CERIAS Tech Report 2009-24
Essays on information security from an economic perspective
by Ta-Wei Wang
Center for Education and Research
Information Assurance and Security
Purdue University, West Lafayette, IN 47907-2086

Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By TA-WEI WANG

Entitled ESSAYS ON INFORMATION SECURITY FROM AN ECONOMIC PERSPECTIVE

For the degree of DOCTOR OF PHILOSOPHY

Is approved by the final examining committee:

Jackie Rees Ulmer, Co-Chair
Chair

Karthik Kannan, Co-Chair

Kemal Altinkemer

Susan G. Watts

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): Jackie Rees Ulmer

Approved by: Mark Bagnoli 7/19/09
Head of the Graduate Program Date

Graduate School Form 20
(Revised 6/09)

PURDUE UNIVERSITY
GRADUATE SCHOOL

Research Integrity and Copyright Disclaimer

Title of Thesis/Dissertation:
ESSAYS ON INFORMATION SECURITY FROM AN ECONOMIC PERSPECTIVE

For the degree of DOCTOR OF PHILOSOPHY

I certify that in the preparation of this thesis, I have observed the provisions of *Purdue University Executive Memorandum No. C-22,* September 6, 1991, *Policy on Integrity in Research.**

Further, I certify that this work is free of plagiarism and all materials appearing in this thesis/dissertation have been properly quoted and attributed.

I certify that all copyrighted material incorporated into this thesis/dissertation is in compliance with the United States' copyright law and that I have received written permission from the copyright owners for my use of their work, which is beyond the scope of the law. I agree to indemnify and save harmless Purdue University from any and all claims that may be asserted or that may arise from any copyright violation.

TA-WEI WANG
Printed Name and Signature of Candidate

7/10/2009
Date (month/day/year)

*Located at http://www.purdue.edu/policies/pages/teach_res_outreach/c_22.html

ESSAYS ON INFORMATION SECURITY FROM AN ECONOMIC PERSPECTIVE

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Ta-Wei Wang

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2009

Purdue University

West Lafayette, Indiana

ACKNOWLEDGMENTS

Many people have provided guidance and supports along the process of the completion of this dissertation. This is a great opportunity to express my gratitude to all these people who have made this dissertation possible. I would like to thank my chair Dr. Jackie Rees and co-chair Dr. Karthik Kannan. I have been fortunate to have the advisors who gave me the freedom to explore the topics on my own, were patient to brainstorm with me, commented on the manuscript and provided possible solutions to all kinds of problems I faced. I am grateful to my committee members, Dr. Kemal Altinkemer and Dr. Susan Watts. Dr. Kemal Altinkemer's encouragement and suggestions have made the process smoother and the papers in the dissertation better. I am thankful to Dr. Susan Watts for support on different research issues and the experience I had when writing a paper with her. Also, her insightful comments have made the dissertation more complete.

I am indebted to all other the MIS faculty members, Dr. Prabuddha De, Dr. Alok Chaturvedi, Dr. Jungpil Hahn, Dr. Jeffrey Hu, and Dr. Zulei Tang, for their suggestions on the dissertation and the supports for different research topics. I am also grateful to Dr. Jungpil Hahn for his insightful comments on the DSS paper. I would like to thank Dr. Mark Bagnoli for providing suggestions on my research, the guidance on the web disclosure paper, and career planning. I am also grateful to Dr. Roy Dejoie for providing useful tips and suggestions for my teaching at Purdue. Last, many thanks to everyone in

the Ph.D. office, Kelly Felty, Marcella Vansickle, and Cynthia Madden, for their helps starting from the first day I joined the program.

I would like to thank the faculty members and Ph.D. students who participated in the AMCIS and the ICIS doctoral consortium. Their insightful comments and suggestions have made the papers better and broaden my view of different types of researches. Also, the conversation with other Ph.D. students at similar stages has made the process of completion a dissertation less stressful. The Ph.D. colleagues at Krannert and many friends also have helped me review as well as discuss my work. I am grateful to their support through my stay at Purdue.

Most importantly, the dissertation could not be done without the support from my family. This dissertation is dedicated to my beloved parents, Karl H. Wang and Frances Su, my wife, Wen-Ming Chen, and the cutest daughter, Clara Wang. They have provided constant support of love and strength in all these years. I would like to express my gratitude to all of them.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Wang, Ta-Wei. Ph.D., Purdue University, August, 2009. Essays on Information Security from an Economic Perspective. Major Professors: Jackie Rees and Karthik Kannan.

Information security risks are becoming a critical issue to organizations given the significant impact of security related incidents. In this dissertation, we seek to further our understanding of how information security incidents and security practices affect information security risks.

The first essay proposes a decision tree classification model to investigate how the nature of security risk factors disclosed in financial reports is associated with breach announcements in the subsequent period. We construct and evaluate the model based on the design science principles in Hevner et al. (2004). The model shows that security risk factors with action-oriented terms are less likely to be related to future incidents. We evaluate the model by showing that market participants could better interpret security disclosures at the time when financial reports are released.

The second essay studies how general investors can make better investment decisions regarding security breaches. We explore the association between the textual contents of the news articles about security breach reports and both the stock price and trading volume reactions to breach announcements. The results suggest that general

breach announcements lead to different assessments of the impact of security incidents. However, specific news articles and those about confidential information result in a more consistent negative belief of the impact of security incidents on a firm's future performance. Interestingly, sophisticated investors do not react immediately to breach announcements. By taking advantage of the different perceptions among investors, we show that, on average, one can make about 300% annual profit around the breach announcement date

The third essay investigates the cost and benefit tradeoffs when selecting two-factor authentication systems. We generalize authentication systems into four cases based on the probability of system failure and compare different systems to determine the key factors managers need to consider. This essay proposes that a firm can lower the impact of customer switching by following the larger provider's decision. Also, regulators can encourage the adoption of a more secure authentication system by changing the penalty when the system fails. Finally, it could be preferable to have both one-factor and two-factor authentication systems depending on the customers' characteristics.

CHAPTER 1. INTRODUCTION

Business nowadays relies heavily on information technology to perform daily operations. Because of this increasing reliance on information technology, information security related incidents could result in a tremendous impact on a firm's operation and significant financial losses. For example, a series of Denial of Service (DoS) attacks in 2000 resulted in online retailers and portals such as Amazon.com and Yahoo! losing service for hours (Sandoval and Wolverton 2000). Also, the estimated average loss caused by security breach is approximately $290,000 US dollars per respondent in the CSI/FBI computer crime and security report in 2008 (CSI/FBI 2008). This evidence highlights the importance of information security to organizations which also raises organizational concerns about information security. In order to resolve the concerns and better manage information security risks, researchers and managers have strived to better understand and assess information security risks. For example, companies such as AOL Time Warner and Merrill Lynch have assigned a chief security officer to better understand security risks and to determine the resources needed to manage such risks (Lohmeyer et al. 2002). Furthermore, prior studies have investigated information security from different perspectives such as security standards (e.g., Siponen 2006), security investment (e.g., Gordon and Loeb 2002), and the association between security

breach announcement and business value (e.g., Campbell et al. 2003; Cavusoglu et al. 2004).

This dissertation attempts to further our understanding of information security risks by focusing on two perspectives: the impact of security incidents and the impact of security practices. Specifically, we first investigate the association between security risk factors disclosed that reflect a firm's security practices and the subsequent impact of future breach announcements. Then we further our understanding of the impact of breach announcements on a firm's future performance given investors' reactions to security incidents. Last, we further explore the practice dimension by considering the costs and benefits of implementing two-factor authentication systems.

First, firms disclose information security risk factors based on the firms' internal information regarding vulnerabilities and the firm's security policies and practices. Based on the literature, the disclosures may reduce the uncertainty of the firm's future performance by showing that the firm is well prepared for future breaches. On the other hand, the disclosures could be announced to avoid lawsuits associated with future security breaches. The above two arguments could be valid in the information security context. Accordingly, the security risk factor disclosures could reflect the internal information that is associated with future security incidents and affect the market's assessment about these security risk factors. Therefore, the first essay attempts to understand how the nature of security disclosures is associated with future breach announcements. A classification model is built to investigate the association between the textual contents of security risk factors disclosed and the possibility of future breach announcements. The model is built and evaluated based on the design science guidelines

discussed in Hevner et al. (2004). Different disclosure patterns are explored to provide insights about how market participants could form their perceptions regarding these disclosures and assess the firm's future information security uncertainties.

The second essay further examines investors' reactions to security breaches. Investors' reactions provide explanations to managers and researchers about what leads to the price and volume reactions to security incidents. Also, understanding investors' reactions could help general market participants make better investment decisions by lowering information asymmetry among investors. The association between the textual contents of breach announcements and the price and volume reactions is explored. The result shows how general market participants can adjust their investment decisions regarding breach announcements given the sophisticated investors' reactions. A trading strategy is performed to demonstrate profitable short-term investment opportunities given the information asymmetry among investors.

The third essay focuses on the cost and benefit tradeoffs when selecting two-factor authentication systems. The shift to two-factor authentication system could possibly lower the probability of system failure but might be accompanied with possible privacy concerns and inconvenience. This essay defines the probability of system failure and generalizes all possible combination of authentication systems into four different cases. By comparing the expected costs and losses under these four cases, this essay provides suggestions on whether the new authentication system is more preferable.

This dissertation contributes to the field of information security in the following ways. The three essays provide different perspectives when assessing and understanding information security risks. In particular, essay one emphasizes on what annual report

users should consider when assessing a firm's future uncertainty regarding information security based on the disclosed security risk factors. Essay two formally investigates how investors react to security incidents based on their assessments of the firm's future performance. Furthermore, essay two shows how investors could adjust their investment decisions based on the contents of news articles regarding breach announcements and possible profitable short-term investment opportunities by taking advantage of the information asymmetry among investors. The third essay is the first study that formally considers the selection of authentication system from a generalized and economic perspective. By boiling down the probability of system failure into two broad sets, the third essay is able to compare the authentication system through four different cases and provides suggestions to managers.

The remainder of the dissertation is organized as follows. Chapter 2 describes the first essay. The research framework and both the qualitative and quantitative results are discussed. Chapter 3 presents the second essay where the classification model and the results are discussed in the subsections. The third essay is included in Chapter 4. The basic setting of the model and the propositions are elaborated in the subsections. Chapter 5 concludes the dissertation.

CHAPTER 2. THE ASSOCIATION BETWEEN INFORMATION SECURITY RISK FACTORS AND BREACH ANNOUNCEMENTS: A DESIGN SCIENCE APPROACH

2.1. Introduction

Firms often recognize that information security breaches can impact their performance. Some firms announce risks related to information security publically. For example, Kohl's disclosed in its 2006 annual report that "…. [the company's] facilities and systems…may be vulnerable to security breaches… [which] could severely damage its reputation, expose it to the risks of litigation and liability, disrupt its operations and harm its business" (Kohl's, 2007, p.8). There are two competing motivations from the literature for why firms disclose risk factors. On one hand, the disclosure of risk factors may help reduce the uncertainty that investors have regarding the firm's performance (Jorgensen and Kirschenheiter 2003). On the other hand, a firm may disclose risk factors in order to reduce its future litigation costs associated with adverse events (e.g., Skinner 1994). In the information security context, either motivation may be valid. Some firms are inclined to disclose to indicate preparedness, which corresponds to the first motivation, whereas other firms disclose in order to head off lawsuits, which is the second motivation. Prior literature (e.g., Verrecchia 1983; Dye 1985; Skinner 1994; Kasznik and Lev 1995; Verrecchia 2001) states that a firm's disclosure may vary depending on its internal information which is often reflected in the textual contents of

the disclosure as shown in prior works (e.g., Bettman and Weitz 1983; Abrahamson and Park 1994; Li 2008). Building upon this body of work, we attempt to study how the textual content, or the nature, of information security risk factors disclosed in annual reports is associated with breach announcements.

However, given the lexical nature of the disclosures, it often requires a detailed content analysis to understand the textual contents of the disclosures (e.g., Abrahamson and Park 1994). The detailed content analysis could prevent users from applying and implementing the analysis in their organizational contexts. The model we propose is designed to overcome this problem and to further our understanding of disclosures. Specifically, we propose a classification model for market participants to understand the association between the textual contents of information security risk factors disclosed in financial reports and breach announcements. When constructing and evaluating the proposed model, we follow the seven guidelines of design science research suggested by Hevner et al. (2004). Also, as we are going to show, this model can be operationalized by using readily available software packages.

To the best of our knowledge, the proposed model is the first model that can be used to understand the textual contents of disclosures and associate the disclosure characteristics with the events that might reflect a firm's internal information. By proposing this model, we seek to develop insights into the security attitude of the firm based on the nature of its disclosures. These insights are directly beneficial to investors and debtors, who can take into account this association when evaluating a firm's future uncertainty regarding information security. The cross-sectional analysis validates the usefulness of our model and helps explain how investors update their beliefs of a firm's

future uncertainty regarding information security after breach announcements. Taken together, this study draws upon a diverse set of tools and features both quantitative and qualitative to provide a comprehensive analysis on the nature of information security risk factors.

The rest of the paper is organized as follows. We review the literature on the management and the economics of information security and disclosures in Section 2.2. In Section 2.3, we summarize how the seven guidelines presented in Hevner et al. (2004) are used in our paper. The modeling framework and the data collection process are elaborated in Section 2.4. Next, in Section 2.5, we analyze the textual data of the disclosures and propose our model. We further present the evaluation of the usefulness of the proposed model in Section 2.6. In Section 2.7, we conclude with discussion of contributions, limitations and avenues for future research.

2.2. Literature Review

There are two major streams of literature that are directly related to our study. One is the research stream on management and the economics of information security. The other is the literature on disclosures in accounting.

2.2.1. Management and Economics of Information Security

There is a limited but growing body of knowledge in this stream of research. A few papers have analyzed security investment decisions while a few others have studied the management of information security policies and procedures. Gordon and Loeb (2002),

Gordon et al. (2003), Schechter and Smith (2003), and Gal-Or and Ghose (2005) employ analytical frameworks to study security investment decisions. Tanaka et al. (2005) empirically analyze how vulnerabilities of the firm affect security investments. Goodhue and Straub (1991) show that security concerns vary by industry, company actions and individual awareness. Also, studies (e.g., Straub 1990; Kotulic and Clark 2004; Siponen and Iivari 2006; Siponen 2006) demonstrate the critical role played by information security policies and standards in managing security risks. Often, such investment decisions, policies and actions are closely guarded by organizations in order to avoid exposing their vulnerabilities. By revealing security risk factors in annual reports, but not specific policies, firms convey their internal assessment of the risk factors to the market, as mentioned previously.

