Winter 12-4-2009

Analyzing The Risk and Financial Impact of Phishing Attacks Using a Knowledge Based Approach

Xi Chen

Indranil Bose

Alvin Chung Man Leung

Chenhui Guo

ANALYZING THE RISK AND FINANCIAL IMPACT OF PHISHING ATTACKS USING A KNOWLEDGE BASED APPROACH

Xi Chen[1], Indranil Bose[2], Alvin Chung Man Leung[3] and Chenhui Guo[4]
[1,4]School of Management, Zhejiang University, Hangzhou, China
[2,3]School of Business, The University of Hong Kong, Hong Kong, China
[1]chen_xi@zju.edu.cn; [2]bose@business.hku.hk; [3]alvincml@gmail.com;
[4]Julianchguo@gmail.com

Abstract

We assess the severity of phishing attacks in terms of their risk levels and the potential loss in market value to the firms. We analyze 1,030 phishing alerts released on a public database as well as financial data related to the targeted firms using a hybrid text and data mining method that predicts the severity of the attack with high accuracy. Our research identifies the important textual and financial variables that impact the severity of the attacks and determine that different antecedents influence risk level and potential financial loss associated with phishing attacks.

Key Words: Phishing, Data Mining, Financial loss, Risk, Text Mining, Variable Importance.

Introduction

Phishing is a major security threat to the online community. It is a kind of identity theft that makes use of both social engineering skills and technical subterfuge to entice the unsuspecting online consumer to give away their personal information and financial credentials [1]. Phishing caused an estimated financial loss of US $3.2 billion affecting 3.6 million people from September 2006 to August 2007, showing its tremendous financial impact [2]. Phishing attacks not only cause financial loss, but also shatter the confidence of customers in conducting e-commerce. 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 [3]. Another study also reported that the customer fear psychosis has resulted in a 20% decrease in the rate of opening of genuine emails [4].

To make customers aware of latest phishing attacks, some international organizations, such as Anti-phishing Working Group (APWG) and Millersmiles, and government statutory bodies have published phishing alerts on their respective Web sites. Apart from contextual information such as apparent sender, return email address, content of phishing email, URL of phishing server, and location of phishing location, an anti-phishing Web site Millersmiles has announced the associated risk level of phishing attacks and provided security advice to the general public. However, 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 based on past research [5]. The financial loss resulting from a phishing attack is always of great concern to security administrators, investors and consumers of an organization. In fact, both risk level and indirect financial loss are complementary measures because the two indicators may not be correlated and a high risk level of a phishing alert does not necessarily imply that the phishing attack will result in a high loss in market value [5]. Therefore, assessing the severity of phishing alerts using both these indicators helps to build a complete picture of the impact of phishing attacks.

This research has several objectives. By analyzing data related to phishing alerts using data mining techniques, we aim to identify the key characteristics of phishing attacks that determine the risk level of phishing attacks. Secondly, we predict the magnitude of loss in the market value of a firm when it is targeted by phishers. Since direct financial loss due to a phishing attack is difficult to calculate, we look into indirect financial loss in market value caused by phishing alerts.

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 [6]. Workman found that the critical success factors for some marketing strategies were applicable to phishing attacks as well [7]. 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 [8].

Among economic studies related to phishing, 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 [9]. 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 [10]. 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 [5]. 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 [11].

In the area of technical research, information security researchers have toiled to discover better countermeasures of phishing. Data mining techniques have been used to filter out phishing emails that contained fraudulent messages [12]. 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% [13]. To authenticate the URL embedded in the emails, logistic regression [14] and decision trees have also been used [15].

Text mining has also gained popularity as a research tool due to its ability to mine digital content available on the Internet. The most typical application of text mining is in document management involving tasks such as text segmentation, key words extraction, indexing, and text categorization. Wei et al. have used clustering techniques and integrated information on personal preferences for document management [16]. A hybrid methodology that combined text mining with data mining has been adopted by some researchers as well. Ma et al. used text mining to analyze company news and discover social networks among companies and utilized the discovered characteristics of the social networks to predict the revenue of the associated companies using decision trees and logistic regression [17]. Although text mining has been frequently used in a number of domains, its application in the area of information security is not so common. Wang et al. used text mining to analyze disclosures about information security incidents in financial reports and determined if they impacted the valuation of the firm [18]. We believe that text mining techniques can be used to analyze text-based phishing alerts in the same way for identification of important textual variables that characterize phishing attacks.

