Group (APWG), have published phishing alerts on their websites. To assess the risk level of each phishing attack, some firms have sought help from information security experts who evaluated reported phishing incidents based on the contents of the phishing email and the phishing websites. However, as phishing incidents continue to increase at a tremendous rate, the manual risk assessment method involving experts may be too slow. Data mining techniques can improve the assessment of phishing attacks. They can discover the knowledge embedded in the traits of prior phishing attacks and identify the inherent characteristics that contribute to the different risk levels of a phishing attack. This can help predict the associated risk level of a new phishing incident in a short period of time with a reasonable accuracy. Furthermore, the risk level, which is based on the technical sophistication of phishing attacks, may not be directly related to financial loss caused by an attack. Past research has shown that the impact of sophisticated phishing alerts on stock markets is not as significant as phishing alerts whose risk level is considered to be moderate [33]. However, the financial loss resulting from a phishing attack is always of great concern to security administrators as well as consumers of an organization. Therefore, a warning mechanism that can identify the phishing incidents that are either very risky or likely to cause a large financial loss will be of great interest to shareholders and senior managers of the targeted companies.

In this research we use supervised classification techniques, which is a major stream of data mining, to assess the severity of phishing attacks. At the same time, we identify the key antecedents that contribute to a high risk level or a high financial loss generation by a phishing attack. We use a hybrid approach which combines key phrase extraction and supervised classification methods that makes use of the textual data description of the phishing attack as well as financial data of the targeted company to assess the severity of a phishing attack according to its risk level or financial loss generating

* Corresponding author.
E-mail addresses: chen_xi@zju.edu.cn (X. Chen), bose@business.hku.hk (I. Bose), aleung@mail.utexas.edu (A.C.M. Leung), chguo@email.arizona.edu (C. Guo).

0167-9236/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2010.08.020

potential. The three classifiers used for this purpose result in a classification accuracy of up to 89%. Our results also show that the key identifying variables for risk level and potential financial loss of phishing attacks are different from each other. High risk level is associated with phishing emails that ask customers of large firms to update their accounts whereas high financial loss is characterized by phishing attacks targeted to customers of large firms that have high total liabilities.

2. Literature review

Phishing has aroused great interest among information security researchers. Understanding the critical success factors of phishing and determining methods that can prevent or detect such a crime has been a popular area of research. We can roughly split current research on phishing into three streams, namely, phenomenal studies, economic analysis, and technical research.

As an example of a phenomenal study related to phishing, Jagatic et al. found that the social engineering skill of the adversary was a critical success factor for phishing [26]. Dhamija and Tygar discovered that lack of knowledge, inability to control visual deception, and lack of attentiveness to detail are the major weaknesses of people who fall prey to phishing attacks [14]. Interestingly, Workman found that the critical success factors for some marketing strategies were applicable to phishing attacks as well [51]. Researchers also found that education of customers, standardization of technology, and sharing of phishing information were among the most important policies that could deter phishing attacks [35]. Some researchers conducted experimental studies and confirmed that if a user was trained to identify phishing attacks, the chance of being cheated in future was significantly lowered [32].

Among economic studies related to phishing, Jakobsson classified the costs of phishing into three categories, namely, direct cost, indirect cost, and opportunity cost [27]. Singh studied a number of international phishing incidents and found that the direct financial loss per incident ranged from US $900 to 6.5 million pounds [45]. However, it is widely believed that as companies are quite reluctant to disclose information related to direct financial loss caused by phishing, the actual financial loss might be ten times more than the estimated numbers that appeared in research reports [21]. In their attempt to estimate the indirect financial loss caused by phishing, Leung and Bose found that phishing related announcements caused a significant negative reaction among investors of targeted companies [33]. It is interesting to note that a significant negative investor reaction of 2.1% loss in market value within two days of the announcement was reported in the broader context of analyzing the economic impact of information security breaches [30].