Research has also investigated the impact of information security breaches on a firm's business value. Based on different methodologies and different datasets, some papers show that there exists a significant negative impact (e.g., Ettredge and Richardson 2003; Garg et al. 2003; Cavusoglu et al. 2004; Aquisti et al. 2008), while others do not find such impact (e.g., Campbell et al. 2003; Hovav and D'Arcy 2003; Kannan et al. 2007). Although our paper also considers security breach events, we focus on proposing a model to investigate the association between the nature of information security risk factors and subsequent security incidents and show the usefulness of our model by examining how market reactions to security breaches vary with the nature of the disclosure.

2.2.2. Disclosure in Accounting

There is a rich body of literature in accounting that examines disclosures. When there is no disclosure cost, full disclosure exists because investors believe that non-disclosing companies have the worst possible information (e.g., Grossman 1981; Milgrom 1981). However, if disclosure costs or uncertainty exist, companies will disclose only when the benefits exceed the costs (e.g., Verrecchia 1983; Dye 1985). Disclosure may also be used to reduce ex post legal and reputation costs from bad news, or when the firm faces earnings disappointments (e.g., Skinner 1994; Kasznik and Lev 1995; Field et al. 2005). Specific to risk disclosures, one recent study by Jorgensen and Kirschenheiter (2003) formally models managers' decisions on voluntarily disclosing a firm's risks, and they find that firms with smaller future uncertainty will choose to disclose risk factors. Additionally, studies have focused on the quality and credibility of the disclosures (e.g., Lang and Lundholm 1993; Penno 1997; Stocken 2000), the usefulness of disclosures (e.g., Francis et al. 2002; Landsman and Maydew 2002), and other aspects of voluntary disclosures such as expectation adjustment, costs, analysts following, and signaling rationale (e.g., Ajinkya and Gift 1984; King et al. 1990; Lev and Pennman 1990; Elliott and Jacobson 1994; Lang and Lundholm 1996).

In this paper, we link both the aforementioned streams of research. To the best of our knowledge, Sohail (2006) is the only study that has also linked these two streams. In Sohail's paper, he demonstrates that the market values security disclosures, by showing that such disclosures are positively related to stock price at the time when financial reports are released. However, our paper has a different focus. We focus on proposing a model to understand the relationship between security risk factors disclosed in financial

reports (10-K or 20-F for foreign firms) and breach announcements. Specifically, we investigate how the nature of security risk factors disclosed in financial reports is associated with the possibility of future breach announcements. In addition, our paper analyzes how the market reaction to information security breach announcements is dependent upon the nature of disclosure to show the usefulness of our proposed model.

2.3. A Design Science Approach

In this section, we present a summary of how our study is related to the seven guidelines discussed in Hevner et al. (2004) which has been widely applied in different contexts such as customer-centric web sites (e.g., Albert et al. 2004). We will refer to these seven guidelines when we discuss the details about our model in the following sections.

- *Design as an Artifact*. In our study, we propose a classification model as an artifact that is built to show the association between the textual contents of security risk factors disclosed in financial reports and breach announcements.
- *Problem Relevance*. Our classification model provides important insights on understanding the textual contents within security risk factors disclosed in financial reports. The understanding of these risk factors can clarify that a firm discloses security risk factors because the firm either is well prepared for future incidents or attempts to avoid future lawsuits. This clarification helps market participants better evaluate the breached firm's future uncertainty at the time when the financial reports are released.

- *Research Rigor.* To build the model, we draw upon the concepts and theories from information security, economics, accounting, and machine learning to build a classification model and to evaluate the performance and the usefulness of our model. In particular, the theories in economics and accounting about disclosures provide solid ground on our model. The concepts in information security and the methodologies in machine learning and accounting help us reliably construct and evaluate the model. The model construction and evaluation processes are summarized in the following two points.
- *Design as a Search Process*. We build our model based on a three-step process and search for different settings to ensure a best design of our model. We first search for the number of clusters which results in the smallest error rate. Then, we vary the number of observations and perform a 10-fold cross validation to check the robustness of our model. Third, we consider different classification models and compare the results with ours. Last, part of the model building processes is repeatedly performed to further understand the textual contents which demonstrate the relevance of our model.
- *Design Evaluation.* We evaluate our model by using the descriptive method suggested in Hevner et al. (2004). In particular, we build on previous literature in information security, economics and accounting, and machine learning to show (1) the robustness of the model and (2) how the market values these security risk factors at the time when the financial reports are released and after breach announcements. The former evaluates the performance and the robustness of the model while the latter demonstrates the usefulness of our model.

- *Research Contribution*. Our study adds to information security literature by showing that how market participants can better assess a firm's future uncertainty regarding information security based on the contents of security risk factors at the time when financial reports are released by using commercial software packages. Also, the settings of our model can be applied to different voluntary disclosure contexts in order to further understand the impact of public announcements on market participants' decisions.
- *Research Communication*. Consistent with Hevner et al. (2004), this paper aims at both technology- and managerial-oriented audience. For technology-oriented audience, we provide detailed information about the model building processes which enables the model to be implemented. Also, management-oriented audience can easily determine how this model can be used in their investment decisions based on our description.

2.4. Modeling Framework and Data Collection

In this section, we present our modeling framework and discuss the data collection processes. This section communicates to users about the context and the framework of the model (*research communication*) and shows that the data for our model was collected and processed carefully (*research rigor*).

2.4.1. Modeling Framework

Figure 2.1 provides the timeline for our model. The disclosure at time *t* in Figure 2.1 is a list of information security risk factors that may adversely affect the firm's future performance, as reported in the annual report. See Appendix A for an example of a security risk factor disclosed in an annual report. The announcements at time *t* + 1 in Figure 2.1 are the breaches reported in news articles.

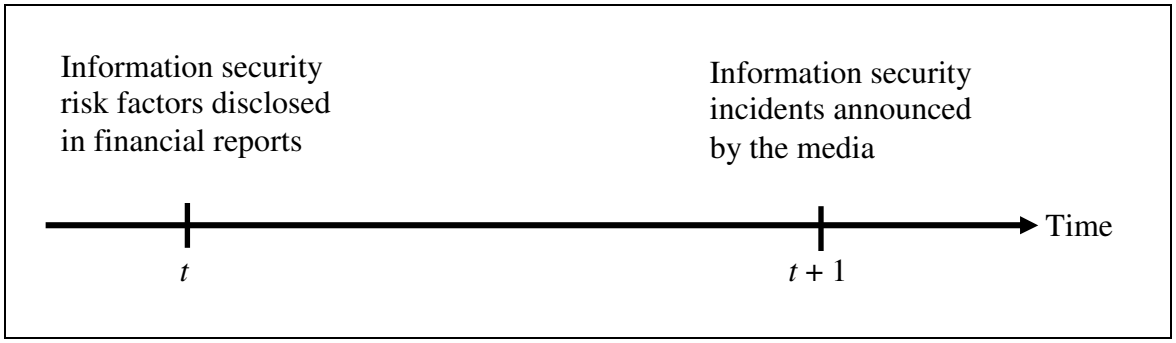

Figure 2.1 Timeline for Two Information Sets

We expect the security risk factors disclosed at time *t* in Figure 2.1 to contain clues regarding the possibility of future security incidents occurring at time *t* + 1. In particular, our proposed model explores the textual content of information security risk factors to show the association between disclosure patterns and the occurrence of future breach announcements. To demonstrate the usefulness of our model, we further analyze the relationship between market reactions to security incident announcements and security risk factors disclosed.

2.4.2. Data Collection

We employ an endogenous stratified sampling method (Cosslett 1981; Manski and McFadden 1981; Cameron and Trivedi 2007) for our data collection. This method is commonly used when the event is rare (compared to nonevents), such as in international relations, wars, venture capital investments, and epidemiological infections (e.g., King and Zheng 2001; Sorenson and Stuart 2001). Estimations using an endogenous stratified sample are more efficient than using a full sample (e.g., Cosslett 1981; Donkers et al. 2003; Imbens 1992). In our context, we employ this sampling method because, as we show below, information security breach announcements regarding publically traded firms are rare. Fortunately, the decision tree method, which we propose in our analysis, can be used with this sampling method. In fact, prior works have shown that decision tree models have a better accuracy rate with endogenous stratified sampling (e.g., Goto et al., 2008; Long et al., 1993; Rudolfer et al., 1999) than logistic regressions. Also, studies such as Zadrozny (2004) and Fan et al. (2005) compare various classification models. They find that that decision tree models perform well even with biased samples.

Our data collection is a three step process. As a first step, we collected the data regarding publically traded firms which have breach announcements between 1997 and 2007 in major media outlets. We searched the *Wall Street Journal*, *USA Today*, the *Washington Post*, and the *New York Times* using the Factiva database as well as the *CNet* and *ZDNet* websites. We used the following search terms: (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 search terms were similar to those

used in prior studies (e.g., Campbell et al. 2003; Garg et al. 2003; Kannan et al. 2007). We screened the news articles and collected only those in which the breach announcement identified the specific date for the security incident, and the breached firm did not have any confounding events, such as earnings announcements, or mergers and acquisitions, around that date. The above process resulted in 101 firm-event observations from 62 firms. The number of observations confirms our argument that information security breach announcements regarding publically traded firms are rare. These incidents correspond to events occurring at time $t + 1$ in Figure 2.1.

As a second step, for each event in the previous step, we gathered the information security risk factors disclosed in the breached firm's annual report (10-K or 20-F filings for foreign firms) published immediately *prior* to the breach announcement using EDGAR Online.[1] Note that some firms did not have any security risk factors disclosed in the annual report while others had several. By this process, we collected 43 risk factors, each corresponding to a breach announcement.[2] These disclosures correspond to period t in Figure 2.1.

In the third step, we need to collect security risk factors from firms that did not have any breach announcements (nonevents). However, one of the main questions with endogenous stratified sampling is how big should the sample size of nonevents be. There is considerable variation in the literature regarding how the total sample should be split between events and nonevents. Breslow and Day (1980) use a 20%-80% split of events and nonevents; Pinczowski et al. (1994) use a 30%-70% split; Rudolfer et al. (1999) use a

[1] http://www.sec.gov/edgar.shtml

[2] Suppose, in a particular year, if a firm has two events, we collected only the disclosure in the previous annual report and counted it as one disclosure in our dataset. Additionally, we counted each of the disclosures separately and ran our analysis, but our results were consistent.

60%-40% split; and Steinberg et al. (2006) and Steyerberg et al. (2007) use a 50%-50% split. Lancaster and Imbens (1991) show that a 50%-50% split is optimal for estimation purposes. Consistent with this work, we also used a 50%-50% split. To check for robustness with respect to the splits, we also studied the performance of our decision tree when subjected to a progressive sampling method (e.g., John and Langley 1996; Provost et al. 1999; Frey and Fisher 1999; Morgan et al. 2003). The details of the robustness check are discussed in Section 2.5.3.

For this third step, we randomly chose 62 firms without any breach announcement between 1997 and 2007. For each of these firms, we randomly picked the annual report from one of the years in the 11 year period (1997-2007) and collected information security risk factors in that annual report. Through this process, we collected 34 risk factors. As before, not all firms had security risk factors in the annual report and a few firms had several. We did not use security risk factors from all 11 years because firms typically tend to add new risk factors to the earlier ones; as a result, using risk factors from all 11 years would lead to oversampling and biasing of our results.

From the above three steps, our dataset involves 124 (62 + 62) firms and 77 (43 + 34) information security risk factors. These firms are distributed across 28 different industries (two-digit SIC code). At the end of 2007, the firms had an average age of 22 years (standard deviation of 19 years), which was calculated based on the year range in Compustat, and average total assets of $2.8 billion (standard deviation of $8 million).

2.5. Classification Model

In this section, we focus on mining the textual data to understand the information conveyed by security risk factors through a decision tree classification model. Text mining, in general, has proven to be a useful tool to extract information through finding nontrivial patterns and trends (e.g., Tan 1999; Feldman and Sanger 2006). For example, text mining techniques have been used in different contexts, such as to classify news stories, summarize banking telexes, detect fraud, and improve customer support (e.g., Young and Hayes 1985; Masand et al. 1992; Han et al. 2002; Fan et al. 2006; Cecchini et al. 2007). In the information security context, we apply text mining techniques to the contents of disclosed security risk factors so as to identify and categorize the elements of security risk factors that might associate with future incident announcements by using a decision tree. We chose to use a decision tree because of the following two reasons. First, the inherent transparency and interpretability of decision tree models help users follow the path of the tree and understand the classification rules step by step (e.g., Kim et al. 2001; Baesens et al. 2003; Zhou and Jiang 2004; Brandãn et al. 2005; Zhang and Zhu 2006). Second, studies such as Goto et al. (2008), Long et al. (1993), and Rudolfer et al. (1999) have shown that decision tree models have a better accuracy rate for our sampling method than logistic models. We also tested other classification models, such as neural networks, and obtained similar results.

2.5.1. Decision Tree Classification Model

In this sub-section, we present our decision tree classification model which is the artifact we propose in this paper (*Design as an Artifact*). The processes we used to build

our classification model demonstrate a rigorous model building procedure by searching different possible settings (*Research as a Search Process* and *Research Rigor*) and provide detail information for users to implement and to apply to different contexts (*Research Communication*). Specifically, we built a classification model by adopting a three-step procedure given in Figure 2.2 and detailed below.

