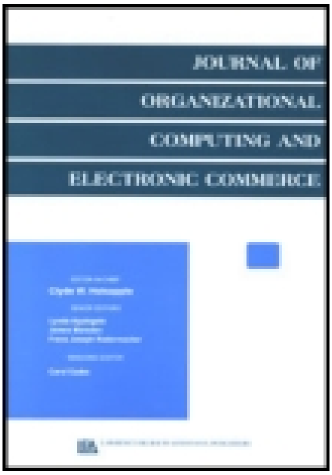

Journal of Organizational Computing and Electronic Commerce

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hoce20

The Textual Contents of Media Reports of Information Security Breaches and Profitable Short-Term Investment Opportunities

Tawei Wang [a] , Jackie Rees Ulmer [b] & Karthik Kannan [b]

[a] School of Accountancy, Shidler College of Business, University of Hawaii at Manoa , Honolulu , HI , USA

[b] Krannert Graduate School of Management, Center for Education and Research in Information Assurance and Security (CERIAS), Purdue University , West Lafayette , IN , USA

Accepted author version posted online: 30 May 2013.Published online: 18 Jul 2013.

To cite this article: Tawei Wang , Jackie Rees Ulmer & Karthik Kannan (2013) The Textual Contents of Media Reports of Information Security Breaches and Profitable Short-Term Investment Opportunities, Journal of Organizational Computing and Electronic Commerce, 23:3, 200-223, DOI: 10.1080/10919392.2013.807712

To link to this article: http://dx.doi.org/10.1080/10919392.2013.807712

Journal of Organizational Computing and Electronic Commerce, 23: 200–223, 2013
Copyright © Taylor & Francis Group, LLC
ISSN: 1091-9392 print / 1532-7744 online
DOI: 10.1080/10919392.2013.807712

Taylor & Francis
Taylor & Francis Group

THE TEXTUAL CONTENTS OF MEDIA REPORTS OF INFORMATION SECURITY BREACHES AND PROFITABLE SHORT-TERM INVESTMENT OPPORTUNITIES

Tawei Wang,[1] Jackie Rees Ulmer,[2] and Karthik Kannan[2]

[1]*School of Accountancy, Shidler College of Business, University of Hawaii at Manoa, Honolulu, HI, USA*
[2]*Krannert Graduate School of Management, Center for Education and Research in Information Assurance and Security (CERIAS), Purdue University, West Lafayette, IN, USA*

Information security-related incidents continue to make headlines. Interestingly, researchers have found mixed results when attempting to associate reports of information security breaches with changes in the affected firm's stock price. This research delves further into this puzzle by investigating the association between the textual contents of information security breach media reports and the stock price, as well as the trading volume reactions of the affected firm(s) around the breach announcement day. Our findings suggest that when the textual contents of breach reports provide more detailed information regarding the incidents, a more consistent belief is formed by the market about the negative impact of the reported security incident on the firm's business value. However, when there is a lack of specific information regarding the reported breach, the market does not seem to reach consensus on the impact of reported security incidents. We further demonstrate that different perceptions exist among general and sophisticated investors regarding the impact of reported information security incidents on a firm's future performance as demonstrated by changes in trading volume. By exploiting the different perceptions among investors, we form a trading strategy to demonstrate that, on average, one can make about 300% annual profit around the breach announcement day.

Keywords: *information security; breach announcements; text mining; decision tree; sophisticated investors*

1. INTRODUCTION

Information security-related incidents often lead to a disruption of business and cause significant losses (CSI/FBI 2007). Recently Sony's PlayStation Network, an online community and gaming service, was offline for nearly one month after an attack by hackers. Early estimates of Sony's losses from the attack run as high as $1B (Wakabayashi 2011). Given the potential threats posed by information security incidents to a firm's operations, as well as legal costs and negative impact to the firm's reputation, it is important for investors

Address correspondence to Tawei Wang, 2404 Maile Way, BusAd E602C, Honolulu, HI 96822, USA. E-mail: twwang@hawaii.edu

to understand how information security breaches could affect a firm’s future performance so that they can make well-informed investment decisions.

A primary information source for all investors is media reports on information security breaches, or what we term “breach announcements.” Breach announcements are typically found in major media outlets, blog posts, etc. Different terminology and other characteristics of specific media reports could affect investor perceptions of the impact of the breach. For example, some articles might contain terms that are more specific, whereas other articles (and their corresponding authors) might use more ambiguous language, leading to different reactions among investors. Furthermore, different types of information security incidents could have different implications for the affected firm. Information security incidents are often classified as affecting the confidentiality of information, the integrity of information, or the availability of information. Availability-type incidents or attacks could cause temporary revenue losses for a firm, whereas confidentiality-type incidents could result in legal actions, depending on the type of information affected as well as the legal jurisdiction of the involved parties.

Typically the only information sources available to general investors (or so-called unsophisticated investors) are those major media reports. So-called sophisticated investors, such as analysts and investment institutions, are the investors having firm-specific knowledge about the firm’s operations, more information sources, and superior capabilities for processing information (e.g., Bhushan 1989; Francis, Hanna, and Philbrick 1997; Lakonishok, Shleifer, and Vishny 1992; Roulstone 2003). Accordingly, sophisticated investors might be able to assess the impact of security incidents on a firm’s future performance more accurately than unsophisticated investors.

All investors observe the corresponding stock price, as well as trading volume reactions, in the immediate aftermath of a breach announcement. The stock price and the trading volume reactions to breach announcements provide both the aggregate reaction of the market and investors’ individual reactions to security incidents (e.g., Bamber and Cheon 1995). The stock price reaction should provide information on the aggregate market reaction to an incident, for example, by moving stock prices lower by an amount corresponding to the perceived impact of the breach. The trading volume behavior should demonstrate whether different market participants have conflicting assessments as to the impact of information security incidents on a firm’s future performance. For example, if most investors share similar beliefs as to the impact of a security breach on a firm’s future performance, there might be a negative change in stock price for the firm and a change in trading volume. If investors have very divergent beliefs, the trading volume might be higher than normal, but the share prices might be relatively unchanged, indicating that the investors with negative beliefs are, in effect, “cancelled out” by those with positive or unchanged beliefs in the future performance of the firm.

Therefore, if sophisticated investors have different or additional information about a firm that has been reported as breached (e.g., information indicating how prepared/unprepared a firm is to respond to an attack), they may react differently than the overall market. Could general investors take advantage of this information and trade by considering this potential for difference in information and subsequent beliefs?

In this study, we address the following two questions. First, do different information security breach announcements lead to different investors’ assessments of the impact of security incidents on a firm’s future performance? Specifically, do certain characteristics within the textual contents of the breach announcements result in a consistently negative belief of the impact of security incidents, while other characteristics do not? Second, by taking into account the sophisticated investors' reactions to breach announcements, are there any profitable short-term investment opportunities around the breach announcement day for general investors?

In order to approach our research questions, we use text mining techniques to explore the characteristics within news articles reporting information security breaches. The characteristics are later associated with the corresponding stock price and trading volume reactions by using a decision tree classification model. The classification results are compared to the sophisticated investors' reactions to breach announcements to show the possibility of profitable short-term investment opportunities after the breach announcement. Our findings suggest that market participants could reevaluate a firm's future uncertainties regarding information security from the sophisticated investors' perspective and the textual contents of the news articles about security breaches. In addition, although managers need to know the impact of security breach announcements on a firm's business value from the investors' perspectives, the temporary drop of business value may not be a good indicator of the impact of security breaches on business value, which, in turn, cannot be incorporated when forming information security investment/deployment decisions. Also, firms could focus more on conveying the breached information to the public, which might reduce the magnitude of the temporary drop of the firm's stock price around the breach announcement day.

The remainder of the article is organized as follows. We review related literature on information security, trading volume behavior, and analyst forecasts in Section 2. The theoretical framework and our data collection process are presented in Section 3. We text mine the breach announcements articles and investigate the association between the contents of the articles and the price and trading volume reaction in Section 4. In Section 5, we examine sophisticated investors' reactions to security breach announcements and demonstrate our trading strategy for profitable short-term investment opportunities. Last, we conclude with discussion, limitations, and possible future research avenues in Section 6.

2. LITERATURE REVIEW

There are three major streams of literature that are directly related to our study. The first and the second streams of literature are related to information security and trading volume behavior corresponding to information announcements. The third stream of literature is about analyst forecasts.