In the area of technical research, information security researchers have toiled to discover better countermeasures of phishing. A number of anti-phishing toolbars and phishing filters have been developed. Data mining based approaches have been frequently adopted in the development of such countermeasures. Data mining techniques have been used to filter out phishing emails that contained fraudulent messages [1]. By analyzing the headers of emails, researchers were able to prevent the spread of malicious emails containing virus/worms/Trojans, and stop crimes such as phishing and distributed denial of service attacks with an accuracy of 99% [52]. Among the various data mining techniques that have been adopted for determination of phishing emails are support vector machines [12], random forest [18], one-step ternary and repeated binary classification techniques [19], and ensemble methods [43]. To authenticate the URL embedded in the emails, logistic regression [20] and decision trees have also been used [37]. The focus of the extant research was on the analysis of the characteristics of the emails and determination of the malicious nature of the emails. However, the focus of the current research is on the assessment of the influence of such phishing emails.

The use of data mining techniques in research related to information security is not new. Zhu et al. had used rough sets, neural networks, and decision trees for detection of network intrusion [55]. They tried different combinations of classification tools and data representation format, and experimented with variation in the proportion of training and testing data. They showed that rough sets performed best when the data was presented in binary format, and the proportion of training and testing data was balanced [55]. Zhao et al. proposed a hybrid system for network intrusion detection [54]. The system consisted of three main components: service pattern databases that were used to detect the network traffic patterns for different services, anomaly detection module that was based on unsupervised clustering and detected anomalous network traffic patterns, and a random forest module that was used to differentiate between intrusion cases and normal cases of service usage. Zhao and Huang proposed a data mining approach that mimicked the human immune system and detected network intrusion [53]. Ansari et al. detected misuse and anomalies using a soft computing method like fuzzy logic [3]. From the various examples cited in this paragraph we can see that data mining techniques have been favored by researchers in the area of information security. For a comprehensive review on this topic the interested reader may refer to Tsai et al. [47]. However, past research mainly focused on the detection of security events such as misuse, anomalies, intrusion, and other types of security breaches. Research on the use of data mining techniques to assess the influence of security events such as phishing attacks was limited.

In this paper, we used neural network (NN), decision tree (DT), and support vector machine (SVM) to classify the risk levels. The three classifiers have different characteristics. NN consists of three interconnected layers, namely, input layer, hidden layer, and output layer. Each layer contains interconnected nodes than can process the data. The interconnections are assigned weights that continue to change as the NN 'learns' the pattern from the input data. Inputs and intermediate results are passed from the input layer to the output layer to produce the final classification results [9]. Because of the structure, NN is good at learning non-linear relationships between input data and output data. SVM views data sets as vector spaces and performs classification by constructing a hyperplane that maximizes the separation in order to divide the vectors into different classes. SVM can perform either linear or non-linear classification [11]. DT can tolerate the presence of outliers and missing data, and so minimum effort is required for data preprocessing using DT. When processing categorical data with more than two levels of value, NN and SVM create dummy variables for each level of value of the related input variable, and this adds to the computational burden. In contrast, DT can derive rules directly from categorical data without creating dummy variables. However, DT cannot use continuous variables directly, and has to convert them to categorical data. The DT model adopted in this research was C5.0, an upgraded version of C4.5 developed by Quinlan [41]. Compared to C4.5, C5.0 is faster, more accurate, and less memory intensive [42]. Furthermore, C5.0 allowed multiple splits at any level of the DT.

Similar to data mining, text mining has also gained popularity as a research tool due to its ability to mine digital content available on the Internet. Text mining was used to convert free text of the phishing alerts to structural data in the form of a document-term matrix. Since a document is composed of various terms, if we used all terms as attributes, the dimensionality of the document-term matrix would be very high. We grouped similar terms together so that the dimensionality of the document-term matrix was significantly reduced following the example of prior research [16]. In fact, we found that some of the frequently occurring words had almost similar meaning, and thus it was more efficient to group such words together under a higher level concept. For example, the terms 'cash', 'refund', and 'savings' could be grouped under the concept 'money'. Usually, a dictionary which contained the linguistic and semantic relationships