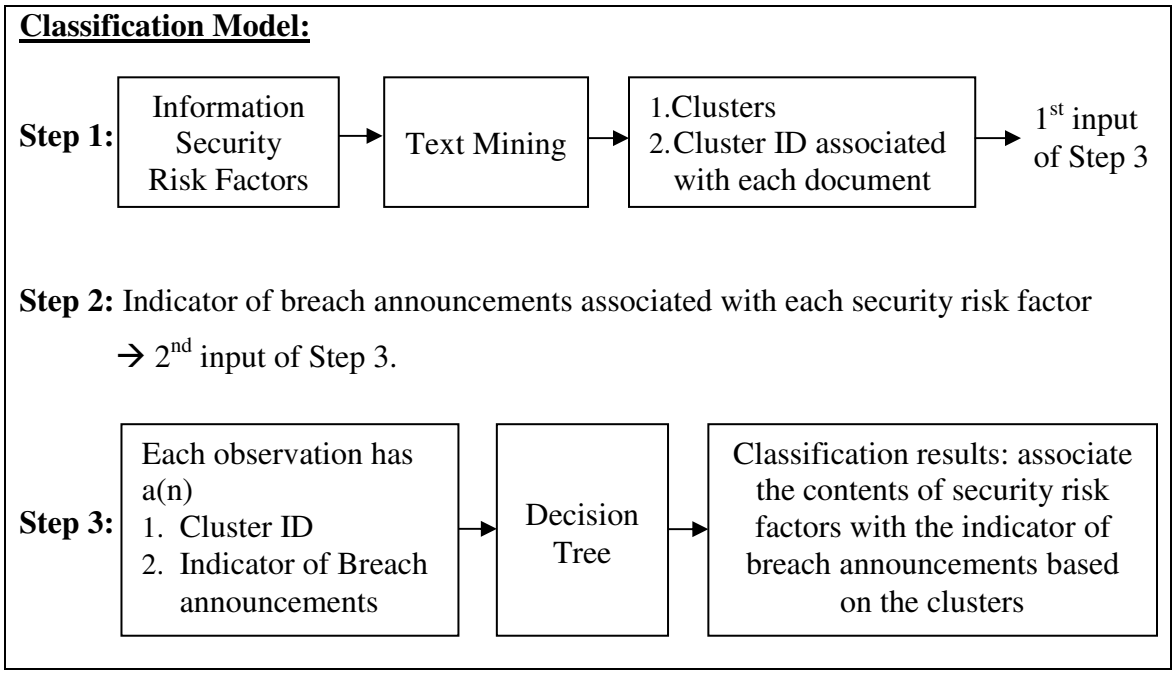

Figure 2.2 Process Flow for the Classification Model

In order to perform the analysis, we used the 77 security risk factors collected. In the first step, we used SAS Text Miner to extract the terms and the associated clusters of these terms for the textual data in the information security risk factors disclosed (the process of identifying the clusters is a standard one and is detailed in Appendix B). Based on these clusters, we assigned each document a cluster ID.

In the second step, we pre-processed the data to make it conducive for the decision tree model. Specifically, we associated each disclosed security risk factor with an

indicator showing that whether the corresponding firm had breach announcement or not. If the security risk factor was from the breached firm, the indicator shows “yes”, and shows “no” otherwise.

In the third step, a decision tree was built to classify the indicator of breach announcements (from step 2) based on the cluster ID (from step 1). In order to perform the classification task, several worth noting settings are as follows. First, we set the clusters that can be found in the first step as four. This optimal number of clusters being four was determined through an iterative process of experimentally varying the number of clusters and repeating the three steps in Figure 2.2 until the error rate of the decision tree model in step 3 is the smallest (e.g., Smyth 2000; Still and Bialek 2004; Tibshirani et al. 2001). Second, the dataset was partitioned into two parts: training (80%), validation and testing (20%). Furthermore, when setting up the classifier, we set the prior probability of the classifier as the proportion of the number of related documents in the whole dataset. The classification model was trained, validated, and tested using a decision tree in SAS Enterprise Miner.

Based on the three steps in Figure 2.2 and after the decision tree model was trained, validated and tested, we found that the resulting tree has two leaves from the root (see Figure 2.3 for an instance).

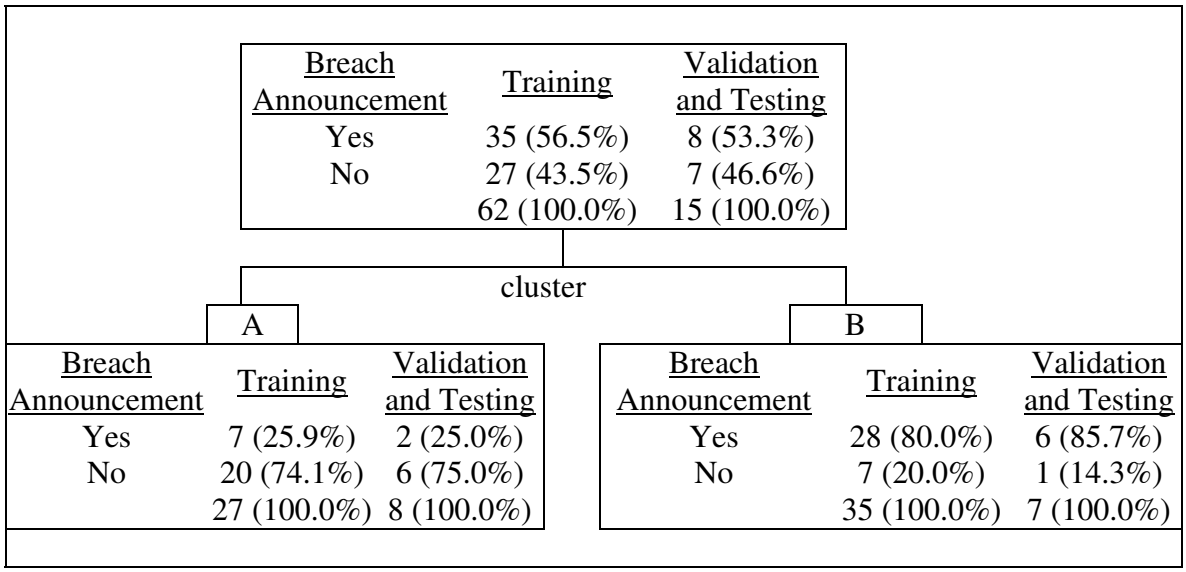

Figure 2.3 An Instance of the Decision Tree

As shown in Figure 2.3, in this instance, 62 and 15 documents were used for training, and validation and testing respectively. Furthermore, documents associated with cluster A were classified into the left sub-tree and about 75% of them in the validation and testing dataset were associated with "no breach announcement". Documents related to cluster B were classified into the right sub-tree and about 86% of them in the validation and testing dataset were associated with "breach announcement". Note that, instead of presenting the results for four clusters, cluster A and B were aggregated clusters and each of them consisted of 2 small clusters. We present our classification tree results by aggregating the four clusters because each of these four clusters has very few data points and is not amenable for further analyses in Section 2.5.2.

To further validate our results above, we used a commonly adopted procedure called 10-fold cross validation (e.g., Weiss and Kapouleas 1989; Kohavi 1995). Accordingly, we repeated the processes 10 times by randomly drawing 80% of the data and averaged

the results from ten different randomly chose training dataset, it was still about 75% of the documents in cluster A that were associated with "no breach announcement" and about 85% of the documents in cluster A that were associated with "breach announcement". The result from one of our 10-fold cross validation is given in Table 2-1. Table 2-1 demonstrates that the overall accuracy rate for this validation result is 77.42% (45.16% + 32.26%). Again, we performed the same process ten times and the average accuracy rate of all ten validation results is about 76%.

Table 2-1 Confusion Matrix of the Cross Validation Results

Frequency Percentage Row Percentage Column Percentage		Predict		
		Breach Announcement	No Breach Announcement	Total
Actual	Breach Announcement	28 45.16 80.00 80.00	7 11.29 20.00 25.93	35 56.45
	No Breach Announcement	7 11.29 25.93 80.00	20 32.26 74.07 74.07	27 43.55
	Total	35 56.45	27 43.55	62 100.00

This model demonstrates that there indeed exist textual differences among security risk factors which associate different possibility of future incidents. Also, it shows that there are cluster A and cluster B that might relate to this different possibility. Two interesting aspects of this model are worth noting. First, the high accuracy rate of the model suggests that the market might be able to predict the impact of the disclosed risk

factors based on the contents disclosed. Accordingly, to evaluate the usefulness of our model to market participants, we show how market participants' can change their perception of security risk factors at the time when financial reports are released based on our model. Details will be discussed in Section 2.6. Second, the model further leads us to explore the characteristics of these two sets of clusters in order to provide detailed explanations of the underlying factors that associate with different future uncertainty in the next sub-section.

2.5.2. Comparison of the Disclosure Groups

We explore the textual content within the security risk factors associated with Disclosure Group A and Disclosure Group B in this sub-section. By doing so, we show what market participants can focus when they look at the security risk factors disclosed in order to better interpret the security risk factors (*Problem Relevance*) and better assess a firm's future uncertainty regarding information security (*Research Contribution*). The exploring of text has long been widely used in different psychological constructs such as therapy transcript (e.g., Peterson et al. 1983) and personality (e.g., Winter 1987). We apply the same concept in the information security context to explore the textual content within security risk factors.

Specifically, we pooled together all the security risk factors from each of the Disclosure Group (Group A or Group B). Then we repeated step 1 in Figure 2.2 twice but now the input was the security risk factors associated with Disclosure Group A and B separately. Through this step, we identified the terms and the associated clusters of textual content that commonly occurred in that group as shown in Table 2-2. Instead of

limiting the number of clusters as in Section 2.5.1., here we explored all the possible clusters in order to have a detailed understanding of the characteristics of these two groups. In Table 2-2, each row represents one cluster and each Disclosure Group has many clusters. Within each cluster, there are five terms with the highest calculated frequency in the cluster (see Appendix C for detail information). A term with the plus (+) sign represents a group of equivalent terms. For example, both "ability" and "abilities" are considered equivalent. The percentage is the frequency of a set of terms divided by the total frequency. The root mean squared standard deviation (RMS Std.) for cluster k equals to $\sqrt{W_k / [d(N_k - 1)]}$, where W_k is the sum of the squared distances from the cluster mean to each of the N_k documents in cluster k, and d is the number of dimensions.

Table 2-2 Text Mining Results of Information Security Related Risk Factors

Cluster	Terms	Percentage	RMS Std.
Disclosure Group A			
1	+damage, ***+implement***, +require a, +resource, +virus	55.7%	0.1113
2	***+act***, +customer, +disruption, ***+prevent***, ***+process***	44.3%	0.1127
Disclosure Group B			
1	+breach, confidential, +harm, +liability, +transmission	19.0%	0.1547
2	+affect, +product, reputation, software, +vulnerability	17.7%	0.1642
3	catastrophic, +earthquake, +facility, +fire, +interrupt	13.9%	0.1523
4	company, +customer, +disaster, +disrupt, infrastructure	13.9%	0.1625
5	+blackout, data capacity, +disaster, terrorism, +virus	12.7%	0.1444
6	basis, +disrupt, +lose, +problem, +system	12.7%	0.1410
7	adversely, code, +program, +sale, +store	3.8%	0.1314
8	+assurance, fraud, internal controls, policy, +statement	3.8%	0.1092
9	+business, +cause, identity, +risk, +theft	2.5%	0.1137
Note: For readers' convenience, we highlight the examples discussed in the text as bolded and italicized.			

From Table 2-2, we compare the clusters of Disclosure Group A and Disclosure Group B and assess the similarity between the clusters. First, Disclosure Group A has only two clusters while Disclosure Group B has nine clusters. So it seems that there are more distinct security risk factors (more clusters) provided by Disclosure Group B than Disclosure Group A. We further look at the terms within these clusters. Since these disclosures are about security risk factors, we do observe terms with negative meanings about risks in the clusters within both disclosure groups such as "damage" and "disruption" for Disclosure Group A and "harm" and "disrupt" for Disclosure Group B. Also, in the clusters within both disclosure groups, we observe the terms about the type of incidents and the subjects that could be affected such as "virus" and "customer". After eliminating these common terms across these two groups, there are still several action related terms in the clusters of Disclosure Group A which are not included in the clusters of Disclosure Group B. Recall that Disclosure Group A corresponds to the *no breach*

announcement group while Disclosure Group B is related to *breach announcement group*. Therefore, it possibly implies that the lack of terms about operations and actions such as "act", "prevent", and "process" (bolded and italicized in Table 2-2) in Disclosure Group B associate with a negative interpretation of the disclosed risk factors. Last, as we observed above, Disclosure Group B provides more information in the disclosures than Disclosure Group A (nine clusters vs. two clusters). As given in Table 2-2, the terms in the clusters of Disclosure Group B point out various information security risks such as "confidential", "virus", "fraud", "identity" and "theft". It seems that, by disclosing various information security risk factors, the firms in Disclosure Group B could possibly avoid future lawsuits which is consistent with our argument that some firms disclose information security risk factors in order to avoid future lawsuits. We will discuss the association between the firms in Disclosure Group B and the litigation risks of omitting material information later in Section 2.6.

In order to provide context to and to better understand the terms in the clusters, we further connected the terms in the cluster with other phrases in the disclosures. This co-occurrence relationship can be captured by concept links (see Appendix C). Concept links provide contexts to the terms in the clusters which help us better explain the terms within each cluster. For example, if we observe the terms "attack" and "denial" are often disclosed together, we are able to understand that the term "attack" in the cluster refers to the context of denial-of-service attacks. We checked the concept links for all the terms in clusters for both groups. For Disclosure Group B, 60% (3 out of 5) of the terms with concept links are general concepts, such as "breach" (see Figure 2.4 for an example), or specific subjects that might be affected such as "data capacity" and "infrastructure". The

rest 40% are terms with negative meanings such as “disaster” and “interrupt”. That is, in the risk factors disclosed by Disclosure Group B (correspond to the breached firms), general security terms or the subjects that might be affected play an important role in conveying information to the public (i.e., generally co-occur with other phrases in security risk factors). However, for Disclosure Group A, all the terms (2 out of 2) with concept links are action terms such as “implement” and “prevent” (see Figure 2.4 for an example). Thus, in the risk factors disclosed in Disclosure Group A, action terms generally co-occur with other phrases in the risk factors. The results from the concept links confirm our findings that the major disclosure characteristic difference between these two groups is: Disclosure Group A uses action terms to disclose security risk factors while Disclosure Group B does not. This characteristic difference can be used by market participants as the focus to distinguish these two groups and determine whether the there is any association between the security risk factors disclosed and future breach announcements.