2.1. Information Security

Studies have investigated information security–related issues from several perspectives, such as information security policies (e.g., Siponen 2006; Siponen and Iivari 2006; Straub 1990), information security investments (e.g., Gal-Or and Ghose 2005; Gordon and Loeb 2002; Gordon, Loeb, and Lucyshyn 2003; Schechter and Smith 2003) and the association between the board, as well as the top management team and security management (Kwon, Rees, and Wang 2013; Wang and Hsu 2010a, 2010b). However, studies that are directly related to our article are about the impact of information security breaches on a firm's performance and uncertainty. For example, Glover, Liddle, and Prawitt (2001) discussed the impact of information security breaches on business operations, including physical and intangible impacts. Also, various papers have investigated the association between security breach announcements and a firm's business value. Some results show that there exists significant negative impact (e.g., Aquisti, Friedman, and Telang 2006; Cavusoglu, Mishra, and Raghunathan 2004; Ettredge and Richardson 2003; Garg, Curtis, and Halper 2003) while others do not find such impact (e.g., Campbell et al. 2003; Hovav and D'Arcy 2003; Kannan, Rees, and Sridhar 2007). The inconclusive results of the impact of security breaches on a firm's future performance (or business value) from the previously mentioned studies point out the need to explore in more detail the investors' reactions to security incidents and information asymmetry among investors. Moreover, because sophisticated investors have more information sources and a more comprehensive understanding of the firms, the analysts' perspective on security breach announcements could help us better understand the impact of security incidents. Previous research assumes that market participants are homogenous in their information processing capabilities, so by examining the responses of different types of market participants, we hope to shed more light on this issue.

2.2. Trading Volume

The discussion of trading volume can be traced back to Beaver (1968) who shows that earnings announcements generate not only abnormal price changes, but also high trading volume. According to the literature, the stock price change reflects the change in the market's average beliefs in aggregate, while the trading volume behavior is the sum of all individual investors' trades (e.g., Bamber 1987; Bamber and Cheon 1995; Kim and Verrecchia 1991). That is, the trading volume behavior reveals counterbalanced beliefs among individual investors (e.g., Bamber and Cheon 1995). Accordingly, the association between the inconsistency of beliefs and trading volume demonstrates that a subset of investors have an advantage in processing the information or different beliefs regarding the information announcements (e.g., Bamber, Barron, and Stober 1997; Bhattacharya 2001; Easley and O'Hara 1987; Hasbrouck 1988, 1991; Kim and Verrecchia 1991; Morse 1981). For example, the analytical model in Kim and Verrecchia (1991) shows that the trading volume behavior results from the differing quality of information acquired and initial beliefs among investors. In our research, we apply this concept in the context of announcements of information security incidents in the major media and investigate the different reactions among investors based on their different information processing capabilities. Furthermore, it has been previously investigated whether different price and volume reactions are associated with different earnings announcement characteristics, such as the standard deviation of analyst forecasts and the market value of the firm (Bamber and Cheon 1995). In this article, we similarly investigate whether different price and volume reactions are associated with various textual characteristics within the news articles about security incidents.

2.3. Analyst Forecast

Analysts collect information about a firm from various sources, and provide information such as transaction recommendations and the prospects of the firm to certain market participants in a timely manner (e.g., Bhushan 1989; Francis et al. 1997; Lev and Thiagarajan 1993; Roulstone 2003). In the literature, the role played by analysts in the market can be used as proxies for informed traders as well as signals of information asymmetry because of their superior information processing capabilities and detailed communication with firms (e.g., Core 2001; Francis, Schipper, and Vincent 2002; Roulstone 2003). In our research, the analysts' superior capabilities of processing information and their understanding of the firm are used in the context of interpreting information security breaches. In particular, given the analysts' capabilities and their understanding of the firm, we argue that unsophisticated investors could make better decisions by further considering sophisticated investors' reactions to security breaches.

The number of analysts following a firm can be determined by several firm characteristics, such as firm size and return variability (Bhushan 1989). The number of analysts following the firm can also be used as a proxy for the amount of publicly available information (e.g., Atiase and Bamber 1994; Roulstone 2003). Many other studies also focus on the relationship between analyst following and the valuation of a firm (e.g., Lang and Lundholm 1993), market liquidity (e.g., Roulstone 2003), and analysts' communication with firms (e.g., Francis et al. 1997).

The analyst forecasts have also been widely investigated in terms of how analysts formulate their expectations about firms' earnings, how to improve the forecasts, and the determinants of analyst research (e.g., Barth, Kasznik, and McNichols 2001; Brown 1993; Elgers and Murray 1992; Frankel, Kothari, and Weber 2006; Kross, Ro, and Schroeder 1990; Stickel 1990). Analyst forecasts are also commonly used as a reference point when calculating earnings surprises (e.g., Ayers, Jiang, and Yeung 2006; Barron, Byard, and Yu 2008; Kasznik and Lev 1995) and when investigating whether firms attempt to manipulate their earnings (e.g., Beneish 2001; Degeorge, Patel, and Zeckhauser 1999; Matsumoto 2002; McNichols 2000). Therefore, analyst forecasts can be a good proxy and reference point of a firm's future performance. Accordingly, in this article, analyst forecasts serve as the reference point for the impact of security incidents on a firm's future performance from the sophisticated investors' perspective.

3. THEORY AND DATA COLLECTION

In this section, we first describe the rational expectation model, the theory used in this research. Then we present the data collection processes. The data collected serve as the input for the classification model in Section 4 and are also used to investigate the sophisticated investors' reactions as well as the possible profitable trading strategy in Section 5.

3.1. Rational Expectation Model

This article draws on the rational expectation model as our theoretical model. Rational expectation models describe the investment behavior of investors and how stock price incorporates and reveals information to investors. These models are commonly used to understand both the stock price and the trading volume reactions to public disclosures of information (e.g., Karpoff 1986; Kim and Verrecchia 1991, 1994, 1997). The main concept of rational expectation models applied in our article is as follows (see the papers cited previously for the mathematical models and a detailed description). In the rational expectation model, each investor has his or her own initial belief about the firm's value before the public announcement. The public announcement changes his or her beliefs so investors trade again. Given each investor is different from his or her initial belief, and based on how good the information regarding the public announcement is, investors respond to the announcement differently.

Our arguments build on prior literature regarding market reactions to public announcements, in order to show why the market reacts to breach announcements.

Specifically, the news article (i.e., the breach announcement) is the public announcement that possibly changes investors' assessments about the impact of information security events on a firm's valuation. Based on prior studies regarding security breach announcements and business value (e.g., Campbell et al. 2003; Cavusoglu et al. 2004), the breach announcement may change the market's expectation about a firm's future cash flows. For example, the breached firm may suffer from reputation loss, which results in a reduction in revenues, a need to settle lawsuits, compensate customers, etc. These factors would reduce a firm's future cash flows. In addition, lawsuits and reputation loss may further result in an increase in the cost of capital, also reducing a firm's business value. As the breach announcement is the main information source of the market for understanding the breach, and the textual contents of the breach announcement reflect the nature of the breach, we believe that the textual contents of the breach announcement should be associated with market reactions. Accordingly, in this article, we explore how the textual contents of the media reports of security breaches are related to the market reactions. General investors may make their investment decisions based on these public announcements and the associated price and volume reactions to the announcements. In contrast, sophisticated investors have a better understanding of the firm's operations (different initial belief) and have superior capabilities for processing information (better information) than general investors. The response to security breach reports by sophisticated investors could be different than that of general investors, which could be useful for general investors when making investment decisions around the breach announcement day. Therefore, as discussed in the Introduction, we would like to understand how the textual contents of the news articles regarding the breach affect the price and volume reactions by further considering sophisticated investors' reactions to security breaches in order to help general investors make better decisions.

3.2. Sample

To approach our research questions, we searched for news articles between 1997 and 2008 about information security breaches in the major news media, such as *the Wall Street Journal*, *USA Today*, *the Washington Post*, and *the New York Times* in the *Factiva* database. We also search on *CNet*, *ZDNet* and *Yahoo! Finance*. The keywords used in our search are: (1) security breach, (2) hacker, (3) cyber-attack, (4) virus or worm, (5) computer break-in, (6) computer attack, (7) computer security, (8) network intrusion, (9) data theft, (10) identity theft, (11) phishing, (12) cyber fraud, and (13) denial of service. These keywords are similar to those used in prior studies (e.g., Campbell et al. 2003; Garg et al. 2003; Kannan et al. 2007; Kwon et al. 2013; Wang, Kannan, and Rees 2013). We only included news articles about publicly traded firms with specific event dates, after ruling out the observations with confounding events, such as earnings announcements and mergers and acquisitions. For the following analyses, we excluded consecutive-attack observations except the first day, such as the series of denial-of-service (DoS) attacks in 2000, and the observations without trading data or analyst forecast data. The resulting sample size was 89 firm-events.[1]

[1] We consider the following criteria for our sample selection. First, it must be related to publicly traded firms. We exclude government agencies, private organizations, schools, etc. from our analyses. Second, the breach announcement must be from a national media source, such as the W*all Street Journal, USA Today, Washington Post*, and the *New York Times*. We do not consider other sources because we investigate the market reactions to media reports. It is hard to argue that the market is affected by a report released by local media outlets. Third, we only consider the first event in our sample period. Last, the observations are not included if there exist confounding events. Based on the above criteria, only 89 observations remain in our sample.