Prior research has demonstrated that phishing as an online crime is growing in terms of frequency of occurrences, financial loss imparted to firms and their customers, as well as technical sophistication. As there is a lack of research in the area of assessment of phishing attacks, we are motivated to construct a warning system based on a knowledge based approach. In the context of security breaches, past research has evaluated the impact of the characteristics of the attack on the financial loss generated by the security breach [11,19] but did not find any significant relationship between them.

Data Collection and Analysis

In this section we describe how we collect, prepare, and analyze phishing alerts to assess their severity and determine important antecedents that influence the classification.

Data Collection and Preparation

To determine the severity of phishing attacks, we utilized phishing alerts available from the database Millersmiles and financial data available from the financial statements of the firms. The phishing alerts data used in this research is the largest available data at the time of research and was collected from mid-2005 to mid-2008.

As phishing is primarily motivated by financial gains, corporate financial data may be relevant for the assessment of severity of phishing. Relevant financial data that was reported in the last month of the year prior to the release of the phishing alert was retrieved from The Center for Research in Security Prices. In the raw dataset, there were 168 financial variables. The authors conferred with each other and an expert in the area of finance to choose relevant financial variables that were appropriate for the context of this research. This resulted in the choice of 75 attributes related to the financial performance of a firm. Then we used the Pearson's Chi-square statistic to determine the strength of the relationship between those 75 financial variables and the target variables. The top 25 variables for the classification tasks (in terms of the Pearson's Chi-square statistic) were selected. The list of those 25 financial variables appears in Table 1. As some targets of phishing attacks did not have publicly available financial data, (e.g., Internal Revenue Service) some sample data was discarded at this stage.

Table 1. List of Financial Data

The technical sophistication of the phishing attack was measured in terms of the risk level of the attack that was determined by the information security specialists of Millersmiles. As for financial impact, an event study was conducted to determine the change in market value of the firm after the release of the phishing alert, similar to the research done by Leung and Bose [5]. First, all events related to

Variables	Mean	Std. Dev
Advertising_Expense	528.41	639.80
Assets_Total	65481.96	130713.00
Book_Value_Per_Share	24.61	30.93
Common_Equity_Tangible	14641.81	19161.09
Cost_of_Goods_Sold	14004.00	20477.22
Debt_in_Current_ Liabilities_Total	52018.47	102523.14
Earnings_Before_ Interest_and_Taxes	10021.76	11791.06
Employees	61.20	113.79
Income_Before_ Extraordinary_Items	4262.38	5304.67
Inventories_Total	7679.02	20266.69
Invested_Capital_Total	86918.49	111025.53
Liabilities_Total	453071.57	635203.23
Long_Term_Debt_Total	60586.83	86909.57
Market_Value_Total_Fiscal	47621.20	56221.96
Net_Income_Loss	4271.03	5330.99
Notes_Payable_Short_Term _Borrowings	45027.88	94363.77
Operating_Expenses_Total	20771.45	27889.17
Other_Intangibles	3281.63	6536.10
Preferred_Preference_Stock_ Capital_Total	407.25	1033.62
Price_High_Annual_Fiscal	51.19	21.34
Price_Low_Annual_Fiscal	35.90	16.60
Receivables_Total	244662.65	317712.00
Revenue_Total	31582.61	38778.60
S_P_Core_Earnings	4212.09	5058.42
Selling_General_and_ Administrative_Expense	7645.00	8144.54

private firms were removed. Then events that were affected by some confounding events such as mergers, acquisitions, dividend announcements, and changeovers were eliminated from further consideration. Then the stock return of the firm was compared with that of a market index to determine the cumulative abnormal return (CAR) of stock prices of firms due to the release of the phishing alert. We used CAR in this study because the change in the stock price of a firm is a synthesized reflection of various consequences due to phishing attacks, such as loss of clients, shrinkage in market share, and reduced confidence of consumers as well as investors. A total of 1,030 phishing alerts in our sample data had relevant CAR data and were subsequently used for classification of risk level and CAR. The CAR for these 1,030 phishing alerts ranged from -7.9% to 5.7% with an average of 0% and standard deviation of 1.3%.