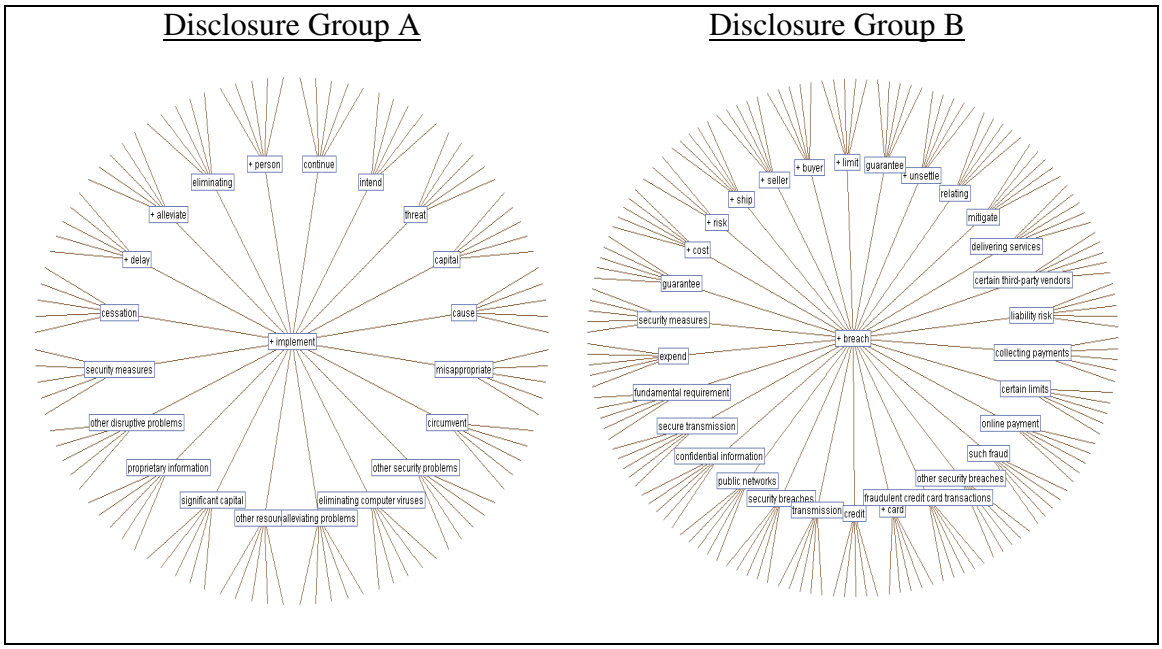

Figure 2.4 Examples of Concept Links

In addition to the comparison we perform above, we also investigated whether there is any change in the security risk factors disclosed before and after the breach announcement for Disclosure Group B to explore how breach announcements affect disclosed security risk factors. Note that since Disclosure Group A is associated with no breach announcement, we can only focus on Disclosure Group B. The results are given in Table 2-3.

Table 2-3 Characteristics of Information Security Risk Factors Before and After Breach Announcements

Cluster	Terms	Percentage	RMS Std.
Before Security Breach Announcements			
1	+breach, confidential, +harm, +liability, +transmission	19.0%	0.1547
2	+affect, +product, reputation, software, +vulnerability	17.7%	0.1642
3	catastrophic, +earthquake, +facility, +fire, +interrupt	13.9%	0.1523
4	company, +customer, +disaster, +disrupt, infrastructure	13.9%	0.1625
5	+blackout, data capacity, +disaster, terrorism, +virus	12.7%	0.1444
6	basis, +disrupt, +lose, +problem, +system	12.7%	0.1410
7	adversely, code, +program, +sale, +store	3.8%	0.1314
8	+assurance, fraud, internal controls, policy, +statement	3.8%	0.1092
9	+business, +cause, identity, +risk, +theft	2.5%	0.1137
After Security Breach Announcements			
1	+business, information, not, security, +service	45.3%	0.177
2	+computer, experience, +failure, +interruption, +result	25.0%	0.171
3	+disruption, +interruption, +loss, +telecommunication, +system	23.4%	0.164
4	+attack, + harm, + have, other, + type	6.3%	0.152

When we compare the clusters before and after the breach announcements in Table 2-3, interestingly, our findings suggest that breached firms still do not disclose with action or process related terms. Similarly, the concept links after breach announcements are still pertain to general concepts, such as “breach” but the number of concept links reduces to one. This observation demonstrates that the risk factors become more diversified after breach announcements (i.e., fewer terms are co-occurred with other phrases in the paragraph) and further validate our observation about the disclosure characteristics of the two groups.

In summary, our proposed model shows that different disclosure characteristics are associated with different indication of future uncertainties. Specifically, we find that when disclosures involve action terms or terms about processes, the disclosures are less likely to associate with the occurrence of future incidents. This result highlights the

importance of our model to market participants. That is, market participants can look for action terms or terms about processes and assess the firm's future uncertainty regarding information security differently than the firm's with general breach information disclosed in financial reports. Also, the high accuracy rate for our classification model indicates that the market can assess the potential impact of disclosures on a firm's future uncertainty regarding information security. However, is the market aware of this link between disclosed information and the possibility of future incidents? We address this question and show the usefulness of our classification model in Section 2.6 by performing a cross-sectional analysis.

2.5.3. Classification Model Robustness Tests

To validate our classification results, we performed the following robustness tests. First, in addition to a binary indicator of breach announcement as the classifier, we also considered using the textual contents from the breach announcements as the classifier. However, we did not find any distinct pattern across breach announcements which might result from the way how the media reported security breaches. Second, we created another "empty cluster" for non-disclosing firms and re-performed the decision tree analysis. The results were consistent but with higher accuracy rate (about 85%). Third, we controlled for (1) industries, (2) the type of breach, i.e., confidentiality, integrity, or availability (e.g., Bowen et al. 2006; Gordon et al. 2006), and (3) historical security risk factor disclosures and our results remained similar. Fourth, we considered controlling for the disclosure patterns for other non-security related risk factors. However, we also did

not find any distinct patterns which might due to the complexity nature of different business risk factors faced by firms across industries or even within the same industry.

Finally, as mentioned in Section 2.4.2., we examined our results by using a progressive sampling method. We randomly sampled 38, 138, and 238 non-breached firms from Compustat to form a total sample of 100, 200, and 300 firms. Similar to the method used in the progressive sampling literature (e.g., Frey and Fisher 1999; John and Langley 1996), we then built our decision tree model based on these three different dataset. We noticed that the accuracy rate for our model increased but in a decreasing rate as the number of observations for non-breached firms (total sample size) increased from 38 to 238 (100 to 300). The accuracy rate only improved less than 2% for last 100 observations. That is, additional observations did not provide significant improvement of the model in terms of the accuracy rate. Therefore, the results from the equal number of firms with and without breach announcements presented above do not qualitatively different from the cases when we had different choice of nonevent sample.

2.6. Design Evaluation: Usefulness of the Model

In this section, we evaluate our model by investigating the usefulness of the classification results through a cross-sectional analysis (*Design Evaluation*). This section also relates to the design science guideline *Research Contribution* by showing how market participants can better interpret the security risk factors disclosed in financial reports.

To investigate the usefulness of our model, we focus on whether the market is aware of the link between the disclosed security risk factors and the possibility of future

incidents. Accordingly, we investigate (1) whether the market values security risk factors disclosed in financial reports at the time when the reports are released, and (2) the association between market reactions to security incidents and security related disclosures. That is, by investigating the market's reactions at time *t* (when the financial reports are released) and time *t* + 1 (when the breach is announced) in Figure 2.1, we are able to understand whether the investors' perception of the information conveyed by the security risk factors is different after the realization of security incidents. Details of our analyses are described below.

2.6.1. Empirical analysis

We first examined the market's reaction at the time when the financial reports are released by replicating Sohail's (2006) model (without the year factor and industry factor since we do not have enough observations each year and for different industries). We found, consistent with Sohail (2006), a positive association between stock price and security risk factors disclosed in financial reports at the time when the reports are released (0.94, about a 2% cumulative abnormal return for a two-day window). However, our result is not significant which might result from our small sample size for this type of regressions. The association was still positive even after considering whether the security risk factors are with or without action oriented terms. This positive association shows that firms with security risk factors are perceived to be prepared to future breaches (the first motivation in Introduction) regardless of the characteristics of the textual contents.

Next, we focused on the time when there is a breach announcement to investigate the association between market reactions to security incidents and security risk factors disclosed. *EVENTUS* was used to estimate the average market reaction (cumulative abnormal return, CAR) around the breach announcement date by applying the market model (see Appendix B). The result shows that the average market reaction to the incidents 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. This market reaction was later used as the dependent variable for the following analyses.

$$CAR_{i,t+1} = \beta_0 + \beta_1 Size_{i,t} + \beta_2 Industry_{i,t} + \beta_3 DAct\_Sec\_Dis_{i,t} + \beta_4 DSec\_Dis_{i,t} + \beta_5 Other\_Dis_{i,t} + \varepsilon_i \quad \text{Eq. 2.1}$$

$$CAR_{i,t+1} = \beta_0 + \beta_1 Size_{i,t} + \beta_2 Industry_{i,t} + \beta_3 DSec\_Dis_{i,t} + \beta_4 Other\_Dis_{i,t} + \varepsilon_i \quad \text{Eq. 2.2}$$

Eq. 2.1 focuses on the association between whether firm *i* had security risk factors with action-oriented terms at time *t* in Figure 1 (*DAct_Sec_Dis*) and the market reactions to security incidents (*CAR*) at time *t* + 1 in Figure 1, where *DAct_Sec_Dis* is a dummy variable, equals 1 if the firm disclosed security risk factors with action-oriented terms at time *t*, 0 otherwise. Eq. 2.2 focused on the relation between market reactions to security incidents and whether the firm had any security risk factor disclosed (*DSec_Dis*). *DSec_Dis* equals 1 if a firm discloses security risk factors, and 0 otherwise. Also, three control variables were used. Firm size (*Size*) and the industry of the firm (*Industry*) were commonly used as control variables since firm size and industry could affect the market reactions. Firm size was measured by the logarithm of the firm's total assets (data item AT in COMPUSTAT) while the industry of the firm controlled for the firms in the

industry of SIC code 73 which were collected from Compustat. We chose to control for the SIC code 73 because about 41% of the breached firms were in this industry category while the rest 60% belongs to 20 other different industry categories. Also, since it seems that the firms in this industry are more frequently breached, they might have different market reactions and disclosure patterns. Last, we controlled for the risk factor disclosing tendency of a firm by counting the total number of risk factors other than security risk factors in financial reports (10-K or 20-F for foreign firms) (*Other_Dis*). These risk factors reflected not only a firm's disclosure policy but also a firm's future uncertainty in general which might also affect an investor's perception regarding the impact of security incidents.

Also, since the results from the classification model suggest disclosure patterns could imply the occurrence of future incidents, we further considered the following two cases when the disclosed concerns were realized in the subsequent incidents (i.e., imply future incident) as in Eq. 2.3 and Eq. 2.4.

$$CAR_{i,t+1} = \beta_0 + \beta_1 Size_{i,t} + \beta_2 Industry_{i,t} + \beta_3 DMatch_{i,t} + \beta_4 DSec\_Dis_{i,t} + \beta_5 Other\_Dis_{i,t} + \varepsilon_i \quad \text{Eq. 2.3}$$

$$CAR_{i,t+1} = \beta_0 + \beta_1 Size_{i,t} + \beta_2 Industry_{i,t} + \beta_3 PMatch_{i,t} + \beta_4 DSec\_Dis_{i,t} + \beta_5 Other\_Dis_{i,t} + \varepsilon_i \quad \text{Eq. 2.4}$$

where *DMatch* is a dummy variable representing whether the disclosed concerns are realized subsequently, equals 1 if there is a match, 0 otherwise; *PMatch* measures the percentage of the disclosed factors that are realized subsequently. The results for the above equations are given in Table 2-4.

Table 2-4 Results for the Cross-Sectional Analysis—Ordinary Least Square (OLS)

Variables	Eq. 2.1	Eq. 2.2	Eq. 2.3	Eq. 2.4
Intercept	-0.0612	-0.0600	-0.0290	-0.0421
Size	0.0027	0.0027	0.0013	0.0019
Industry	-0.0070	-0.0068	-0.0075	-0.0076
DMatch			**-0.0513*****	
PMatch				**-0.0718****
DAct_Sec_Dis	0.0117			
DSec_Dis		**-0.0234***	-0.0048	-0.0109
Other_Dis	0.0006	0.0006	0.0007	0.0006
Adj R^2	0.04	0.04	0.13	0.09
N	88	88	88	88

* significant at 10% ** significant at 5% ***significant at 1%
Note: Since the impacts of consecutive events are not clear, we exclude the observations of consecutive events and follow-up reports such as the denial-of-service attack in February 2000.

In Table 2-4, the insignificant positive coefficient in column 2 (0.0117) for *DAct_Sec_Dis* shows that when the firm disclose action-oriented term in security risk factors, there is not statistically significant association between security risk factors disclosed and market reactions to security incidents which is expected from our classification results. Our results also show a significant negative coefficient of *DSec_Dis* in column 3 (-0.0234), *DMatch* in column 4 (-0.0513) and *PMatch* in column 5 (-0.0718) in Table 4. Comparing this negative result and the positive association between disclosed security risk factors and stock price we found at the time when the financial report are released previously, it demonstrates that the market is not aware that the security risk factors disclosed could associate with the occurrence of future incidents as shown in the classification model section. Instead, the market realizes the interpretation of the security risk factors disclosed at the time when financial reports are released needs to be adjusted with the help of the breach announcements. Also, given this negative association and the firm is still willing to disclose, we believe that the disclosing of

security risk factors is used to avoid future lawsuits which could be larger than this negative association.

Table 2-5 Results for the Cross-Sectional Analysis—Two-Stage Least Square (2SLS)

Variables	Eq. 2.1	Eq. 2.2	Eq. 2.3	Eq. 2.4
Intercept	-0.0612	-0.0544	-0.0266	-0.0376
Size	0.0027	0.0027	0.0012	0.0018
Industry	-0.0070	-0.0066	-0.0075	-0.0077
DMatch			**-0.0535*****	
PMatch				**-0.0822*****
DAct_Sec_Dis	0.0117			
DSec_Dis		-0.0042	-0.0013	0.0013
Other_Dis	0.0006	0.0000	0.0006	0.0003
Adj R^2	0.04	0.01	0.13	0.08
N	88	88	88	88

* significant at 10% ** significant at 5% ***significant at 1%
Note: Since the impacts of consecutive events are not clear, we exclude the observations of consecutive events and follow-up reports such as the denial-of-service attack in February 2000.