Table 1 Descriptive statistics.

Panel A. Firm characteristics

	Mean	Std. dev.	Median
Total assets	89,985.12	258,469.93	6,366.95
Debt/Total assets	0.60	0.31	0.58
EPS	1.05	2.57	1.04
Institutional ownership	0.57	0.26	0.64
ROA	0.03	0.16	0.05
Tangible ratio	0.43	0.29	0.39

Panel B. Industry breakdown

2-digit SIC code	Description	Percentage
48	Communication	9.8%
60	Depository institutions	8.2%
73	Business services	26.2%
Other 18 industries		55.8%

Panel C. Year breakdown

Year	# of Observations	Percentage
1997	2	2.25%
1998	2	2.25%
1999	16	17.98%
2000	16	17.98%
2001	10	11.24%
2002	6	6.74%
2003	9	10.11%
2004	8	8.99%
2005	6	6.74%
2006	7	7.87%
2007	3	3.37%
2008	4	4.49%
Total	89	100%

We retained the content of the news articles for our analyses in Section 4. The descriptive statistics, such as firm characteristics, industry, and year breakdowns, are given in Table 1.

3.3. Price and Volume Reactions

In addition to the news articles collected previously, we investigated the price and trading volume reactions to breach announcements as the other inputs to our classification model in Section 4. We considered both the stock price and the trading volume behavior because these two measures provide both the aggregate and individual difference information as discussed in the literature review. We used the commonly adopted approach in the literature to calculate the stock price reactions, which are provided in detail in Appendix A. The results show that the average stock price reaction to security incident reports in our sample is -0.15% ($p < 0.10$) in the window $(-1, +1)$, where -1 $(+1)$ denote one day before (after) the breach announcement date. That is, on average, the stock price reacts negatively to the breach announcement. The parametric test statistic is -1.956 ($p < 0.05$), while the nonparametric test statistic is -1.964 ($p < 0.05$).

For the trading volume behavior, we consider two measures. The first measure considers the trading volume changes over time around the breach announcement date (i.e., a three-day window as in the stock price reaction case) by controlling for the market effect as detailed in Appendix A. The significant increase ($p < 0.05$) in trading volume at the breach announcement day demonstrates that the breach announcements indeed induce more trading volume. Similarly, the second measure controlling for firm-specific effects also shows that, on average, the trading volume is 13.62% more ($p < 0.05$) than the usual trading volume after a breach announcement. The parametric test statistic is -1.850 ($p < 0.05$.) and the nonparametric test statistic is -1.677 ($p < 0.05$). As discussed in the literature review, the significantly increased trading volume behavior shows that investors have different beliefs regarding the impact of security breaches on a firm's future performance, and some investors are able to better process the information regarding the impact of security breaches. Accordingly, general investors could take advantage of this difference and have profitable investment opportunities which will be investigated in Section 5.

4. CLASSIFICATION MODEL

4.1. Decision Tree Classification Model

In this section, we first apply a text-mining algorithm to the textual contents of the news articles about security breaches (the breach announcements). Using these results, we then associate the characteristics within such breach announcements with the price and volume reactions by using a decision tree classification model. Text mining has been widely used in different contexts, such as to classify news stories, summarize banking telexes, detect fraud, and to improve customer support (e.g., Cecchini, Aytug, Koehler, and Pathak 2010; Fan, Wallace, Rich, and Zhang 2006; Han et al. 2002; Masand, Linoff, and Waltz 1992; Young and Hayes 1985). In our context, we apply text mining techniques to the breach announcements to investigate how this publicly available information regarding security breaches is associated with the stock price and trading volume reactions. As we shall show, the tool we use for the association is a decision tree model. We chose a decision tree model first because of its inherent transparency and interpretability. Decision tree models help users follow the path of the tree and understand the classification rules step-by-step (e.g., Baesens, Setiono, Mues, and Vanthienen 2003; Brandãn, Dyer, and Hahn 2005; Kim et al. 2001; Zhang and Zhu 2006; Zhou and Jiang 2004). Second, the literature has shown that decision tree models have been used in different small sample contexts and performs reasonably well compared to other classification models (e.g., Goto et al. 2008; Masand et al. 1992; Sordo and Zeng 2005). Because this study also has a small sample size, decision tree models should also perform reasonably well. We also tested other classification models, such as neural networks, and obtained similar results.

We use a three-step process (as provided in Figure 1) to build the decision tree model, which is presented in detail in the following paragraphs. First, recall that from our data collection process, we collected 89 breach announcements reporting security incidents at publicly traded U.S. firms. These 89 breach announcements were input into SAS Text

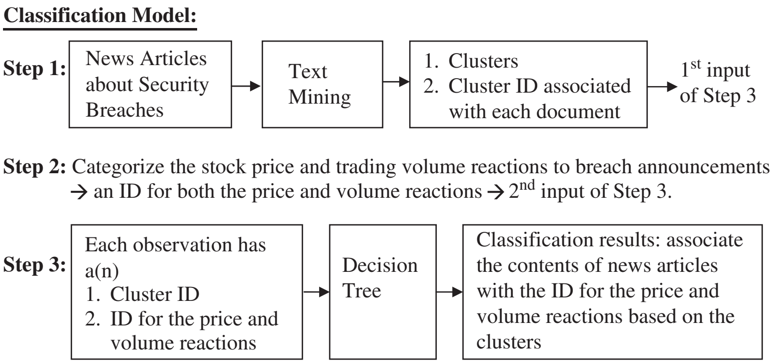
Classification Model:
Step 1:
News Articles about Security Breaches
Text Mining
1. Clusters
2. Cluster ID associated with each document
1st input of Step 3
Step 2: Categorize the stock price and trading volume reactions to breach announcements → an ID for both the price and volume reactions → 2nd input of Step 3.
Step 3:
Each observation has a(n)
1. Cluster ID
2. ID for the price and volume reactions
Decision Tree
Classification results: associate the contents of news articles with the ID for the price and volume reactions based on the clusters

Figure 1 Building process of the Classification Model.

Table 2 The Change of accuracy rate given the number of clusters.

Number of clusters for breach reports	Accuracy rate based on decision tree	Accuracy rate based on neural network	Accuracy rate based on logistic regression
2	64.3%	65.2%	58.8%
3	67.2%	70.9%	64.1%
4	71.4%	74.3%	69.5%
5	65.6%	69.0%	68.2%
6	61.7%	68.5%	66.5%

Miner and then categorized into clusters.[2] Please see Appendix B for a detailed description of the process. Here, we determined the number of clusters to be four by experimentally varying the number of clusters until the error rate of the decision tree model (discussed next) was the smallest (about 0.39, see Table 2 for the accuracy information under different models and different number of clusters) (e.g., Smyth 2000; Still and Bialek 2004; Tibshirani, Walther, and Hastie 2001). However, because three of the four clusters did not converge when we further explore the textual characteristics of the contents (discussed next), we chose to group the four clusters into two "super-clusters" (Cluster A and Cluster B) when we present our textual analysis. The output was a cluster ID (from one of the two clusters) associated with each breach announcement. This cluster ID (again, Cluster A or Cluster B) will be the classifier in our decision tree model.

The second step is to associate the stock price and trading volume reactions with the breach announcement. We used the standard K-means cluster analysis to classify our observations based on both the stock price and trading volume reactions around the publication date of the breach announcement. We observed that the stock price and trading volume reactions only converged into two or three major clusters. That is, the clustering process cannot converge into other number of clusters. When we ran the model resulting in three

[2] We choose to form clusters instead of using words (or phrases) directly to predict market reactions because of the following reasons. We have more than 100,000 words/phrases in our dataset. Although it is not a big dataset, when we use all these words/phrases to predict the market reactions, it becomes very difficult to understand the results, not to mention the low predictability. That is, we are not able to summarize 50,000 words/phrases, for instance, to explain why the textual contents may be associated with different market reactions.

clusters, one of the three clusters had only one observation, and the breach announcements in other two clusters were the same when we ran the model resulting in only two clusters. Therefore, we used two clusters when presenting the results, labeled as Reaction Group 1 and Reaction Group 2.

Reaction Group 1 has 63 observations with an average (standard deviation) of stock price reaction of -0.002 (0.0317) lower than and trading volume of 78.470% (17.539%) than usual. Reaction Group 2 has 26 observations with an average (standard deviation) of stock price reaction of 0.021 (0.0502) higher than and trading volume of 145.008% (31.816%) higher than usual. Further analysis shows that the breach announcements in Reaction Group 2 result in a significant higher trading volume ($p < 0.05$), but a slightly positive stock price reaction, which is not significantly different from zero. However, the breach announcements in Reaction Group 1 result in a significant negative stock price reaction ($p < 0.05$), but an insignificant and small trading volume behavior.