Numerical Experimentation

We used a 3×3×2 experimental design in this research incorporating three sets of input data, three classifiers, and two classification tasks. The design included:

- Textual data from phishing alerts, financial data of the targeted companies, and combined textual and financial data. Text mining techniques were used on the phishing alerts to determine important semantic concepts that could act as input variables to the classifiers.
- Three classifiers – decision trees (DT), support vector machines (SVM), and neural networks (NN).
- Classification of risk level and CAR.

After the models were built, their performances were compared using top decile lift as performance metric. In addition, we also evaluated the relative importance of the different input variables for the various models. Further details about the experimental design are provided in the following sub-sections.

Textual Content Analysis Using Text Mining

Text mining was used to convert free text of the phishing alerts to structural data in the form of a document-term matrix. We grouped similar terms together so that the dimensionality of the document-term matrix was significantly reduced. 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 between words is used for grouping of concepts. We used the text mining module of the SPSS Clementine data mining suite to extract the key semantic concepts from the phishing alerts that had its own built-in dictionary. After grouping various terms under the broader semantic

concepts, a document-concept matrix was built. Each cell of the matrix represented the frequency of occurrence of the concepts within a document (i.e., a phishing alert). By performing this analysis, the natural language of phishing alerts was converted to structural data that could be used as input variables to the classification models.

Development of Classification Models for Risk Level and CAR

We first categorized phishing alerts according to the risk level assigned by Millersmiles. There were several predefined risk levels, namely, Low, Low-Medium, Medium, Medium-High, and High. For the sake of simplicity, we grouped risk levels Low and Low-Medium to form a new group 'Low' and Medium-High and High to form a new group 'High'. Next, we categorized phishing alerts according to the CAR generated by them. Positive CAR indicated that the market responded favorably to the phishing alert whereas a negative CAR indicated unfavorable market response. Although CAR is a continuous variable we categorized it into three groups, namely, positive, stable, and negative. The positive group consisted of phishing alerts that resulted in CAR greater than 3%, while the negative group consisted of phishing alerts associated with a CAR less than -3%. This method of creating groups with the choice of 3% as a threshold value was also used in prior research [18].

In the subsequent modeling phase, we classified risk level and CAR using input variables obtained from textual categories or financial data or both. NN, SVM, and DT were used in this research due to their history of superior performance in other applications related to information security [15]. The three classifiers have different characteristics. NN consists of three inter-connected 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. 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. 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.

The risk levels and the CAR for the phishing alerts were not evenly distributed. Table 2 shows the distributions of the two variables. Therefore, for classification of risk level, we oversampled the high risk and low risk instances of data but kept the medium risk instances the same so that the distribution of the three groups became 1:1:1 in the training and testing data sets. For classification of CAR, we repeated the process by oversampling the negative and positive instances while retaining the stable instances in its original form. To build the classification model, 70% of the oversampled data was used for training and 30% was used for testing. However, in the validation data sets, we retained the original distribution of data. We also used 10-fold cross validation and calculated the average accuracy of the model from the cross-validation models.

Table 2. Distributions of the risk levels and the CAR

Category	Count	Proportion
Risk Level		
High	86	8.37%
Low	23	2.24%
Medium	919	89.45
CAR		
Negative	24	2.33%
Positive	28	2.72%
Stable	978	94.95%

Results

In this section, the results obtained by applying the trained classification models on the validation data are presented. We evaluated the decile lift of the models and then identified the important variables discovered by the models for the two classification tasks.

Decile Lift

In Tables 3 and 4, we showed the lift values obtained for the two classification tasks. For classification of risk level, the models assigned likelihood scores to phishing alerts that indicated how likely it was for the phishing alerts to be high risk. The top decile lift was equal to the ratio of true high risk phishing alerts among the top 10% of phishing alerts in terms of the likelihood score of high risk divided by the ratio of high risk phishing alerts in the whole population of phishing alerts. The higher the top decile lift, the better was the model. We used lift values to compare

the model's ability to capture high risk phishing alerts. As shown in Table 3, the combined textual and financial data always performed best in terms of top decile lift up to the 7th decile. For SVM, the use of only textual data was consistently better than the use of only financial data in terms of top decile lift. For DT, the performance using textual data was not as good as that using financial data in the first decile but was consistently better up to the 6th decile and for NN the performance using textual data was better than that using financial data up to the 4th decile. The results indicated that analyzing the textual content of the phishing alerts was important for the classification of risk levels of the phishing alerts. The results also illustrated that combining textual data with financial data made the classification more accurate. Among the three classifiers, the performance of SVM was the best for the top decile. The top decile lift of the SVM classification model using hybrid textual and financial data as inputs was 6.40. This meant that this particular model was 6.4 times more likely to capture true high risk phishing alerts than random selection.