The cross-sectional analysis was validated by using a two-stage least square (2SLS) as pointed out by Core (2001) and Leuz and Verrecchia (2000). The result for 2SLS is also given in Table 2-5. As shown in Table 2-5, the variable *DSec_Dis* in column 3 becomes insignificant while *DMatch* in column 4 (-0.0535) and *PMatch* in column 5 (-0.0822) are still significantly negative which confirm our results. However, given the limitation of the number of observations, we acknowledge that this two-stage analysis does not have enough statistical power and need to be interpreted with caution. Last, given our sample size, the relatively high adjusted R^2 for this type of cross-sectional study (e.g., 0.00 to 0.03 for Sivakumar and Waymire (1993) and 0.01 to 0.04 for Brown and Han (2000)) suggests a high explanatory power of our results.

The results so far demonstrate that the market values the disclosures at the time when the financial reports are released but realizes some disclosures are actually released in order to avoid future litigation costs after the breach occurs. In order to further verify this argument, we investigated whether there is any relationship between high litigation risk industry mentioned in the literature (e.g., Francis et al. 1994) and security risk factors disclosed in financial reports.

For our purpose, high litigation risk refers to the lawsuits under SEC rule 10b-5 associated with the situation where the managers "fail to disclose material adverse information" (see Francis et al. 1994, p.1). According to Francis et al. (1994), high litigation risk industries are: (1) pharmaceutical/biotechnology (SIC codes 2833-2836, 8731-8734) (1.86% of the breached firms in our sample), for example, Pfizer, (2) computer (SIC codes 3570-3577, 7370-7374) (47.22% of the breached firms in our sample), for example, IBM, (3) electronics (SIC codes 3600-3674) (2.78% of the breached firms in our sample), for example, Intel, and (4) retail (SIC codes 5200- 5961) (4.63% of the breached firms in our sample), for example, Amazon. Our finding suggests that firms in the high litigation risk industry, on average, disclose more security risk factors than the firms not in such industry ($t = 1.69$, $p < 0.10$). Also, the number of security disclosures can increase the probability that a firm is in a high litigation risk industry by 0.818 ($p < 0.01$). These results somehow confirm our argument that some firms disclose in order to avoid future lawsuits.

The findings in this section suggest that our proposed model can be useful for market participants to understand the security risk factors disclosed in financial reports at the time when the financial reports are released. Specifically, market participants can

determine the disclosing motivation of the firms without the information of future breach announcements and better evaluate the firm's future uncertainty regarding information security.

2.6.2. Empirical Model Robustness Tests

We performed several robustness tests to verify our cross-sectional results. First, since the average market reaction is not zero, we also used the Fama-French three factor model (see Appendix B) to estimate the market reaction and perform the same set of analyses (e.g., Brown and Warner 1985; Fama and French 1992). Our results were largely the same. Second, we additionally controlled for the following variables that could potentially affect market responses to security incidents: attack history, incident types (namely, confidentiality, integrity, and availability type incidents), previous disclosure patterns, i.e., the number of security risk factors disclosed one year before the annual report we considered, and the time (in months) between annual report release date and breach announcements. Our results remained similar. Last, we validated our results by verifying if our results also hold for other firms without any reported incidents (see, for example, Shadish et al. 2002). We determined, for every firm in the experimental group, one of its publicly-traded competitors that did not have any breach announcements from Yahoo! Finance and the Hoover's Database. We then performed the same analyses but did not find any significant results. Therefore, we can rule out other possible explanations and make sure that we have captured the relationship between security disclosures and incidents.

2.7. Conclusions and Discussion

We have often observed that firms disclose information security risk factors in the financial reports. However, as mentioned in the Introduction, it was not ex ante clear whether the disclosures have a positive (e.g., preparedness for such threats) or a negative (e.g., indicates potential litigation/reputation costs) impact to the firm's business value. Given the complexity of the nature of disclosures, it is often difficult to perform content analyses but keep the tool applicable to different organizational contexts. In order to clarify the issue mentioned above with the easiness of implementation, our paper proposes a classification model following the design science guidelines presented in Hevner et al. (2004) to investigate the relationship between the textual contents of information security risk factors disclosed in the financial reports and the possibility of future incidents. The proposed model demonstrates that the textual content of security risk factors is a good predictor of future breaches. Building on this, we further consider the characteristics of security risk factors. We argue that firms, which disclose more actionable information when they provide information security risk factors, are less likely to be associated with security incidents in subsequent period.

Next, we evaluate the usefulness of our model by examining how the market reacts to these disclosures and how the classification results can help the market better interpret the security disclosures. Through cross-sectional analyses, we find that, the market is not aware of the link between security disclosures and future incidents as shown in our model. Instead, the market values these disclosures at the time when financial reports are released. However, after security breaches occur in the subsequent period, the market realizes that the disclosures are not all credible as it initially perceives. These results

indicate that some disclosures are actually warnings of future incidents in order to avoid future litigation costs.

Our results and analyses suggest that the market participants could re-consider the meaning of these security disclosures when evaluating a firm's future performance and uncertainty regarding information security. Also, the proposed classification model could be applied to other disclosures in order to understand the impact of disclosures on a firm's business value. Last, the results and analyses shed light to a manager on how they can convey security practices to their customers and investors more effectively. By properly reflecting possible security concerns, a firm should be able to convey its security practices and concerns to investors without being considered as a warning of subsequent incidents.

Our paper is not without its limitations. One of the major limitations of our study is sample size for security incidents. Although we attempt to capture as large of a sample as possible, it is still problematic to collect a larger dataset base on our filtering processes. A larger dataset for security incidents might allow us to have better estimates in the cross-sectional analysis section. Furthermore, many firms might suffer from information security incidents that are not disclosed to the public. Obviously, we are unable to incorporate this information into our sample. Second, we implicitly assume that the stock price truly reflects a firm's business value. Although the stock price for high-tech firms might be biased, we only look at the price change in a short time period. Thus, we believe that our results still hold even with this possibility that the high-tech firms' stock price is not fairly reflected. Third, we adopt a simple coding scheme for the disclosures. Although we believe that a more complicated coding scheme does not alter

our main results, a finer coding scheme for all the disclosures that can be applied to different industries may provide more details than the present scheme. Last, our model for the cross-sectional analysis implicitly assumes that the disclosures affect CARs which is typical in the literature. However, the disclosures can affect the CARs and the CARs also affect a firm's subsequent disclosure decisions. Our model does not capture this interaction effect which is still an open question in the disclosure literature.

Possible future extensions are as follows. First, in our paper, we implicitly assume that the disclosures are creditable and truly reflect a firm's practices. However, some firms might disclose lots of information but invest little. On the other hand, some other firms might invest substantially in information security but refuse to disclose such investments to the public. Therefore, this anomaly is worth further investigation. Second, a larger dataset can be used to provide more meaningful text mining results for both information security risk factors and business risk factors. The text mining analysis of business risk factors can also provide a first glance on how these risks affect different businesses. Last, as different media becomes popular information sources for investors, we can further consider other media sources, such as blogs, to investigate the relationship among different information sources, information security incidents, and stock price reactions.

CHAPTER 3. THE TEXTUAL CONTENTS OF INFORMATION SECURITY BREACH REPORTS AND PROFITABLE SHORT-TERM INVESTMENT OPPORTUNITIES

3.1. Introduction

Information security related incidents often lead to a disruption of business and cause significant losses (CSI/FBI 2008). For example, security incidents could affect business operations and result in a loss of a firm's reputation (e.g., Glover et al. 2001; Warren and Hutchinson 2000). Given the potential threats posed by information security incidents on a firm's operations, it is important for market participants to understand how information security breaches would affect a firm's future performance in order to make investment decisions. However, for general investors (or so-called "unsophisticated investors"), the only information source they can use to determine the impact of security incidents on a firm's future performance around the breach announcement day is the media information about security breaches, such as news article and blogs, and the corresponding stock price as well as the trading volume reactions. The stock price and the trading volume reactions to breach announcements provide both the aggregate reaction of the market and the different individual investor's reaction to security incidents (e.g., Bamber and Cheon 1995). Therefore, the stock price and the trading volume reaction could let the unsophisticated investors understand the aggregate market reaction and whether market participants assess information security incidents differently.

"Sophisticated investors", different from unsophisticated investors, such as analysts and investment institutions, are the investors that have firm specific knowledge about the firm's operations, more information sources and superior capability of processing information (e.g., Bhushan 1989; Lakonishok et al. 1992; Francis et al. 1997; Roulstone 2003). Accordingly, the sophisticated investors might be able to assess the impact of security incidents on a firm's future performance more accurate than general investors. Therefore, if the sophisticated investors react differently than the overall market, could general investors take advantage of this information and trade by considering the sophisticated investors' perspective?

Based on the discussion above, in this study, we address the following two questions. First, do different information security breach reports lead to different investors' assessments of the impact of security incidents on a firm's future performance? Specifically, do certain characteristics within the news articles 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, for general investors, are there any profitable short-term investment opportunities around the breach announcement date?

In order to approach our research questions, we first collect the news articles about security breaches and estimate the corresponding stock price and trading volume reactions to security breaches. Then, we use text mining techniques to explore the characteristics within the news articles. The characteristics are later associated with the corresponding stock price and trading volume reactions by using a decision tree classification model. The sophisticated investors' reactions to breach announcements are

investigated and compare to the classification results. The comparison result shows profitable short-term investment opportunities after the breach announcement.

Our results demonstrate that the stock price and the trading volume behavior around the breach announcement day are associated with the textual contents of the news article. In particular, news articles that have specific information regarding the incident such as the subject affected, or news articles about confidentiality type incidents or about identity theft often lead to a negative stock price reaction but small trading volume reactions. However, breach announcements with unclear incident information could result in different beliefs of 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 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 findings suggest that market participants could re-evaluate a firm's future uncertainties regarding information security from the sophisticated investors' perspective and the textual contents of the news articles about security breaches. Also, firms could focus more on conveying the breached information to the public which might affect the magnitude of the temporary drop of the firm's stock price around the breach announcement date.

The remainder of the paper is organized as follows. We review related literature in information security, trading volume behavior and analysts' forecasts in Section 3.2. The theoretical framework and our data collection process are presented in Section 3.3. We

text mine the news articles about breach announcements and investigate the association between the contents of the news articles and the price and voluem reaction in Section 3.4. In Section 3.5, 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 3.6.

3.2. Literature Review

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

3.2.1. Information Security

Studies have investigated information security related issues from different perspectives such as information security policies (e.g., Straub 1990; Siponen and Iivari 2006; Siponen 2006) and information security investments (e.g., Gordon and Loeb 2002; Gordon et al. 2003; Schecter and Michael 2003; Gal-Or and Ghose 2005). However, studies that are directly related to our paper are about the impact of information security breaches on a firm's performance and uncertainty. For example, Glover et al. (2001) discuss the impact of information security breaches on business operation, including physical and intangible impacts. Also, various papers have investigated the association

between security breach announcements and a firm's business value. Some of the results show that there exist significant negative impacts (e.g., Aquisti et al. 2008; Cavusoglu et al. 2004; Ettredge and Richardson 2003; Garg et al. 2003), while others do not find such impact (e.g., Campbell et al. 2003; Hovav and D'Arcy 2003; Kannan et al. 2007). The inconclusive results of the impact of security breaches on a firm's future performance (or business value) from the above studies point out the need to explore in more detail the investors' reactions to security incidents and information asymmetry among investors. Moreover, since the sophisticated investors have more information sources and more understanding of the firms, their viewpoint of security breach announcements could help us better understand the impact of security incidents.

3.2.2. Trading Volume

The discussion of trading volume can be traced back to Beaver (1968) who shows that earnings announcement generates not only abnormal price changes but also high trading volume. According to the literature, the stock price change reflects the change in market's average beliefs aggregately while the trading volume behavior is the sum of all individual investors' trades (e.g., Kim and Verrecchia 1991; Bamber 1987; Bamber and Cheon 1995). That is, the trading volume behavior keeps the counterbalanced beliefs among individual investors (e.g., Bamber and Cheon 1995). Accordingly, the association between the inconsistent of beliefs and trading volume demonstrates that a subset of investors have the advantage in processing the information or different beliefs regarding the information announcements (Morse 1981; Karpoff 1986; Kim and Verrecchia 1991, 1994, 1997; Bamber and Cheon 1995; Bamber et al. 1997; Easley and O'Hara 1987;

Hasbrouck 1988, 1991; Bhattacharya 2001). For example, Kim and Verrecchia (1991) analytically show that the trading volume behavior results from different quality of the information acquired and initial beliefs among investors. In this essay, we apply this concept in the context of the announcements of information security incidents and investigate the different reactions among investors based on their different information processing capabilities. Also, Bamber and Cheon (1995) investigate whether different price and volume reactions are associated with different earnings announcement characteristics such as the standard deviation of analyst forecast and the market value of the firm. In this essay, we adopt a similar concept to investigate whether different price and volume reactions are associated with various textual characteristics within the news articles about security incidents.

3.2.3. Analyst Forecast

Analysts collect information of a firm from various sources and provide information such as transaction recommendations and the prospects of the firm to some market participants in a timely manner (e.g., Bhushan 1989; Lev and Thiagarajan 1993; Francis et al. 1997; Roulstone 2003). In the literature, the role played by analysts in the market can be used as proxies such as informed traders and signals of information asymmetry because of their information processing capabilities and communication with the firms (e.g., Francis et al. 2002; Roulstone 2003; Core 2001). In this essay, the analysts' superior capabilities of processing information and their understanding of the firm are used in the context of information security breaches. In particular, given the analysts' capabilities and the 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 of a firm can be determined by several firm characteristics such as firm size and return variability (Bhushan 1989). The number of analysts following could 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 et al. 2004), 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 such as how analysts formulate their expectations about firms' earnings, how to improve the forecasts or the determinants of analyst research (e.g., Kross et al. 1990; Stickel 1990; Elgers and Murray 1992; Brown 1993; Barth et al. 2001; Frankel et al. 2006). Analyst forecasts are also commonly used as a reference point when calculating earnings surprises (e.g., Ayers et al. 2006; Barron et al. 2008; Kasznik and Lev 1995) and when investigating whether firms attempt to manipulate their earnings (e.g., Beneish 2001; Degeorge et al. 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 paper, analyst forecasts are served as the reference point of the impact of security incidents on a firm's future performance from the sophisticated investors' perspective.