The last step is to build a decision tree for the price and volume reactions (namely Reaction Group 1 and Reaction Group 2) based on the cluster ID identified from the breach announcements (from step 1). The dataset was randomly partitioned into two parts: training (80%) and validation and testing (20%), and the classification model was trained, validated, and tested using a decision tree in SAS Enterprise Miner.

For the decision tree, there are 71 documents in the training set (80% of 89 announcements) and 18 documents in the validation and testing set (20% of 89 announcements). The resulting tree only has two branches. We considered other factors, such as firm size (total assets of a firm) and an industry indicator in the model, but these did not result in any new branches. For the two branches, one is associated with cluster A and Reaction Group 2, 71.4% of the time in the validation and testing dataset, while the other branch is associated with cluster B and with Reaction Group 1, 100% of the time in the validation and testing dataset. We further verify our results by a commonly adopted procedure, 10-fold cross validation (e.g., Kohavi 1995; Weiss and Kapouleas 1989). When we repeated our procedure 10 times by randomly drawing 80% of the data and averaged the classification results across 10 different runs, the associations are similar for both the left and the right branch. Next, for the rest of the 20% (the validation dataset), we predict the categories for the textual contents and market reactions. We then estimate the association again. We repeat these steps 10 times. The results are similar. The result from one of our 10-fold cross validations is given in Table 3. Table 3 demonstrates that the overall accuracy rate for this model is 71.83% (21.13% + 50.70%). Similarly, we repeated the process 10 times, and the average accuracy rate of all 10 validation results is about 70%.

Recall that events within Reaction Group 1 depict significantly negative stock price reactions, but an insignificantly small trading volume. Reaction Group 2 events reflect significantly large trading volume, but an insignificant slightly positive stock price reaction. Therefore, it seems that the textual contents, i.e., cluster A and cluster B, in the breach announcements result in different market reactions. This result leads us to further explore the breach announcements in Reaction Group 1 and Reaction Group 2. The exploration of text has long been widely used in psychological constructs, such as therapy transcription (e.g., Peterson, Luborsky, and Sligman 1983) and personality (e.g., Winters 1987). In this research, we apply the same concept and explore the terms within the breach announcements.

We further explore the textual differences between these two datasets (namely Dataset A and Dataset B). Then we performed a cluster analysis by repeating the first step in Figure 1 and using SAS Text Miner again to obtain all the possible groups of words based

Table 3 Confusion matrix for the cross validation results.

Reaction group		Predict A	Predict B	Predict Total
Actual				
A	Frequency	15	7	22
	Percentage	21.13	9.86	30.99
	Row percentage	68.18	31.82	
	Column percentage	50.00	17.07	
B	Frequency	13	36	49
	Percentage	18.31	50.70	69.01
	Row percentage	26.53	73.47	
	Column percentage	43.33	87.80	
Total	Frequency	30	41	71
	percentage	42.25	57.75	100.00

Table 4 Terms in Dataset A and Dataset B.

Group of words	Terms	Percentage	RMS std.
	Dataset A		
1	+***breach***, ***compromise***, computer, ***security***, +***threat***	42%	0.2059
2	+***attacker***, +computer, +disable, +***infect***, +system	58%	0.2093
	Dataset B		
1	+affect, ***credit card***, +***customer***, ***operation***, +***site***	28%	0.1333
2	+***account***, +***amount***, ***data***, +***employee***, +victim	72%	0.1342

Note: For readers' convenience, we highlight the examples discussed in the text as bolded and italicized.

on these two datasets. The settings and procedures are the same in the first step when building the decision tree. In Table 4, each row is a group of words. Within each group, there are five terms that have the largest frequency. The terms with plus (+) signs mean equivalent terms. The percentage is the frequency of a set of terms divided by the total frequency. The root mean squared standard deviation (RMS Std.) for group k is equal to $\sqrt{W_k / [d\,(N_k - 1)]}$, where W_k is the sum of the squared distances from the group mean to each of the N_k documents in group k, and d is the number of dimensions.

We then compare the groups of words associated with Dataset A and Dataset B in Table 4. Each dataset has two groups of words. However, when we investigate the terms within the groups, most of the terms (60%) associated with Dataset A are general terms about security breaches, such as "breach," "compromise," "security," "threat," "attacker," and "infect." That is, these terms are commonly used in breach announcements and are not specific to certain incidents. Accordingly, when looking at the terms in Dataset A, the information regarding the incident seems rather vague. The information contained within the announcements does not appear to shed much light on the specific nature of the incident, particularly how the incident might affect the firm's future performance. On the other hand, 80% of the terms associated with Dataset B are about specific subjects, such as "credit card," "customer," "operation," "site," "account," "amount," "data," and "employee." Furthermore, for Dataset B, terms such as "credit card," "account," and "data" are related to confidentiality type incidents or possibly identity theft.

Recall that the differences between these two datasets are the stock price and trading volume reactions around the breach announcement date. Therefore, it seems that subject-specific terms or terms about confidentiality-type incidents result in a more consistent negative price reaction. This result is intuitive because, with the specific description in the breach announcements, the details of the security breach and how the loss of confidential information might affect a firm's customers are more obvious, resulting in a negative impact on a firm's future performance. However, ambiguous information and general descriptions in the breach announcement leads to different interpretations and assessments of the impact of the security breach.

We perform an additional content analysis by using the General Inquirer software (see http://www.wjh.harvard.edu/~inquirer/) on the two sets of documents to provide additional insights. In the following analysis, we only retain the categories of words that have more than 5% frequency in any of the documents out of the total number of words. For example, if a document has 100 words, we show only the categories with more than 5 words. The results are given in Table 5.

From Table 5, we notice that there are three significantly different dimensions that distinguish these two datasets. First, on average, Dataset B has significantly more words than Dataset A. Second, Dataset B has more words from the ECON category than does Dataset A. The ECON category contains words related to the economy, industries, and businesses, including roles and acts. That is, the documents in Dataset B, on average, describe more about the link between security breaches and the businesses in more detail, such as how security breaches affect business, customers, revenues, etc. Third, Dataset B has more words in the SV category than does Dataset A. The SV category contains state verbs related to mental or emotional status, such as think, understand, anticipate, appreciate, need, can, and so on. From the definitions given by General Inquirer, these words demonstrate the opinions or comments toward certain subjects, for example. That is, on average, the documents in Dataset B provide more information (comments or opinions, for instance)

Table 5 Content analysis results.

	Dataset A	Dataset B	Difference (*p* value)
Avg. number of words	388.04	565.34	−177.30 (0.01)
Categories			
Strong	10.25%	10.47%	−0.22% (0.75)
Active	10.22%	10.14%	0.08% (0.87)
ECON	4.94%	6.85%	−1.91% (0.00)
HU	5.63%	6.10%	−0.47% (0.29)
IAV	7.75%	8.11%	−0.36% (0.39)
DAV	5.17%	5.06%	0.11% (0.67)
SV	4.52%	5.53%	−1.01% (0.02)
PowTot	5.77%	5.89%	−0.12% (0.81)
EnlTot	8.17%	8.32%	−0.15% (0.76)

Please refer to http://www.wjh.harvard.edu/~inquirer/ for detail information about the categories. The definitions of the categories in Table 5 are as follows. Strong: words relate to strength; Active: words implying active orientation; ECON: words of an economic, commercial, industrial, or business orientation; HU: refers to humans including roles; IAV: verbs giving an interpretative explanation of an action; DAV: straight descriptive verbs of an action or feature of an action; SV: state verbs describing mental or emotional states; PowTot: power related words, a valuing of having the influence to affect the policies of others; EnlTot: Enlightenment refers, according to Lasswell, to "knowledge, insight, and information concerning personal and cultural relations."

regarding the breaches. It seems that the breach announcements in Dataset B provide more descriptive details of the breaches, which also reflects why the announcements in Dataset B have more words than those in Dataset A, on average. Therefore, we believe Dataset B appears to contain more information pertaining to how the breach could be affecting business than does Dataset A.

In summary, our findings suggest that general investors could estimate the price and volume reactions to breach announcements based on the textual contents of the announcements. However, given this information, what investment decisions could they make? To address this question, general investors can further consider sophisticated investors' reactions to breach announcements and adjust their investment decisions based on the more "informed" investors' reactions as discussed previously.