Table 3. Lift Values for Classification of Risk Level

Deciles	Combined DT	Text. DT	Fin. DT	Combined SVM	Text. SVM	Fin. SVM	Combined NN	Text. NN	Fin. NN
1	5.26	4.07	4.09	6.40	4.40	2.44	4.77	4.19	2.75
2	4.35	3.17	2.98	3.95	3.52	1.82	3.49	3.07	2.51
3	3.02	2.50	2.40	2.91	2.54	1.98	2.55	2.36	1.97
4	2.33	2.02	1.99	2.31	2.00	1.76	2.06	1.80	1.69
5	1.93	1.73	1.70	1.93	1.75	1.52	1.67	1.51	1.58
6	1.61	1.55	1.50	1.61	1.59	1.37	1.47	1.38	1.46
7	1.38	1.35	1.38	1.43	1.38	1.24	1.28	1.24	1.33
8	1.21	1.24	1.25	1.25	1.21	1.16	1.13	1.15	1.18
9	1.11	1.11	1.11	1.11	1.11	1.09	1.06	1.06	1.07
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4. Lift Values for Classification of CAR

Deciles	Combined DT	Text. DT	Fin. DT	Combined SVM	Text. SVM	Fin. SVM	Combined NN	Text. NN	Fin. NN
1	5.91	4.76	2.90	8.52	7.72	2.86	7.62	7.14	5.63
2	3.10	2.75	2.86	4.76	5.00	4.03	4.76	4.29	4.07
3	2.75	2.70	2.68	3.17	3.33	3.17	3.17	3.17	3.17
4	2.38	2.20	2.50	2.50	2.50	2.50	2.38	2.38	2.50
5	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
6	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67
7	1.43	1.43	1.43	1.43	1.43	1.43	1.43	1.43	1.43
8	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25
9	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

The lift values obtained for the classification of CAR are shown in Table 4. The results shown are consistent with those in Table 3. Again, the combined textual and financial data performed best in terms of lift in most cases and the use of only textual data was better than the use of only financial data for SVM but not for DT and NN. Table 4 again illustrated the importance of combining textual data with financial data for the purpose of classification. As in the case of risk level classification, the SVM model using combined textual and financial data as inputs obtained the highest lift value of 8.52.

Comparison of Important Variables

In order to understand the antecedents that governed the classification of risk level and CAR of phishing alerts, we calculated the importance of all input variables. As the combined textual and hybrid data gave rise to good decile lift in general, we listed the top five most important textual variables and the top five most important financial variables identified for each of the classification tasks using the three classifiers with this data as input. The variables are listed in order of their importance in Tables 5 and 6 where the column 'Identifying Classifier(s)' showed which classifiers rated the variable as a top five variable.

We can observe that there was no general agreement among the classifiers about the most important textual category. For classification of risk level, 'update' was identified as an important textual category by all three classifiers. This implied that phishing attacks with messages requesting recipients to update their personal information were of high risk level. For classification of CAR, 'consumers' was identified as an important textual category by all

three classifiers whereas 'information' and 'writing' were identified as top five categories by two of the three classifiers. When phishers pretended to be authenticated service providers and requested their customers to reveal personal information, then such attacks became likely to cause financial loss to the customers, hurt brand reputation, and affect present and future revenues of the company.

Table 5. Textual Concepts Listed in Order of Importance with Identifying Classifiers

Risk Level	Identifying Classifier(s)	CAR	Identifying Classifier(s)
Update	DT, SVM, NN	Consumers	DT, SVM, NN
Security	DT, SVM	*Information*	DT, SVM
Email	DT, SVM	Writing	SVM, NN
Bank account	DT, NN	eBay	DT
Bank	SVM, NN	*Confirmation*	DT
Confirmation	DT	Warning	DT
Account	SVM	Person	SVM
Information	NN	*Account*	SVM
Computers	NN	Work	NN
		Computers	NN
		Assets	NN