3.3. Theory and Data Collection

In this section, we first describe the theory we use. Then the data collection processes are presented. Specifically, we first identify the breached firms. Based on the breached firms identified, we investigate both the stock price and trading volume reactions to breach announcements. The data collected will be the inputs for the classification model (Section 3.4). Also, based on the breached firms identified, in Section 3.5, we investigate the sophisticated investors' reactions and propose a profitable short-term investment opportunity for unsophisticated investors.

3.3.1. Rational Expectation Model

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

In our context, the breached news article is the public announcement that changes investors' assessments about a firm's future performance regarding information security. General investors make their investment decisions based on this public announcement and the associated price and volume reactions to the announcement. In contrast, sophisticated investors have better understanding of the firm's operation (different initial belief) and have superior capability of processing information (better information) than general investors. Their response to security breaches could be useful for general investors when make investment decisions around breach announcement day. Accordingly, as discussed in the Introduction, we would like to understand how the textual contents of the breached news articles affect the price and volume reactions by considering sophisticated investors reactions to security breaches in order to help general investors make better decisions.

3.3.2. Sample

To approach our research questions, we searched for news articles between 1997 and 2008 about breach announcements 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 were similar to those used in prior studies (e.g., Campbell et al. 2003; Garg et al. 2003; Kannan et al. 2007; Wang et al. 2008). We only included the news

articles about publicly traded firm with specific event date after ruling out the observations with confounding events, such as earnings announcements and mergers and acquisitions. For the following analyses, we excluded the consecutive-attack observations except the first day, such as the series of denial-of-service (DoS) attack in 2000, and the observations without trading data and analyst forecast data. The resulting sample size was 89 firms. We stored the content of the news articles for our analyses in Section 3.4.

3.3.3. Price and Volume Reactions

In addition to the breached news articles stored above, we investigated the price and volume reactions to breach announcements as the inputs to our classification model. We considered both the stock price and the trading volume behavior since 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 market reactions which are discussed in detail in Appendix B.

The result shows that the average market reaction to the incidents 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, there is a negative stock price reaction to security breach announcement.

For the trading volume behavior, we first investigated the trading volume changes across time around the breach announcement date (i.e., a three-day window as the stock price reaction) by controlling for the market effect as detailed in Appendix B. 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 that controls for firm-specific effect also shows that, on average, the trading volume is 13.62% more ($p < 0.05$) than the usual trading volume after breach announcements. As discussed in the literature review, the significantly increased trading volume behavior shows that investors have different beliefs of the impact of security breaches on a firm's future performance and some investors are able to better process the information about security breaches. Accordingly, general investors could take advantage of this difference and have profitable investment opportunities which will be investigated in Section 3.5.

Based on the inputs above, in Section 3.4, we explore the textual contents of the breached news articles and show how the contents affect the price and volume reactions by using a decision tree classification model.

3.4. Classification Model

3.4.1. Decision Tree Classification Model

In this section, we first text mine the textual contents of the news articles about security breaches. Then we associate the characteristics within such news articles 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 improve customer support (e.g., Young and Hayes 1985; Masand et al. 1992; Han et al. 2002; Fan et al. 2006; Cecchini et al. 2007). In our

context, we apply text mining techniques to the contents of news articles to investigate how this publicly available information regarding security breaches is associated with the stock price and trading volume reactions to breach announcements. As we are going to 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., Kim et al. 2001; Baesens et al. 2003; Zhou and Jiang 2004; Brandãn et al. 2005; Zhang and Zhu 2006). 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). Since 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.

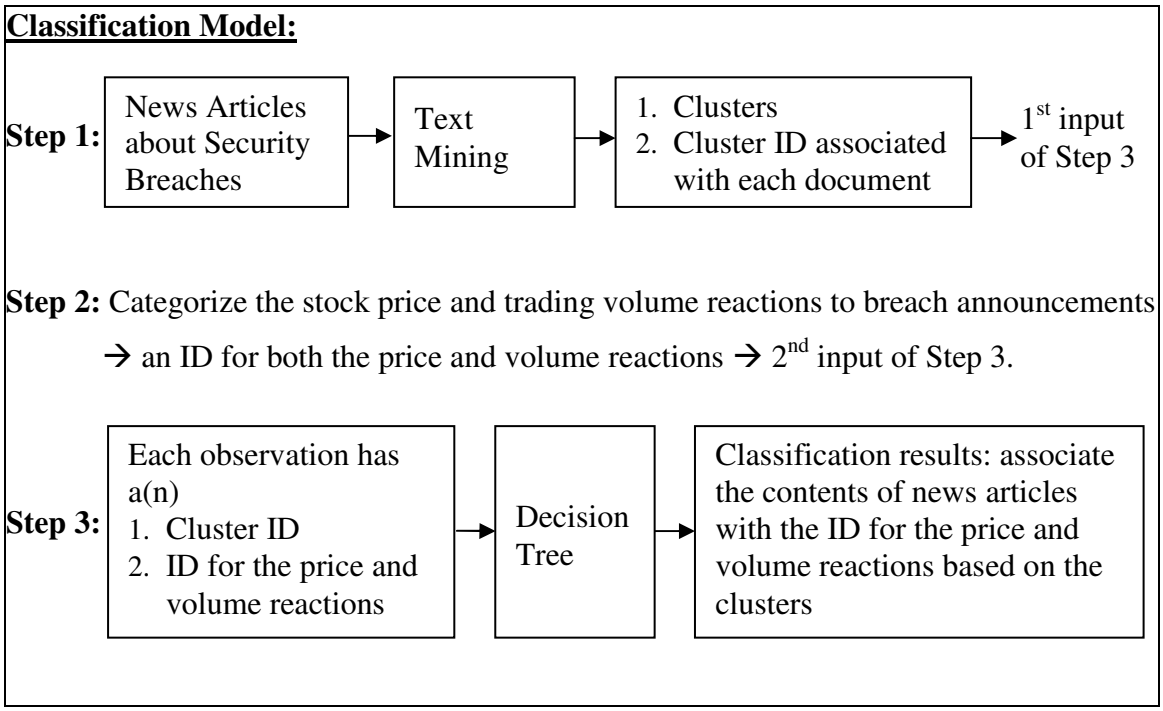

Figure 3.1 Building Process of the Classification Model

We have a three-step process (as given in Figure 3.1) to build the decision tree model which is presented in detail in the following paragraphs. First, recall that from our data collection process, we stored 89 news articles about breach announcements. These 89 announcements were input into SAS Text Miner to categorize them into clusters. The settings of the cluster analysis in SAS Text Miner are summarized as follows. The Text Miner decomposed 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 since 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 $1 + \sum_j \left(p_{ij} \log_2 \left(p_{ij} \right) / \log_2 n \right)$, 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 Inc 2004). We also consider assigning equal weights to different terms and our results are qualitatively similar. Accordingly, we will only present the results based on the logarithm calculation of the weight in the following sections. For dimension reduction, we used the standard single value decomposition (SVD) method. We considered different levels of the reduced dimensions in our analysis and the results are similar. The resulting SVD dimensions were further used for calculating the clusters of news articles by the standard expectation maximization method (SAS Institute Inc 2004). Here, we determined the number of clusters to be four by experimentally varying the number of clusters until the root average squared error of the decision tree model (discussed later) was the smallest (about 0.39) (e.g., Smyth 2000; Still and Bialek 2004; Tibshirani et al. 2001). However, since three of the four clusters did not have enough announcements (12, 14, 18 announcements in each of the three clusters respectively) for further analyses, we chose to group them into two when we present our results. The output was a cluster ID associated with each news article. This cluster ID will be the classifier in our decision tree model.

The second step is about the stock price and trading volume reactions to 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 breach announcement date. We experimentally varied the number of clusters and

observe that the stock price and trading volume reactions can converge into two or three major clusters. However, when there were three clusters, one cluster had only 1 observation and the news articles in other two clusters were the same when there were only two clusters. Therefore, we presented our results when there were two major clusters. Also, we considered using the discriminant analysis and had the same result. The standard K-means cluster analysis converged into two clusters (labeled as Reaction Group 1 and Reaction Group 2) after 7 iterations when there was no change in the cluster center. Reaction Group 1 has 63 observations with an average (standard deviation) of stock price reaction of -0.002 (0.0317) 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) and trading volume of 145.008% (31.816%) than usual. Further analysis shows that the breach announcements in Reaction Group 1 result in a significant high 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 2 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 news articles. Several settings of our classification model are as follows. First, the dataset was randomly partitioned into two parts: training (80%), validation and testing (20%). Second, we set the prior probability of the classifier as the proportion of the number of related documents in the whole dataset. Third, we used the Chi-squared test of a significance level of 0.2 as the splitting criteria. Fourth, as suggested by Berry and

Linoff (1999), the minimum number of observations in a leaf was set to be 1% of the dataset. The classification model was trained, validated, and tested using a decision tree in SAS Enterprise Miner.

A decision tree was built to classify the two market reaction groups based on the cluster ID. The classification results (an instance of the decision tree) are given in Figure 3.2.

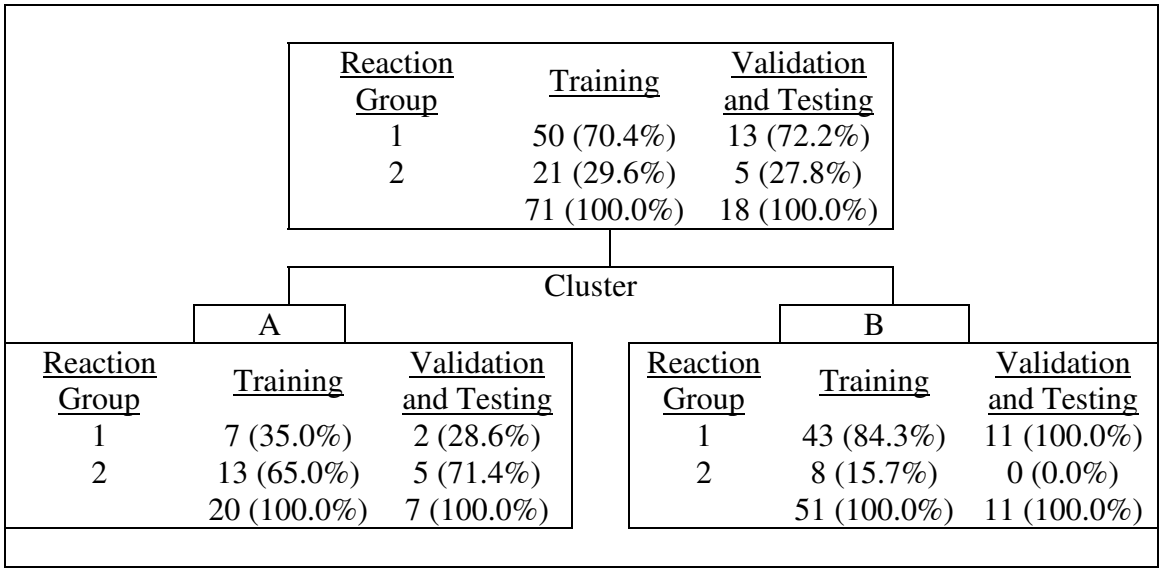

Figure 3.2 An Instance of the Decision Tree

As shown in Figure 3.2, 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). There are two branches in Figure 3.2. The left branch is associated with cluster A and with Reaction Group 2 71.4% of the time in the validation and testing dataset. The right branch is associated with cluster B and with Reaction Group 1 100% of the time in the validation and testing dataset. Since the number of observations in our validation and testing dataset is small, we further verify our results by a commonly

adopted procedure called 10-fold cross validation (e.g., Weiss and Kapouleas 1989; Kohavi 1995) was used. When we repeated our procedure ten times by randomly drawing 80% of the data and averaged the classification results across ten different runs, the associations are similar for both the left and the right branch. The result from one of our 10-fold cross validation is given in Table 3-1. Table 3-1 demonstrates that the overall accuracy rate for this model is 71.83% (21.13% + 50.70%). Similarly, we repeated the process ten times and the average accuracy rate of all ten validation results is about 70%.

Table 3-1 Confusion Matrix for the Cross Validation Results

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

Recall that Reaction Group 1 is with significant negative stock price reactions but an insignificant small trading volume. Reaction Group 2 is with a significant 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 news articles in Reaction Group 1 and Reaction Group 2. The exploration of text has long been widely used in psychological constructs such as therapy transcript (e.g., Peterson et al. 1983) and personality (e.g., Winter 1987). In this essay, we apply the same concept and explore the terms within the news articles about security breaches.

We pooled together all the announcements associated with cluster A or cluster B (labeled as dataset A and dataset B). Then we performed a cluster analysis by repeating the first step in Figure 3.1 and using SAS Text Miner again to obtain all the possible clusters based on these two datasets. The settings and procedures are the same in the first step when building the decision tree except that we do not limit the number of clusters this time. Table 3-2 shows all the possible clusters for dataset A and dataset B. In Table 3-2, each row is a cluster. Within each cluster, there are five terms. The terms with plus (+) signs means 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 cluster k equals to $\sqrt{W_k/[d(N_k-1)]}$, where W_k is the sum of the squared distances from the cluster mean to each of the N_k documents in cluster k, and d is the number of dimensions.

Table 3-2 Terms in Dataset A and Dataset B

Cluster	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.			

We then compare the clusters and terms associated with dataset A and dataset B in Table 3-2. Both datasets have two clusters. However, when we investigate the terms within the clusters, 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 reports and are not specific to certain incidents. Accordingly, by looking at the terms in dataset A, the information regarding the incident is not clear. 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, the terms such as "credit card", "account" and "data" are related to confidentiality type incidents or 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 the specific terms or the terms about confidentiality type incidents result in a more consistent negative price reaction. This result is intuitive because, with the specific description in the news articles, the detail of the security breach and how the loss of confidential information would direct affect a firm's customer are clearer which result in a negative impact on a firm's future

performance. However, the unclear information and the general description in the breach announcement leads to different interpretations and assessment of the impact of security breach. Since based on the general information in the news articles, market participants are not able to have a clear understanding of how the breached firms will be affected.