4.2. Robustness Tests

We performed the following tests to verify our results. First, we considered using industry, incident type, attack history, composition of investors, and market value of the firm as the classifiers in order to rule out possible alternate explanations to our results. For industry, we controlled for the firms with the two-digit SIC code 73, as about 40% of the firms in our sample are within this category, and our results remain similar. For incident type, we considered confidentiality, integrity, and availability-type incidents. However, incident type is not always clear at the time when the breach announcements are made. This result confirms our finding that it is not clear whether the terms in Dataset A refer to which security incidents. We also considered whether the firm had been attacked before (attack history), how many of the shares outstanding were held by institutional investors (the composition of investors), and the firm value which is the market capitalization one day before the breach announcement. We take into account these three factors because they could also affect the market reactions to breach announcements. Our results remain similar.

Second, as pointed out in Wang and colleagues (2013), the textual contents of security risk factors disclosed in financial reports could also affect the market reactions. Accordingly, we also took into account the textual contents of security risk factors disclosed in financial reports as the classifier. However, our results are similar. Last, instead of performing a cluster analysis on Dataset A and Dataset B, we performed the analysis on the documents associated with Reaction Group 1 and Reaction Group 2, and our results are qualitatively similar.

5. INVESTMENT OPPORTUNITY

In this section, we show how investors could use the results in Section 4 and the sophisticated investors' reactions to form short-term trading strategies. We first investigate sophisticated investors' reactions to breach announcements. Then we compare this reaction to the classification results and show profitable short-term investment opportunities for general investors.

5.1. Sophisticated Investors' Reactions to Breach Announcements

For sophisticated investors' reactions to breach announcements, we considered the revision of analyst forecast and change of institutional ownership.

Analyst forecast data was collected from the $I/B/E/S$ database. We calculated (1) the consensus of analyst forecasts of earnings per share (EPS) for the corresponding quarter before and after breach announcement articles, and collected (2) the actual quarterly EPS for each of the breached firms in our sample. The former shows whether there is any forecast change after breach announcements, and the later verifies the actual impact as compared to analysts' forecasts. For the consensus forecasts before the breach announcement, we calculated the median of analysts' forecasts made within one year before the quarter when incidents occurred for each breached firm. This consensus was used as the reference point for the firm's performance for that quarter *without* security breach announcements. We chose this one-year period because the forecasts are more accurate when they are made closer to the end of the reporting period (e.g., Brown 1991; O'Brien 1988).

For the consensus forecasts after the breach announcement, we searched for any forecast revision immediately after the incidents and calculate the median of these revised forecasts. Although studies such as that by Ivkovic and Jegadeesh (2004) showed that about 20% to 26% of analyst revisions of earnings estimates are issued at the earnings announcement date and within the following two days, some studies uses a three-week period (e.g., Bowen, Hash, and Wilson 2002). To be conservative, we also searched for all possible forecast revisions occurring up to three weeks after the breach announcement. If there was any forecast revision, it was attributed to the security incident after controlling for all other announcements, such as announcements of mergers and acquisitions, by searching for related news articles on *LexisNexis* and the firm's website.

For institutional ownership, we searched the 13-F filings of the corresponding quarters through 10-K Wizard before and after breach announcements. Although 13-F filings only provide the shares held by investment institutions at the end of each quarter, if the breach results in a significant impact on the firm's future performance, we should still observe some significant changes in institutional ownership in the quarter before and after breach announcements. Similarly, if there was any change, we searched for news articles on *LexisNexis* and the firm's website to investigate any events that could result in the change of position.

Our results show that about 33% of our observations have some analyst forecast revisions after the breach announcement. Interestingly, none of these forecast revisions can be associated with security incidents. Second, for institutional ownership, we do not observe any significant change ($p > 0.10$) before (about 62% on average) and after (about 64% on average) breach announcements. These findings suggest that the sophisticated investors might not consider information security breaches as an event that will significantly affect a firm's future performance in the time window around the breach announcement day. This observation was further verified by comparing the breached firm's subsequent actual quarterly performance with the analyst forecasts. The comparison results confirm our results and demonstrate that, without other future events, the firms' average performance is \$0.02 greater than the average analysts' forecasts ($p < 0.05$).

In order to rule out alternative explanations to our results, we first performed the same set of analyses on a list of controlled firms that did not have any breach announcements and did not demonstrate any significant increase in trading volume. The actual quarterly performance for these controlled firms was also higher than the forecasts. Second, we considered the time effect, incident types and attack history but our results were similar. Last, we also considered analyst recommendations, sales, ROA, annual forecasts, and two-year forecasts as the performance measures and did not observe any forecast revisions after breach announcements.

Based on the foregoing results, we interviewed two analysts and two investment portfolio managers to investigate their reactions to breach announcements in order to provide more insight into our findings. Two major reasons emerged as to why they do not immediately react negatively to breach announcements. First, although they do care about confidentiality-type incidents, the information regarding the security incident around the breach announcement day is typically ambiguous. It might require more time before the detailed breach information is available and can be clarified. Second, the impact of security breaches should be jointly considered with each breached firm's characteristics, such as the overall business risks, the market share, the competition in the market, and the operational advantages and disadvantages. Therefore, the impact of security breaches on a firm's future performance should be evaluated on a firm-level basis in order to have a better understanding of the impact of the breach. However, because the second point requires further case studies on various firms from different industries to provide additional insight, we leave it as a future research avenue.

For the first point, we focused on confidentiality incidents in our sample and searched for all the analyst reports about the breached firm in the Morningstar database after each breach announcement up to the end of 2008. Among the 32 observations of confidentiality-type incidents in our sample, we found 1 analyst report discussing the security breaches of T J Maxx. Although T J Maxx suffered from credit card data losses in early 2007, one analyst considered this event as a bearish cause to the stock price in the June 2008 report (two months before the alleged hackers were arrested). That is, the event was considered after 18 months when the breach information was clarified. However, we did not find other analyst reports for the remaining observations of confidentiality-type incidents in our sample. This could be due to analysts considering the firm characteristics as mentioned previously or that the complete analyst reports are not covered by the database. We further searched for similar analyst reports on bloggingstocks.com and did not find any. Nevertheless, these findings suggest that sophisticated investors do react to security incidents, but not in the two-day window as in most event studies or the short forecast revision period for earnings announcements. This result suggests that, for information security incidents, the time needed for sophisticated investors to react could be much longer.

5.2. Profitable Short-Term Investment Opportunities

The results demonstrate that the textual contents of the news articles containing breach announcements are associated with both stock price and trading volume reactions. Also, given that sophisticated investors do not typically react immediately after publication of the breach announcement, the negative stock price in Reaction Group 2 in our classification model appears to be driven by unsophisticated investors. Because unsophisticated investors only temporarily affect stock price (e.g., Bamber and Cheon 1995), the negative stock price reaction is indeed only temporary. Therefore, it is possible that the general investor could take advantage of this reaction difference and garner profitable short-term investments.

In order to demonstrate that the profitable short-term investment opportunity exists and to support our argument about the temporary stock price drop as detailed previously, we simulated a trading strategy by buying the breached firm's stock using the closing price on the breach announcement date and selling the stock after three trading days, also using the closing price. The result shows that we are able to make an average of 0.84% daily return (about 300% annually). This trading strategy is validated by investigating the cumulative abnormal return for the window (1, 3), where 1 (3) means 1 day (3 days) after publication of the breach announcement, for those breached firms encountering a negative stock price reaction after breach announcements. We focus only on these firms because, from the results in previous sections, the negative stock price reaction is driven by unsophisticated traders. By focusing on these firms, we are able to take advantage of the different beliefs among investors. The result shows that the average abnormal return is about 2% ($p < 0.10$), which verifies our positive trading strategy and further confirms our observation that the stock price drop around the breach announcement date is only temporary.

The profitable short-term investment trading strategy is further examined by investigating the change in implied volatility before and after the breach announcement. The implied volatility is the theoretical volatility based on the option pricing model (see Appendix C) and has been shown to be a good prediction of the firm's future volatility (e.g., Christensen and Prabhala 1998; Harvey and Whaley 1992; Sheikh 1989). Based on data collected from the database *OptionMetrics*, the implied volatility decreases about 1.26% ($p < 0.05$) after the breach announcement. This decrease suggests that in the long-run, the breached firms' business values will restore to their normal state, other things being equal, which verifies the possibility of our trading strategy.

6. CONCLUSIONS AND DISCUSSION

Our results show that the stock price and the trading volume behavior around the breach announcement day are associated with the textual contents of the corresponding breach announcement(s). In particular, breach announcements containing specific information regarding the incident, such as the subject affected, or news articles about confidentiality-type incidents or identity theft often lead to a negative stock price reaction, but small trading volume reactions. However, breach announcements with unclear or ambiguous incident information could result in different beliefs about the impact of security breaches on a firm's future performance (i.e., a high trading volume but small stock price reactions). Interestingly, sophisticated investors typically do not react to breach announcements around the breach announcement day and the negative stock price reactions we observed are only temporary. By taking into account the differences between the overall market reactions and sophisticated investors' reactions, it is possible to have profitable short-term investment opportunities.