*****Textual concepts common to both classifications are shown in italic

In Table 6, the top five most important financial variables identified by the three classifiers are listed. There were no common financial variables for classification of risk level and CAR. This showed that the underlying financial variables determining the two measures of severity of phishing attacks were significantly different. For classification of risk level, total of inventories was identified as an important financial variable by all three classifiers whereas other intangibles and advertising expense was identified as a top five financial variable by two out of three classifiers. These financial variables indicated the preference of phishers towards launching attacks on large firms. High total inventories and intangibles is a hallmark of a large firm and high advertising expense identified a company that had greater media exposure. This meant that large companies were preferred targets for high risk phishing attacks because they had a strong customer base and their customers were likely to be misled by fake emails due to their inherent trust on these companies. For classification of CAR, number of employees, total invested capital, and total liabilities were identified as top five financial variables by two out of three classifiers. Again, the number of employees and total invested capital indirectly hinted at the large size of the firm. It was interesting to note that firms that already had high total liabilities were at greater risk of being penalized by investors when phishing attacks took place and shook the confidence of the investors.

Table 6. Financial Variables Listed in Order of Importance with Identifying Classifiers

Risk level	Identifying Classifier(s)	CAR	Identifying Classifier(s)
Inventories_Total	DT, SVM, NN	Employees	DT, SVM
Other_Intangibles	DT, NN	Invested_Capital_Total	SVM, NN
Advertising_Expense	SVM, NN	Liabilities_Total	SVM, NN
Price_High_Annual_Fiscal	DT	Receivables_Total	DT
Operating_Expenses_Total	DT	Net_Income_Loss	DT
Income_Before_Extraordinary_Items	DT	Price_Low_Annual_Fiscal	DT
S_P_Core_Earnings	SVM	Long_Term_Debt_Total	DT
Preferred_Preference_Stock_Capital_Total	SVM	Assets_Total	SVM
Market_Value_Total_Fiscal	SVM	Book_Value_Per_Share	SVM
Common_Equity_Tangible	NN	Notes_Payable_Short_Term_Bors.	NN
Earnings_Before_Interest_and_Taxes	NN	Debt_in_Current_Liabilities_Total	NN
		Cost_of_Goods_Sold	NN

Discussion

Keeping in mind that it is important to evaluate the technical sophistication as well as the potential financial impact of phishing attacks, we conducted this research and developed a mechanism to predict the severity of phishing alerts in terms of risk level and potential loss in market share indicated by CAR of stock prices. From the list of top five most important input variables generated using the three classifiers, we found that the overlap for the two types of classifications was consistently low and this implied that risk level of a phishing alert was not indicative of the CAR generated by it. The loss in market value of the targeted firm could be added with the information of the risk level by anti-phishing organizations to give a complete picture of the

impact of a phishing attack. Furthermore, our research results indicated that assessment based on data that consisted of important textual categories discovered from the text of phishing alerts as well as financial data of the targeted companies, outperformed assessment based on any of the above data items alone. Information security specialists usually assess risk level of phishing incidents based on the textual description of phishing alerts. Our results indicated that for assessing severity it was important to consider the financial condition of the targeted company as well.

From an academic perspective, our research made an important contribution in terms of application of a hybrid text and data mining method for solving a problem in the area of information security. Text mining was used in the first stage to extract key semantic concepts from the textual content of the phishing alerts. The performance of the classifiers in terms of top decile lift showed that the hybrid text and data mining model was successful in classifying different levels of risks and different types of financial impact caused by phishing attacks. The results were more or less consistent for the three different classifiers and indicated that a hybrid data mining model was able to generate consistent results of high accuracy. Data mining techniques have been frequently used in the past to filter out phishing emails or thwart access to phishing Web sites and our research showed that the same techniques could be used to assess severity of phishing attacks effectively.

From a managerial perspective, our study paved the way for automating the assessment of severity of phishing attacks. As there are an increasing number of phishing incidents that are reported around the world every day, manual assessment of such incidents could be time consuming as well as inaccurate due to the subjective bias of the evaluator. The method proposed in this paper automated the assessment of phishing incidents using past data and provided a richer assessment of such incidents than what is currently being done by the anti-phishing organizations. We hope that the findings of this study can encourage anti-phishing organizations to adopt our proposed method to predict the risk level as well as potential financial impact of a phishing alert as soon as it is reported on their Web site.

Conclusion

In this research, we adopted a hybrid text and data mining model that used text mining to discover important semantic categories from the textual content of the phishing alerts and combined those discovered categories with financial data of the targeted companies to come up with classification of risk level of the attack and the loss in market value of the firm that it was likely to cause. The performance of the hybrid model was quite superior in terms of top decile lift and demonstrated the need to consider textual data as well as financial data for making prediction about the severity of the phishing alert. Furthermore, our results showed that risk level and CAR were fundamentally different from each other as we discovered that different textual and financial factors impacted them. This implied that it was important to evaluate both for fully assessing phishing alerts – a practice we recommend that all anti-phishing organizations should adopt in future to make their members more knowledgeable about the severity of phishing attacks.