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

3.4.2. Robustness Tests

We performed the following tests to verify our results. First, we considered using industry, incident types, attack history, composition of the investors, and market value of the firm as the classifiers. For industry, we controlled for the firms with two-digit SIC code 73 since about 40% of the firms in our sample are within this category and our results remain similar. For incident types, we considered confidentiality, integrity, and availability type incidents and our results remain similar. 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 and how many of the shares outstanding were held by institutional investors but our results remain

similar. We also took into account the firm value which is the market capitalization one day before the breach announcement and our results remain similar.

Second, as pointed out by Wang et al. (2008), 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.

3.5. Investment Opportunity

In this section, we first investigate sophisticated investors' reactions to breach announcements. Then we compare such reaction to the classification results and show profitable short-term investment opportunities for general investors.

3.5.1. Sophisticated Investors' Reactions to Breach Announcements

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

Analyst forecasts data were 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 announcements, 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 comparing 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 is 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. Though studies such as Ivkovic and Jegadeesh (2004) show that about 20% to 26% of analyst revisions of earnings estimate are issued at the earnings announcement date and the following two days, some studies uses a three-week period (e.g., Bowen et al. 2002). To be conservative, we also searched for all possible forecast revisions within three weeks after the security incidents. If there was any revision, it was attributed to the incidents after controlling for all other announcements such as the announcements of mergers and acquisitions by searching for 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. Though 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 the position.

Our results show that about 33% of our observations have some analyst forecast revisions after the breach announcement. However, interestingly, none of these forecast revisions can be associated to security incidents. Second, for institutional ownership, we do not observe any significant change ($p > 0.10$) before (about 62%) and after (about 64%) breach announcements. The above findings suggest that the sophisticated investors might not consider information security breaches as an event that will significantly affect a firm's future performance 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 comparing results confirm our results and demonstrate that, without other future events, the firms' average performance is greater than the average analysts' forecasts (0.02, $p < 0.05$).

In order to rule out possible 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 find any significant increase in trading volume. Also, 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 results above, we interviewed two analysts and two investment portfolio managers to investigate their reactions to breach announcements to provide more insights to our findings. From their viewpoints, two major reasons why they do not react negatively to breach announcements immediately are as follows. First, though they do care about confidentiality type incidents, the information regarding the security incident around the breach announcement day is not clear. It might require more time before the detailed breached information 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 operation advantages and disadvantages. Therefore, the impact of security breaches on a firm's future performance should be evaluated on a firm basis in order to have a better understanding. Since the second point requires further case studies on various firms from different industries to provide more insights, we leave it as future research avenues. 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 till the end of 2008. Among the 32 observations of confidentiality type incidents in our sample, we found 1 analyst report discussing security breaches of T J Maxx. Though T J Maxx suffered from the credit card data losses in early 2007, one analyst considered this event as a bearish cause to the stock price in the report in June 2008 (two months before the alleged hackers were arrested). That is, the event was considered after 18 months when the breached information was clarified. However, we did not find other analyst reports for the rest observations regarding confidentiality incidents in our sample. The possible reasons for no future analyst reports could be the

firm characteristics mentioned above or 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, the above 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.

3.5.2. Profitable Short-Term Investment Opportunities

The above results demonstrate that there exist different assessments of the impact of security incidents on a firm's future performance among investors around the breach announcement day. Furthermore, the textual contents of the news articles regarding breach announcements are associated with both the stock price and trading volume reactions. Also, given that sophisticated investors do not react to breach announcements immediately after the breach announcement, the negative stock price in Reaction Group 2 in our classification model is driven by unsophisticated investors. Since unsophisticated investors could only temporary affect stock price (e.g., Bamber and Cheon 1995), the negative stock price reaction is only temporary. Therefore, it is possible that the general investors could take advantage of this reaction difference among investors and have profitable short-term investment opportunities.

In order to demonstrate that the profitable short-term investment opportunity exists and to support our argument about the temporary stock price drop above, we first used implied volatility, which is the theoretical volatility used in the option pricing model (see

Appendix D), to examine the change of volatility after breach announcements. Based on our argument, we should observe a decrease in volatility after breach announcements (i.e., the stock price return to its normal state). The Implied volatility has been shown to be a good prediction of the firm's future volatility till the expiration date of the option (e.g., Harvey and Whaley 1992; Sheikh 1989; Christensen and Prabhala 1998), so it can also be a good prediction of the firm's future volatility after breach announcements. We obtained all the call option and put option data for the breached firms in our sample from the database *OptionMetrics*. Consistent with the period we investigated for analyst forecasts, for each firm, we selected the options that have the expiration date close to the end of the quarter when the incidents occur. Then, we calculated the average change in implied volatility after the breach announcement. Specifically, we compared (1) the average of the implied volatility of a firm's option that will expire around the subsequent quarterly end one trading day before breach announcements and (2) the average one trading day after breach announcements. The results show that the implied volatility decreases about 1.26% ($p < 0.05$). The above result leads us to believe that, consistent with our results from previous sections, in the short-run, the breached firm might suffer from a decrease in business value after breach announcements. However, in the long-run, the breached firms' business values will restore to the normal state, other things being equal.

Based on the results of implied volatility and previous sections, we performed 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 the breach announcement, for those breached firms that encountered a negative stock price reaction after breach announcements. We only focus 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 verify our positive trading strategy and further confirms our observation that the stock price fall around breach announcement date is only temporary.

3.6. Conclusions and Discussion

Our results suggest that the contents of the news articles about security breaches are associated with the stock price and trading volume reactions. However, sophisticated investors do not react to security incidents around the breach announcement day. Given the different perceptions among investors, we form an actual investment strategy and show that there exists profitable (on average) short-term investment opportunity for general investors after breach announcements.

This study adds to the literature of information security by further investigating investors' reactions to security incidents. 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 decisions. For managers, our results suggest that allocating lot more resources to information security investment might not be an effective way to lower the

impact of information security incidents on the firm's business value. Instead, response properly to security incidents can lower the information asymmetry among investors which in turn could lower the temporary negative impact of incidents.

There are several limitations of the paper. First, the sample size is relatively small for market reaction estimates and for text mining. Though 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 previous literature, we believe the performance of our model could increase as the sample size increases. Second, we show that the sophisticated investors do not 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 current study. Last, we only consider a short time frame around the breach announcement date. However, some breach announcements have more detailed and some 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 react negatively to security breaches can be further investigated. Second, given that managers and other insiders are more likely to know 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 now becomes popular information sources for investors. We can further consider other media sources, such as blogs, 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 react negatively around the breach announcement day.

CHAPTER 4. COST AND BENEFIT ANALYSIS OF TWO-FACTOR AUTHENTICATION SYSTEMS

4.1. Introduction

Authentication can be used to verify either the content of the message, the origin of the message, or the identity of the user (Liebl 1993). Identity authentication focuses on the process of verifying a person's identity. In general, the information (or factor) people use to identify themselves is (1) something the user is. This is biometric information, such as fingerprints; (2) something the user has, such as an ID card; (3) something the user knows, such as a password (O'Gorman 2003). In some situations, users have to provide two of the above information simultaneously, for instance, an Automatic Teller Machine (ATM) card and a Personal Identification Number (PIN). This is called two-factor authentication. Two-factor or multi-factor authentication, as the name suggests, uses more than one single piece of information (i.e., factor) when granting access right. By using more information, the authentication system could have a smaller probability of system failure (defined in Section 4.3.2) for any online service or product provider.

As the concerns about identity theft have increased its popularity (Baum 2006), people start to argue whether the current authentication system can effectively distinguish imposters from genuine users and consider using two-factor authentication systems. For example, Federal Financial Institutions Examination Council (FFIEC) released guidance on authentication in Internet banking environment on October 12, 2005 (FFIEC 2005).

This guidance asked all the regulated agencies, by the end of 2006, to conduct risk-based assessments and to develop security measures to reliably authenticate (i.e., two-factor or multi-factor authentication) customers remotely accessing their online financial services.

The new system using more factors seems to be more secure, however, for customers, multi-factor authentication could also be accompanied by concerns about the use of additional information collected by the firm. Also, for customers, the new interfaces, new devices, and longer authentication processes could result in inconvenience of the authentication process and prolong the time needed to complete the transaction. These factors could affect the customers' willingness to keep subscribing services or purchasing products from the firm. For the firm, the implementation might require additional implementation costs, such as software, hardware, and training (Wildstrom 2005). All the above issues could at the same time affect a service or product provider's decision of implementing the new authentication system. However, it is not clear in the previous literature about how these inter-related factors could affect a firm's authentication system decisions.

This paper attempts to use a static method as a first attempt to understand the decision of choosing authentication systems. In particular, this paper attempts to address the following questions. From an online service or product provider's perspective, what are the key elements it needs to consider when shifting to another single-factor or two-factor authentication system? What are the conditions that make the new authentication system more preferable? However, given there are all kinds of single-factor authentication technologies and different combinations of two- or multi-factor authentication systems, it is unrealistic to optimize the decision by considering all the

possibilities. Moreover, when comparing different authentication systems, it is not clear whether one system is always more preferable than the other. For example, two-factor authentication systems are not necessarily much more expensive or guaranteed to be more secure than one-factor systems. This lack of clarity also makes the analysis more difficult. Therefore, in order to answer the above questions, this study first generalizes all the authentication systems into two broad types based on the definition of system failure probability. According to this generalization, we are able to compare the conditions that make the new authentication system more preferable regardless of the detail specification of the technology. These conditions allow us to uncover rules existing among the factors which provide rationale for managers' decisions.

The remainder of the paper is organized as follows. Relevant literature on authentication and privacy are discussed in Section 4.2. In Section 4.3, we propose a static model for one-factor and two-factor authentication systems. This model leads to our propositions and managerial implications in Section 4.4. We conclude with contribution, and possible avenues for future research in Section 4.5.

4.2. Literature Review

There are two major streams of literature related to our research. These two streams are authentication and privacy.

4.2.1. Authentication

The literature on authentication has long been discussed from the technical perspective. For instance, Woo and Lam (1992) and Diffle et al. (1992) provide the basic authentication mechanisms and the goals of authentication. Other studies focus on the design of protocols (e.g., Tardo and Alagappan 1991; Aboba et al. 2004) or ways to implement or improve authentication methods (e.g., Sutcu et al. 2005; Bhargav-Spantzel et al. 2006a, 2006b). However, studies about authentication from an economic perspective are limited. These studies are often embedded in the discussion of other issues. For example, Anderson (2001) discusses the role of authentication in information security from an economic perspective while authentication has also been discussed in internal control and EDP auditing literature (e.g., Webber 2001). Different from previous literature, our study focuses on the authentication system decisions from an economic perspective and provides decision rules for managers. Specifically, our study formally models the probability of system failure and generalizes the authentication systems into two broad categories based on the calculation of the probability of system failure. The first type of system failure probability is for the systems using the information someone has and someone knows as discussed in Introduction. The second type of system failure probability builds on the biometric literature and calculates the probability of system failure for biometric authentication systems. This will be discussed in detail in the next section.

To implement an authentication system, it is necessary to obtain users' personal identifiable information, such as names, addresses, and even purchasing history of an identifiable individual (e.g., Nowak and Phelps 1992). In the biometric case, personal

data can be the image captured at the enrollment stage or the result of the matching process (Rejman-Greene 2005). Several studies have discussed the information collected and the techniques to preserve privacy in the context of authentication systems (e.g., Davida et al. 1998; Camenisch and Lysyanskaya 2001; Perrig et al. 2004; Dhamija and Tygar 2005; Bhargav-Spantzel et al. 2006b). These concerns will make some customers choose to purchase the service or product from another provider with higher protection level. Also, some customers might also decide to switch to other providers if the system fails. The above two impacts in opposite directions could in turn affect a firm's decision on implementing a new authentication system.

4.2.2. Privacy from an Economic Perspective

This study, thus, also relates to, though not directly, the literature on privacy from an economic perspective. Privacy is defined as the individual's ability to control the collection and use of personal information (e.g., Westin 1967; Stigler 1980; Goodwin 1991; Foxman and Kilcoyne 1993; Hui and Png 2005). Studies about privacy from an economic perspective include reviews on the economic analyses of privacy (e.g., Hui and Png 2005), how businesses use personal information to customize services and to discriminate consumers (e.g., Varian 1985; Chen and Iyer 2002; Ghose and Chen 2003), and how business use personal information for promotions and cross market information (e.g., Akçura and Srinivasan 2005; Hann et al. 2005). The violation of privacy depends on (1) whether consumers can control the amount and the depth of information collected, and (2) the knowledge of the collection and use of their personal information (e.g., Caudill and Murphy 2000). In the context of authentication systems, the change in

authentication level could imply the need for more information depending on the system a firm chooses and the amount of information that might lose once the system fails. Also, privacy concerns rise with the use of the information collected. For instance, Hoffman et al. (1999) show that about 95% of online users are reluctant to provide personal information to websites because of privacy concerns. Therefore, the privacy concerns are involved in the selection process of authentication system alternatives.

4.3. Model

In this section, we first provide the basic settings for our analysis. Then the definition of system failure and the probability of system failure under different authentication methods are discussed followed by the details of our models for one-factor and two-factor authentication systems. Finally, by comparing the expected losses and costs for the firm when switching to another authentication system, we show the conditions that make the new authentication system preferable.

4.3.1. Basic Settings

We focus on one online service or product provider in this study. This provider currently has a market share of m in the service or product category it provides, where $0 < m < 1$ (see Appendix E for variable definitions). This market share m can also be interpreted as the total value the provider can get from the customers comparing to other providers. In order to complete the transaction process, each of the providers' customers is required to provide a certain level (α, $0 < \alpha \leq 1$) of personal information, such as name,

address, and phone number. If the system fails (define later), the product or service provider might need to compensate its consumers' losses and to pay a legal penalty or fine (L for both the compensation and penalties) for not abiding by the privacy commitment or regulations (e.g., Tang et al. 2008).

The customers are categorized along two dimensions: privacy and convenience. The first dimension is privacy sensitivity. A proportion of customers (ρ, $0 \leq \rho \leq 1$) are privacy sensitive in the market the provider faces. This portion of customers has more concerns about the information collected from them and the use of the information. Therefore, on the one hand, after the provider shifts to another authentication system, this provider might attract some potential customers with high privacy concerns because the new system might protect the information better (e.g., Wildstrom 2005). On the other hand, when the system has been breached, some of the existing customers might choose to subscribe the service or purchase the product from other providers because of the privacy concerns.