Our results have implications for investors and managers. For investors, this study demonstrates that general investors do not have to overreact to security incidents. They can form or adjust their investment strategy based on the breach announcements and could have profitable investment opportunities. This implication for investors is useful to managers. That is, managers need to appreciate the impact of security breach announcements on a firm's business value from investors' perspectives. In our context, managers especially need to consider the sophisticated investors' perspectives.

Our study shows that although there are negative impacts of security breaches on a firm's business value, this is only temporary, on average. This can have important implications for information security investment. Specifically, information security resources are allocated based on the importance of the object that is commonly measured as the potential impact when a breach occurs (e.g., Gordon and Loeb, 2002). As suggested by our findings, the temporary drop of business value may not be a good indicator of the impact of security breaches on business value, which, in turn, cannot be incorporated when forming information security investment/deployment decisions. More importantly, managers need to better publicly respond to incidents in order to lower information asymmetry and possibly lower the temporary negative market reactions. Specific information about the incident may still lead to a temporary drop of stock price but, on average, it does not last long. However, being vague when facing security breaches may cause a delay in the reaction and impose more uncertainty onto a firm's business value.

There are several limitations of this article. First, the sample size is relatively small for market reaction estimates and for text mining. Although we have collected as many observations as possible for our analyses, the number of breach announcements for publicly traded firms is limited based on our data processing criteria. Also, from the previous literature, we believe the performance of our model could increase as the sample size increases. However, the generalizability of the results to a larger dataset may be limited. Second, we show that the sophisticated investors do not tend to react negatively to breach announcements. However, how sophisticated investors evaluate the impact of breach announcements and determine whether to adjust their forecasts or investment portfolios are out of the scope of the current study. Last, we only consider a short time frame around the breach announcement date. However, some breach announcements have more detailed and new information regarding the incidents in follow-up news articles or other media, such as blogs, which are not considered in this study.

Possible future extensions are as follows. First, a detailed understanding of how sophisticated investors assess the impact of security incidents and why these investors do not immediately react negatively to security breaches can be further investigated. Second, given that managers and other insiders are more likely to know about the breach before the media, it is possible that the insiders have traded this information before the market. The insiders' reactions could further explain the impact of security incidents on a firm's future performance. Third, different media, such as Web 2.0 technologies, are now popular information sources for investors. We can further consider other media sources, such as blogs, micro-blogs, and social networking applications, to investigate the relationship among different information sources, information security incidents, and market reactions. Last, detailed case studies of various firms from different industries could further explain the impact of security incidents on a firm's future performance and why sophisticated investors do not immediately react negatively around the breach announcement day.

ACKNOWLEDGMENTS

The authors would like to thank the review team for their valuable comments and suggestions. The authors also thank the Center for Education and Research in Information Assurance and Security (CERIAS) as well as the participants of the 20th Workshop on Information Systems and Economics (WISE 2008) and the 10th Annual Information Security Symposium Poster Session. The authors are grateful for the financial support from the University of Hawaii at Manoa and Purdue University.

REFERENCES

Ajinkya, B. B., and P. C. Jain. 1989. "The behavior of daily stock market trading volume." *Journal of Accounting and Economics* 1(4):331–359.

Aquisti, A., A. Friedman, and R. Telang. 2006. "Is there a cost to privacy breaches? An event study." *The 5th Workshop on the Economics of Information Security*, Robinson College, University of Cambridge, England.

Atiase, A., and L. Bamber. 1994. “Trading volume reactions to annual accounting earnings announcements: The incremental role of predisclosure information asymmetry.” *Journal of Accounting and Economics* 17(3):281–308.

Ayers, B. C., J. Jiang, and P. E. Yeung. 2006. “Discretionary accruals and earnings management: An analysis of pseudo earnings targets.” *The Accounting Review* 81(3):617–652.

Baesens, B., R. Setiono, C. Mues, and J. Vanthienen. 2003. “Using neural network rule extraction and decision tables for credit-risk evaluation.” *Management Science* 49(3):312–329.

Bamber, L. 1987. “Unexpected earnings, firm size, and trading volume around quarterly earnings announcements.” *The Accounting Review* 62(3):510–532.

Bamber, L., O. E. Barron, and T. L. Stober. 1997. “Trading volume and different aspects of disagreement coincident with earnings announcements.” *The Accounting Review* 72(4):575–597.

Bamber, L., and Y. S. Cheon. 1995. “Differential price and volume reactions to accounting earnings announcements.” *The Accounting Review* 70(3):417–441.

Barron, O. E., D. Byard, and Y. Yu. 2008. “Earnings surprises that motivate analysts to reduce average forecast error.” *The Accounting Review* 83(2):303–325.

Barth, M., R. Kasznik, and M. McNichols. 2001. “Analyst coverage and intangible assets.” *Journal of Accounting Research* 39(1):1–34.

Beaver, W. 1968. “The information content of annual earnings announcements.” *Journal of Accounting Research* 6:67–92.

Beneish, M. D. 2001. “Earnings management: A perspective.” *Managerial Finance* 27(12):3–17.

Bhattacharya, N. 2001. “Investors’ trade size and trading responses around earnings announcements: An empirical investigation.” *The Accounting Review* 76(2):221–244.

Bhushan, R. 1989. “Firm characteristics and analyst following.” *Journal of Accounting and Economics* 11(2–3):255–274.

Bowen, P., J. Hash, and M. Wilson. 2002. *Information security handbook: A guide for managers*. Gaithersburg, MD: NIST Special Publication 800–100.

Brandãn, L. E., J. S. Dyer, and W. J. Hahn. 2005. “Using binomial decision trees to solve real-option valuation problems.” *Decision Analysis* 2(2):69–88.

Brown, L. D. 1991. “Forecast selection when all forecasts are not equally recent.” *International Journal of Forecasting* 7(3):349–356.

Brown, L. D. 1993. “Earnings forecasting research: Its implications for capital markets research.” *International Journal of Forecasting* 9(3):295–320.

Campbell, K., L. A. Gordon, M. P. Loeb, and L. Zhou. 2003. “The economic cost of publicly announced information security breaches: Empirical evidence from the stock market.” *Journal of Computer Security* 11(3):431–448.

Cavusoglu, H., B. Mishra, and S. Raghunathan. 2004. “The effect of internet security breach announcements on market value of breached firms and internet security developers.” *International Journal of Electronic Commerce* 9(1):69–105.

Cecchini, M., H. Aytug, G. J. Koehler, and P. Pathak. 2010. “Detecting management fraud in public companies.” *Management Science* 56(7):1146–1160.

Christensen, B. J., and N. R. Prabhala. 1998. “The relation between implied and realized volatility.” *Journal of Financial Economics* 50(2):125–150.

Core, J. E. 2001. “A review of the empirical disclosure literature: Discussion.” *Journal of Accounting and Economics* 31(1–3):441–456.

CSI/FBI. 2007. “The CSI/FBI computer crime and security report in 2006.” Retrieved from http://abovesecurity.com/doc/communiquespdf/fbisurvey2006 (accessed April 9, 2007).

Degeorge, F., J. Patel, and R. Zeckhauser. 1999. “Earnings management to exceed thresholds.” *The Journal of Business* 72(1):1–33.

Easley, D., and M. O’Hara. 1987. “Price, trade size, and information in securities markets.” *Journal of Financial Economics* 19(1):69–90.

Elgers, P., and D. Murray. 1992. “The relative and complementary performance of analyst and security-price-based measures of expected earnings.” *Journal of Accounting and Economics* 15(2–3):303–346.

Ettredge, M. L., and V. J. Richardson. 2003. “Information transfer among internet firms: The case of hacker attacks.” *Journal of Information Systems* 17(2):71–82.

Fama, E. F., and K. R. French. 1992. “The cross-section of expected stock returns.” *Journal of Finance* 47:427–465.

Fan, W., L. Wallace, S. Rich, and Z. Zhang. 2006. “Tapping the power of text mining.” *Communications of the ACM* 49(9):77–82.

Francis, J., J. D. Hanna, and D. R. Philbrick. 1997. “Management communications with securities analysts.” *Journal of Accounting and Economics* 24(3):363–394.

Francis, J., K. Schipper, and L. Vincent. 2002. “Earnings announcements and competing information.” *Journal of Accounting and Economics* 33(3):313–342.

Frankel, R., S. P. Kothari, and J. Weber. 2006. “Determinants of the informativeness of analyst research.” *Journal of Accounting and Economics* 41(1–2):29–54.

Gal-Or, E., and A. Ghose. 2005. “The economic incentives for sharing security information.” *Information Systems Research* 16(2):186–208.

Garg, A., J. Curtis, and H. Halper. 2003. “Quantifying the financial impact of IT security breaches.” *Information Management and Computer Security* 11(2):74–83.