References

[1] APWG, Phishing Activity Trends Report Second Half 2008, Anti-Phishing Working Group, 2009, (http://www.antiphishing.org/reports/apwg_report_H2_2008.pdf) pp. 1-12.

[2] Powell T., Ounce of ID Theft Protection Worth More than Agony of Restoring Good Name, My WestTexas.com, 2008, (http://www.mywesttexas.com/articles/2008/05/26/news/opinion/columns/trish_powell/bbb_5_23.txt).

[3] Ensor B., A. Giordanelli, M.d. Lussanet and T.v. Tongeren, Many Online Banking Users Use Few Features, Forrester Research, 2007, pp. 1-6.

[4] Brandt A., Phishing Anxiety May Make You Miss Messages, PC World, 23(10), 2005, p. 34.

[5] Leung A.C.M. and Bose I., Indirect Financial Loss of Phishing to Global Market, Proceedings of the Twenty-Ninth International Conference on Information Systems, Association for Information Systems, Paris, France, 2008, pp. 1-15.

[6] Jagatic T., Johnson N., Jakobsson M. and Menczer F., Social Phishing, Communications of the ACM, 50(10), 2006, pp. 1-10.

[7] Workman M., Wisecrackers: A Theory-grounded Investigation of Phishing and Pretext Social Engineering Threats to Information Security, Journal of the American Society for Information Science and Technology, 59(4), 2008, pp. 662-674.

[8] Liao Q. and Luo X., The Phishing Hook: Issues and Reality, Journal of Internet Banking and Commerce, 9(3), 2004, p. 1.

[9] Singh N.P., Online Frauds in Banks with Phishing, Journal of Internet Banking and Commerce, 12 (2), 2007, pp. 1-27.

[10] Goth G., Phishing Attacks Rising, but Dollar Losses Down, IEEE Security & Privacy Magazine, 3(1), 2005, p. 8.

[11] Kannan K., Rees J. and Sridhar S., Market Reactions to Information Security Breach Announcements: An Empirical Analysis, International Journal of Electronic Commerce, 12(1), 2007, pp. 69-91.

[12] Airoldi E. and Malin B., Data Mining Challenges for Electronic Safety: The Case of Fraudulent Intent Detection in E-Mails, Proceedings of the

Workshop on Privacy and Security Aspects of Data Mining 2004, IEEE Computer Society, Brighton, UK, 2004, pp. 57-66.

[13] Zhang J., Du Z-H.and Liu W., A Behavior-Based Detection Approach to Mass-Mailing Host, Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, China, 2007, pp. 2140-2144

[14] Garera S., Provos N., Chew M. and Rubin A.D., A Framework for Detection and Measurement of Phishing Attacks, Proceedings of the 2007 ACM Workshop on Recurring Malcode, Alexandria, VA, USA, 2007, pp. 1-8.

[15] Ludl C., McAllister S., Kirda E. and Kruegel C., On the Effectiveness of Techniques to Detect Phishing Sites, In: B.M. Hämmerli and R. Sommer (eds.), Detection of Intrusions and Malware, and Vulnerability Assessment: Fourth International Conference, Springer, Berlin, Germany, 2007, pp. 20-39.

[16] Wei C-P., Chiang R.H.L. and Wu C.-C., Accommodating Individual Preferences in the Categorization of Documents: A Personalized Clustering Approach, Journal of Management Information Systems, 23(2), 2006, pp. 173-201.

[17] Ma Z., Sheng O.R.L. and Pant G., Discovering Company Revenue Relations from News: A Network Approach, Decision Support Systems (forthcoming).

[18] Wang T-W., Rees J., and Kannan K., Reading the Disclosures with New Eyes: Bridging the Gap between Information Security Disclosures and Incidents (February 1, 2008). Available at SSRN: http://ssrn.com/abstract=1083992

[19] Andoh-Baidoo F.K.and Osei-Bryson K-M., Exploring the Characteristics of Internet Security Breaches that Impact the Market Value of Breached Firms, Expert Systems with Applications, 32 (3), 2007, pp. 703-725.