Glover, S., S. Liddle, and D. Prawitt. 2001. *E-Business: Principles and strategies for accountants*. 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Gordon, L. A., and M. P. Loeb. 2002. “The economics of information security investment.” *ACM Transactions on Information and System Security* 5(4):438–457.

Gordon, L. A., M. P. Loeb, and W. Lucyshyn. 2003. “Sharing information on computer systems security: An economic analysis.” *Journal of Accounting and Public Policy* 22(6):503–530.

Goto, M., T. Kawamura, K. Wakai, M. Ando, M. Endoh, and Y. Tomino. 2008. “Risk stratification for progression of IgA nephropathy using a decision tree induction algorithm.” *Nephrology Dialysis Transplantation* 24(4):1242–1247.

Han, J., R. Altman, V. Kumar, H. Mannila, and D. Pregibon. 2002. “Emerging scientific applications in data mining.” *Communications of the ACM* 45(8):54–58.

Harvey, C. R., and R. E. Whaley. 1992. “Dividends and S&P 100 index option valuation.” *Journal of Futures Markets* 12(2):123–137.

Hasbrouck, J. 1988. “Trades, quotes, inventories, and information.” *Journal of Financial Economics* 22:229–252.

Hasbrouck, J. 1991. “Measuring the information content of stock trades.” *Journal of Finance* 46(1):179–207.

Hovav, A., and J. D’Arcy. 2003. “The impact of denial-of-service attack announcements on the market value of firms.” *Risk Management and Insurance Review* 6(2):97–121.

Ivkovic, Z., and N. Jegadeesh. 2004. “The timing and value of forecast and recommendation revisions.” *Journal of Financial Economics* 73(3):433–463.

Kannan, K., J. Rees, and S. Sridhar. 2007. “Market reactions to information security breach announcements: An empirical study.” *International Journal of Electronic Commerce* 12(1):69–91.

Karpoff, J. M. 1986. “A theory of trading volume.” *Journal of Finance* 41(5):1069–1087.

Kasznik, R., and B. Lev. 1995. “To warn or not to warn: Management disclosures in the face of an earnings surprise.” *The Accounting Review* 70(1):113–134.

Kim, J. W., B. H. Lee, M. J. Shaw, H. Chang, and M. Nelson. 2001. “Application of decision-tree induction techniques to personalized advertisements on internet storefronts.” *International Journal of Electronic Commerce* 5(3):45–62.

Kim, O., and R. Verrecchia. 1991. “Trading volume and price reactions to public announcements.” *Journal of Accounting Research* 29(2):302–321.

Kim, O., and R. Verrecchia. 1994. “Market liquidity and volume around earnings announcements.” *Journal of Accounting and Economics* 17(1):41–67.

Kim, O., and R. Verrecchia. 1997. “Pre-announcement and event-period private information.” *Journal of Accounting and Economics* 24(3):395–419.

Kohavi, R. 1995. "A study of cross-validation and bootstrap for accuracy estimation and model selection." *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montréal, Québec, Canada.

Kross, W., B. Ro, and D. Schroeder. 1990. "Earnings expectations: The analysts information advantage." *The Accounting Review* 65(2):461–476.

Kwon, J., J. Rees, and T. Wang. 2013. "The association between top management involvement and compensation and information security breaches." *Journal of Information Systems*, forthcoming.

Lakonishok, J., A. Shleifer, and R. W. Vishny. 1992. "The impact of institutional trading on stock prices." *Journal of Financial Economics* 32:23–43.

Lang, M. H., and R. J. Lundholm. 1993. "Cross-sectional determinants of analyst ratings of corporate disclosures." *Journal of Accounting Research* 31(2):216–271.

Lev, B., and R. Thiagarajan. 1993. "Fundamental information analysis." *Journal of Accounting Research* 31(2):190–215.

Masand, B., G. Linoff, and D. Waltz. 1992. "Classifying news stories using memory based reasoning." *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Copenhagen, Denmark.

Matsumoto, D. A. 2002. "Management's incentives to avoid negative earnings surprises." *The Accounting Review* 77(3):483–514.

McNichols, M. F. 2000. "Research design issues in earnings management studies." *Journal of Accounting and Public Policy* 19(4–5):313–345.

Morse, D. 1981. "Price and trading volume reaction surrounding earnings announcements: A closer examination." *Journal of Accounting Research* 19(2):374–383.

O'Brien, P. 1988. "Analysts' forecasts as earnings expectations." *Journal of Accounting and Economics* 10(1):53–83.

OptionMetrics. 2006. *Ivy DB file and Data Reference Manual*. New York: Author.

Peterson, C., L. Luborsky, and N. E. Seligman. 1983. "Attribution and depressive mood shifts: A case study using the symptom-context method." *Journal of Abnormal Psychology* 92(1):96–103.

Roulstone, D. T. 2003. "Analyst following and market liquidity." *Contemporary Accounting Research* 20(3):551–578.

SAS Institute. 2004. *Getting Started with SAS_9.1 Text Miner*. Cary, NC: SAS Institute.

Schechter, S. E., M. D. Smith. 2003. "How much security is enough to stop a thief? The economics of outsider theft via computer systems networks." *Proceedings of the Financial Cryptography Conference*, Gosier, Guadeloupe.

Sheikh, A. 1989. "Stock splits, volatility increases and implied volatility." *Journal of Finance* 44(5):1361–1372.

Siponen, M. 2006. "Information security standards focus on the existence of process, not its content." *Communications of the ACM* 49(8):97–100.

Siponen, M., and J. Iivari. 2006. "Six design theories for IS security policies and guidelines." *Journal of the Association for Information Systems* 7(7):445–472.

Smyth, P. 2000. "Model selection for probabilistic clustering using cross-validated likelihood." *Statistics and Computing* 10(1):63–72.

Stickel, S. E. 1990. "Predicting individual analyst earnings forecasts." *Journal of Accounting Research* 28(2):409–417.

Still, S., and W. Bialek. 2004. "How many clusters? An information-theoretic perspective." *Neural Computation* 16(12):2483–2506.

Straub, D. W. 1990. "Effective IS security: An empirical study." *Information Systems Research* 1(3):255–276.

Sordo, M., and Q. Zeng. 2005. "On sample size and classification accuracy: A performance comparison." *Proceedings of the 6th International Conference on Biological and Medical Data Analysis*, Aveiro, Portugal.

Tibshirani, R., G. Walther, and T. Hastie. 2001. “Estimating the number of clusters in a dataset via the gap statistic.” *Journal of the Royal Statistical Society B* 63(2):411–423.

Wang, T., and C. Hsu. 2010a. “The impact of a company’s board structure on the effectiveness of information security management.” *Pacific Asia Conference on Information Systems (PACIS)*, Taipei, Taiwan.

Wang, T., and C. Hsu. 2010b. “The composition of the top management team and the effectiveness of information security management.” *Americas Conference on Information Systems*, Lima, Peru.

Wang, T., K. Kannan, and J. Rees. 2013. “The association between the disclosure and the realization of information security risk factors.” *Information Systems Research*, forthcoming.

Wakabayashi, D. 2011. “Sony CEO Warns of ‘Bad New World.’” *The Wall Street Journal*, May 18 2011.

Weiss, S. M., and L. Kapouleas. 1989. “An empirical comparison of pattern recognition, neural nets, and machine learning classification methods.” *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, Detroit, MI.

Winters, D. G. 1987. “Leader appeal, leader performance, and the motive profiles of leaders and followers: A study of American presidents and elections.” *Journal of Personality and Social Psychology* 52(1):196–202.

Young, S. R., and P. J. Hayes. 1985. “Automatic classification and summarization of banking telexes.” *Proceedings of the 2nd IEEE Conference on AI Applications*, Miami Beach, FL.

Zhang, S., and Z. Zhu. 2006. “Research on decision tree induction from self-map space based on web.” *Knowledge-Based Systems* 19(8):675–680.

Zhou, Z., and Y. Jiang. 2004. “NeC4.5: Neural ensemble based C4.5.” *IEEE Transactions on Knowledge and Data Engineering* 16(6):770–773.

AUTHOR BIOS

Tawei Wang is Assistant Professor of Accounting at the Shidler College of Business, University of Hawaii at Manoa. Formerly an Assistant Professor of Accounting at National Taiwan University, he is a Certified Public Accountant in Taiwan and a Certified Internal Auditor. Dr. Wang received his Ph.D. from the Krannert Graduate School of Management, Purdue University in 2009. His research interests are risk management, voluntary disclosures and financial reporting, and IT management. His papers have appeared in several leading journals, such as *Information Systems Research*, *Decision Support Systems*, *Journal of Banking and Finance, Journal of Accounting and Public Policy*, *Journal of Information Systems*, and *OMEGA*, plus several leading conferences including AAA Annual Meetings, Annual Meeting of the Academy of Management, INFORMS, Americas Conference on Information Systems, Pacific Asia Conference on Information Systems, Workshop on the Economics of Information Security, and Workshop on Information Systems and Economics.

Jackie Rees Ulmer is Associate Professor of Management Information Systems at Purdue University. Her current research interests include machine learning and risk management for information security. She has published research articles in *Information Systems Research, Decision Support Systems*, *Decision Sciences, Communications of the ACM*, *Journal of Information Systems, INFORMS Journal of Computing*, *European Journal of Operational Research, IEEE Transactions on Systems, Man, and Cybernetics: Part C,* and *Information Technology and Management.* Dr. Ulmer has taught courses in information security, Java, database management, data mining, and the required MBA MIS core course. She serves on the editorial boards of *Decision Sciences, Communications of the*

AIS, and *Journal of Database Management.* She served as co-chair of the 22nd Workshop on Information Technologies and Systems (WITS) and has been Secretary-Treasurer for the Institute for Operations Research and Management Sciences (INFORMS) Information Systems Society (ISS). Dr. Ulmer is a fellow of the Center for Education and Research in Information Assurance and Security (CERIAS) and is also a member of INFORMS and the Association for Information Systems (AIS). She earned her Ph.D. in Decision and Information Sciences from the Warrington College of Business Administration at the University of Florida.

Karthik Kannan is Associate Professor of Management at Purdue's Krannert School of Management. His research focuses on pricing of information goods/services through auctions, managing data networks, and economics of information security. He has published in several leading conferences and journals in the information systems area, including *Management Science*, *Information Systems Research*, *International Journal of Electronic Commerce*, Workshop on Information Technology and Systems, Workshop on Information Systems Economics, International Conference on Information Systems, and Conference on Information System and Technology. His papers have won the Best Paper Awards in the 10th and the 15th Annual Workshop on Information Technology and Systems. Dr. Kannan serves/has served as an Associate Editor for *Management Science, Information Systems Research, and MIS Quarterly*. He is a member of AIS and INFORMS. He is also a CERIAS Fellow and a Krannert Faculty Fellow. He received his PhD, MS in Electrical and Computer Engineering, M.Phil in public policy and management, all from Carnegie Mellon University. Prior to joining the PhD program, he worked for Infosys Technologies.

APPENDIX A. STOCK PRICE AND TRADING VOLUME REACTIONS TO SECURITY INCIDENTS

We use the market model (Equation A.1) to capture the stock price reaction.

$$R_{it}=\beta_0+\beta_1 R_{mt}+\varepsilon_{it}, \tag{A.1}$$

where R_{it} represents company i's return at time t. R_{mt} is the market return, which is estimated by the CRSP equally weighted index, at time t. We estimate the coefficients by using the ordinary least square (OLS) method in a 255-day periods ending at 45 days before the announcement day. The abnormal returns (AR) are calculated as in Equation A.2.

$$AR_{it} = R_{it} - \hat{\beta}_0 - \hat{\beta}_1 R_{mt} \tag{A.2}$$

We use the mean cumulative abnormal returns to capture the market reactions to an economic event. Mean cumulative abnormal returns is the summation of abnormal returns given the window we choose, i.e.,$\left(\sum_{t=1}^{N}\sum_{t_0}^{t_1} AR_{it}\right)\Big/N$, where t_0 and t_1 are the beginning and the ending days for the window. Cumulative abnormal returns (CAR,$\sum_{t_0}^{t_1} AR_{it}$) are used in our analysis.

In addition to the market model, we use the Fama-French three-factor model (Fama and French 1992) as a robustness test.

$$R_{it} = \alpha + \beta_i R_{mt} + s_i SMB_t + h_i HML_t + \varepsilon_{it} \tag{A.3}$$

The three-factor model considers two other factors in addition to the market return. That is, SMB_t is the difference between the average return of small and large market-capitalization portfolios. HML_t is the difference between the average return of high and low book-to-market equity portfolios. See Fama and French (1992) for a detailed explanation. Again, we estimate the coefficients by using the ordinary least square (OLS) method in a 255-day periods ending at 45 days before the announcement day. Similarly, the abnormal returns (*AR*) is calculated as in Equation A.4. The mean cumulative abnormal returns and cumulative abnormal returns are calculated as described previously.

$$AR_{it} = R_{it} - \left(\hat{\alpha} + \hat{\beta}_i R_{mt} + \hat{s}_i SMB_t + \hat{h}_i HML_t\right) \tag{A.4}$$

The cumulative abnormal daily trading volume percentage (CAV_{it}) for firm i at time t is estimated by Equation A.5.

$$V_{it} = \alpha + \beta V_{mt} + \varepsilon_{it}, \tag{A.5}$$

where V_{it} represents the natural log of one plus the daily trading volume divided by the total number of outstanding shares of firm i at time t, and V_{mt} represents the natural log of one plus the daily trading volume divided by the total number of all the firm's outstanding shares for the S&P 500 Composite Index at time t. The logarithm transforming can make the distribution of the prediction error approximately normal distributed (Ajinkya and Jain 1989). α and β are the parameters and ε is the error term. The parameters are estimated in a 255-day periods ending at 45 days before the two-day estimation window by ordinary least square (OLS) method. Then the abnormal trading volume is calculated by summing $V_{it} - \hat{\alpha} - \hat{\beta} V_{mt}$ over a two-day window (-1, 0) where 0 (-1) represents the day of (one day before) the breach announcement. The mean abnormal trading volume equals to abnormal trading volume divided by the total number of observations which is used to test the significance of the trading volume.

The previous measure for trading volume behavior controls for the market effect. Another measure controls for firm-specific effect and allows us to examine whether the trading volume is different from the general trading behavior of each firm. In particular, the abnormal trading volume equals to the average trading volume of firm i two days around the breach announcement divided by the average trading volume of firm i 30 days before the announcement.

APPENDIX B. CLUSTER ANALYSIS USING SAS TEXT MINER

The settings of the cluster analysis in SAS Text Miner are summarized as follows. Text Miner decomposes the sentences in the news articles into terms and creates a frequency matrix. When decomposing the documents, we chose to rule out definite as well as indefinite articles, conjunctions, auxiliaries, prepositions, pronouns, and interjections, as these terms do not help provide meaningful results in our context. For the frequency matrix, the weight for term i in document j (w_{ij}) was the multiplication of the frequency weight (L_{ij}) and the term weight (G_i). In our study, the frequency weight was the logarithm of the frequency (f_{ij}) of term i in document j plus one, i.e., $L_{ij} = \log_2 (f_{ij} + 1)$. The term weight of term i (G_i) was calculated as $1 + \sum_j \left(p_{ij} \log_2 \left(p_{ij}\right) \big/ \log_2 n\right)$, where $p_{ij} = f_{ij} / g_{ij}$, g_i was the number of times term i appears in the dataset, and n was the number of documents in the dataset. In this regard, we put more weight on words that show in few documents and generally give the best results (SAS Institute 2004). That is, when we put more weight on words that infrequently show, it would be easier to distinguish among clusters. We also consider assigning equal weights to different terms, and our results are qualitatively similar.[3] Accordingly, we only present the results based on the logarithm calculation of the weight in the following sections.

APPENDIX C. IMPLIED VOLATILITY

The implied volatility is calculated based on the Black-Scholes option pricing model through the database *OptionMetrics* (OptionMetrics 2006):

$$c = Se^{-qT}N(d_1) - Ke^{-rT}N(d_2) \tag{C.1}$$

$$p = Ke^{-rT}N(-d_2) - Se^{-qT}N(-d_1), \tag{C.2}$$

where c is the price of a call option, p is price of a put option, S is the current stock price, K is the strike price of the option, T is the time remaining to expiration (in years), r is the continuously-compounded interest rate calculated based on the BBA LIBOR rates and the Eurodollar settlement price (see Ivy DB Reference Manual (OptionMetrics 2006) for a detailed explanation), q is the continuously-compounded dividend yield (see OptionMetrics 2006 for a detailed explanation), and σ is the historical volatility, which equals the standard deviation of historic price change per share. In Equation C.1 and Equation C.2, d_1 equals $\left[\ln\left(S/K\right) + \left(r - q + 1/2\sigma^2\right)T\right]\Big/\sigma\sqrt{T}$ and d_2 equals $d_1 - \sigma\sqrt{T}\Big/2$. Different from the historical volatility in Equation C.1 and Equation C.2, implied volatility is the volatility in the Black-Scholes model calculated based on the option price and the stock price of the firm.

[3] As given in Table 5, Dataset A has an average of about 388 words and Dataset B has an average of about 565 words. The media reports are relatively short. In addition, the most popular word in our dataset, not surprisingly, is "security," which only appears about 2.7 times per announcement. Most of the words only appear once in a particular media article. These may be the reasons (or our intuitions) why the weighting scheme does not affect our results.