The other dimension is convenience sensitivity. A proportion of customers (δ, $0 \leq \delta \leq 1$) emphasize more on the convenience of the transaction such as the time required to complete the transaction. After the provider switches to a new authentication system, a certain portion of these customers might not keep subscribing or purchasing from this provider because the possible inconvenience, such as prolonged transaction time, caused by the new system. This categorization is illustrated through Figure 4.1.

In this paper, system failure is defined as any situation in which non-genuine users (e.g., hackers) are able to access to the information or genuine users are unable to access to the information because of the failure of the software or hardware, compatibility issue

of the software or hardware, for example, or the successful action of the hackers. Based on the definition, we discuss the probability of system failure for different authentication systems.

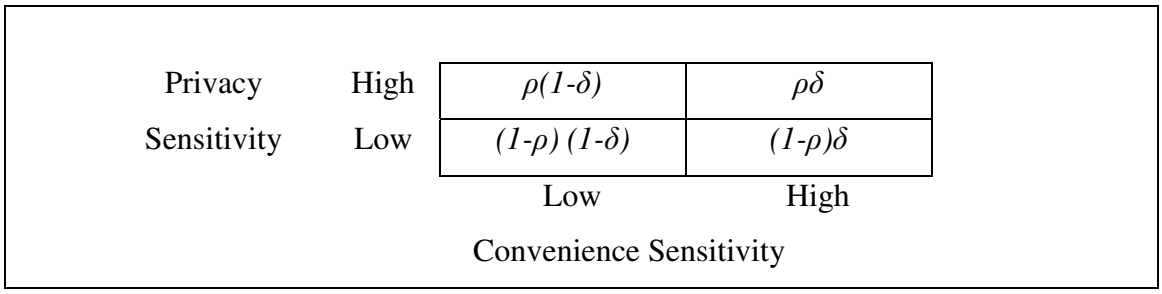

Figure 4.1 Types of Customers

4.3.2. Probability of System Failure

As discussed, we categorize all the authentication systems into two general categories. The first category uses information someone has or someone knows while the other category uses biometric information.

When the information used for authentication is the information someone knows or someone has, the authentication system can be regarded as a non-repairable system with one component. The reason is that, as an analogy to light bulbs, the longer the time we use a light bulb, the higher the chance that we need to replace it. In our context, this means that the longer the time we use a system, the larger the possibility that the system might encounter software or hardware problem due to compatibility issue, for example. Accordingly, based on the concept of reliability analysis (WeiBull.com 2003), the cumulative density function (CDF) of system failure of one non-repairable component across time t equals to $1-e^{-(t/\lambda)^b}$ where λ is the mean-time-to-failure and b is the change

of failure rate. Given there are all kinds of authentication systems within this category, in the following analysis, we do not pose any assumption on λ and b.

However, this probability (i.e., $1-e^{-(t/\lambda)^b}$) only accounts for one part of the probability of system failure. According to our definition of system failure, when an imposter uses the correct information and gains access right to the system should also be considered as system failure. For example, a hacker can obtain the correct login information through phishing. However, when the hacker enters this correct information, the authentication system allows the hacker to login but still functions correctly. This is because when the system uses the information someone knows and someone has, the access decision is dichotomy, i.e., correct or incorrect information. Since the system still functions correctly, the above probability (i.e., $1-e^{-(t/\lambda)^b}$) does not capture the situation when an imposter uses the correct information and gains access right to the system. Therefore, in order to take into account this possibility, we also need to consider the hackers' successful actions. Since hackers' technology is improving with time and the chance of getting the authentication information through other media, such as phishing, is also higher as time passes, the successful rate of the hackers' actions (denoted as $H(t)$) under different authentication methods should be an increasing function of time.

Consequently, from the discussion in the previous two paragraphs, the overall probability of system failure for one non-repairable component system (denote as $F_n(t)$ where the subscript n represents the one non-repairable component) is thus assessed by both $1-e^{-(t/\lambda)^b}$ and $H_n(t)$, i.e., $F_n(t)=\left(1-e^{-(t/\lambda)^b}\right)+H_n(t)-\left(1-e^{-(t/\lambda)^b}\right)H_n(t)$. Note that

since the hackers' successful action could co-occur with software or hardware problems, we need to consider the probability when both occur.

Similarly, if there are two independent non-repairable components, the probability of system failure across time *t* (denote as $F_{nn}(t)$ where the subscript *nn* represents two non-repairable components) is assessed by both $1-e^{-(t/\lambda_1)^{b_1}--(t/\lambda_2)^{b_2}}$ and $H_{nn}(t)$. There are two points worth noting. First, component 1 and component 2 can have different mean-time-to-failure (λ_1 and λ_2) and have different change of failure rate (b_1 and b_2). Although the conditions that make the new authentication system more preferable can be different, our unreported results show that the main propositions in the next section are the same. Therefore, we choose not to have a detailed discussion of these two parameters in the analysis section. Second, these two components could also be dependent. In the case when these two components are dependent, we need to reconsider the failure probability of one component given the other component has failed. Similarly, although the conditions can be different, from the unreported results, our main propositions in the following section remain similar with two dependent components. Therefore, in the following analysis, we only show the case when the two components are independent.

The other information that can be used for authentication systems is biometric information. Biometric authentication systems are different from the systems using the information someone has and someone knows because of the following. Biometric authentication system measures an individual's physical features based on the data stored, and then determines the identity of the user. Biometric systems use "scores" to show the similarity between a pattern and a biometric template (e.g., Braghin 2001; BioID.com

2004; Jain et al. 2004; Bromba Biometrics 2006; Ross et al. 2006). For example, the pattern of someone's fingerprints is matched with the template fingerprints. The higher the score is, the higher the similarity. If the score is higher than a certain pre-determined threshold, access right is granted. Depending on the threshold chosen, the impostor patterns can be falsely accepted by the system. At the current state, The False Acceptance Rate (FAR) is from 0.0001% to 0.1% (e.g., Foxman and Kilcoyne 1993; Panko 2003; Jain et al. 2004). Similarly, if the threshold is too high, some genuine patterns may be falsely rejected. The False Rejection Rate (FRR) is currently within the range from 0.00066% to 1.0% (e.g., Foxman and Kilcoyne 1993; Yun 2002; Panko 2003; Jain et al. 2004). Under the current state of technology solutions, different biometric traits have different accuracy rates and require different implementation costs. For example, fingerprint systems can be relatively cheap to implement with high accuracy at the same time while iris pattern systems could have high accuracy rate and high implementation cost at the same time (e.g., Panko 2003; Jain et al. 2004; Bromba Biometrics 2006;). From the above discussion, it is obvious that the characteristics of biometric authentication systems are different from those of the authentication systems using the information someone is and someone knows. Accordingly, for biometric systems, we should consider a probability of false acceptance (FAR, ψ) and false rejection (FRR, φ) at any given time t based on the pre-determined threshold ($\bar{s}$) and the change of these physical characteristics. The provider can use the receiver operating characteristic (ROC) curve to determine the weights between FAR and FRR that match its needs. However, the decision of the detailed specification of the technology (i.e., the threshold, FAR, FRR) is out of the scope of this study and we do not pose any

assumption on the specification in the following analysis. All these parameters can vary depending on the provider's choice. Once the specification is determined, the probability of system failure given the pre-determined threshold ($\bar{s}$) across time t (denote as $F_{bio}(t;\bar{s})$ where the subscript *bio* represents the biometric system) is assessed by both $1-(1-w_{FRR}\varphi-w_{FAR}\psi)^t$ (e.g., Poh et al. 2002) and $H_{bio}(t)$, where w_{FRR} and w_{FAR} are the weights pre-determined by the provider at the time when it determines the specification of the system.

Similarly, if the provider selects an authentication system that uses both non-biometric and biometric information, the probability of system failure given the pre-determined threshold ($\bar{s}$) across time t (denote as $F_{nbio}(t;\bar{s})$ where *nbio* represents the system with one non-repairable component and one biometric component) is calculated by both $1-e^{-(t/\lambda)^b}(1-w_{FRR}\varphi-w_{FAR}\psi)^t$ and $H_{nbio}(t)$.

4.3.3. Analysis

We start our analysis with the base case: one non-repairable component authentication system. Specifically, the provider is now using the one non-repairable component authentication system and considers switching to other authentication systems. Our analysis aims at showing that the key elements the provider should consider. To do so, we focus on the expected costs and losses the provider faces when implementing an authentication system.

The expected costs and losses (denoted as C) associated with the one non-repairable component authentication system can be expresses as the addition of the implementation

costs (c), the change in customer base when the system fails, and the expected losses when the system fails. The change in customer base is the loss of customers due to the failure in terms of the value these customers can create (V) which equals the market share (m) times a percentage ($0 \leq \varepsilon_1 \leq 1$) of ρ (see Appendix E for the definition of ε_i). The expected loss is the value the provider needs to compensate its customers and to settle possible lawsuits and penalty (L) once the system fails. Formally,

$$C_n = c_n + F_n(t)(V_n + L_n) \qquad \text{Eq. 4.1}$$

where V_n equals $m \times \varepsilon_1 \times \rho$ as discussed and the subscript n represents the one non-repairable component authentication system.

If the firm decides to use a new biometric authentication system to replace this current one non-repairable component authentication system, the associated expected costs and losses consist of four components. The first component again is the implementation costs. The second component reflects the net change of the customer base when the provider shifts to the new system which is measured by the net value these customers can bring. Specifically, the provider might attract a certain number of potential privacy sensitive customers because of this new and possible safer authentication system while losing a certain number of existing convenience sensitive customers because the inconvenience associated with the new method. This net effect consists of the loss of existing customers which equals the current market share (m) times a certain percentage ($0 \leq \varepsilon_2 \leq 1$) of δ while the benefit of attracting new customers equals the potential market share ($1 - m$) times a certain percentage ($0 \leq \varepsilon_3 \leq 1$) of ρ. The last two terms again are the loss of customers after the system fails (V_{bio}, which equals the new market share after considering the net change of the customer base times a certain

percentage ($0 \le \varepsilon_4 \le 1$) of ρ) and the expected losses (L_{bio}) if the system fails which is similar to the base case. Accordingly,

$$C_{bio} = c_{bio} + V_{net\_bio} + F_{bio}(t;\bar{s})(V_{bio} + L_{bio}) \qquad \text{Eq. 4.2}$$

where V_{net_bio} equals $m \times \varepsilon_2 \times \delta - (1-m) \times \varepsilon_3 \times \rho$, V_{bio} equals $[m - m \times \varepsilon_2 \times \delta - (1-m) \times \varepsilon_3 \times \rho] \times \varepsilon_4 \times \rho$, and the subscript *bio* represents the biometric system while the subscript *net_bio* represents the net change of the customer base when the provider shifts to the new system in terms of the value these customers can create without considering the probability of system failure as shown.

In the same vein, if the firm decides to use a two non-repairable component authentication system or the combination of one non-repairable component and one biometric component authentication system, the associated expected costs and losses again consist of four major components which are given in Eq. 4.3 and Eq. 4.4 respectively.

$$C_{nn} = c_{nn} + V_{net\_nn} + F_{nn}(t)(V_{nn} + L_{nn}) \qquad \text{Eq. 4.3}$$

$$C_{nbio} = c_{nbio} + V_{net\_nbio} + F_{nbio}(t;\bar{s})(V_{nbio} + L_{nbio}) \qquad \text{Eq. 4.4}$$

where the subscript *nn* (*nbio*) represents the two non-repairable component authentication system (the combination of one non-repairable component and one biometric component authentication system) and the subscript *net_nn (net_nbio*) represents the net change of the customer base when the provider shifts to the new system in terms of the value these customers can create. Similarly, V_{net_nn} equals $m \times \varepsilon_5 \times \delta - (1-m) \times \varepsilon_6 \times \rho$ and V_{net_nbio}

equals $m \times \varepsilon_8 \times \delta - (1-m) \times \varepsilon_9 \times \rho$. V_{nn} equals $[m - m \times \varepsilon_5 \times \delta - (1-m) \times \varepsilon_6 \times \rho] \times \varepsilon_7 \times \rho$ while V_{nbio} equals $[m - m \times \varepsilon_8 \times \delta - (1-m) \times \varepsilon_9 \times \rho] \times \varepsilon_{10} \times \rho$.[3]

In order to address our research question, we subtract Eq. 4.1 from Eq. 4.2, Eq. 4.3, and Eq. 4.4 in order to understand the factors and the conditions that make the shifting worthwhile. The results are shown in Panel A, Panel B, and Panel C in Appendix F. Since one-factor and two-factor authentication systems are inherently different in terms of the calculation of the probability of system failure, we choose to focus on comparing one-factor with another one-factor system and to compare two-factor with another two-factor authentication system. That is, whether the shift to another one-factor authentication system is worth engaging and which two-factor authentication system the provider should choose. On the one hand, the results given in Appendix F Panel A compare two different types of one-factor authentication systems: a biometric system and a one non-repairable component system. The results demonstrate the conditions that a biometric system is more preferable from five different parameters: additional implementation costs, percentage of privacy sensitive customers, percentage of convenience sensitive customers, market share, and the expected losses when the system fails. On the other hand, we also compare two different types of two-factor authentication systems. In particular, we subtract Eq. 4.4 from Eq. 4.3 to determine the conditions that make a two non-repairable component system more preferable than the system with one non-repairable component and one biometric component system as shown in Appendix F Panel D. Similarly, the conditions are from five different

[3] We do not consider two biometric component systems in our analysis because given that, at the current stage, two biometric solutions for online authentication is still under developing and is not a readily available alternative.

parameters: additional implementation costs, percentage of privacy sensitive customers, percentage of convenience sensitive customers, market share, and the expected losses when the system fails. All these conditions are discussed in the next section with managerial implications.

4.4. Managerial Implications

From the conditions given in Appendix F, the conditions that could make the new authentication system more preferable are essentially similar and can be boiled down to the factors stated in Proposition 1. Again, the focus of the analysis will be on Appendix F Panel A and Panel D.

Proposition 1: *When deciding to shift to a new authentication system from the current one non-repairable component authentication, the service or product provider should consider (1) the additional implementation costs ($c_{bio} - c_n$, $c_{nn} - c_{nbio}$), (2) the net change of the value of its customers (V_{net_bio}, V_{net_nn}, V_{net_nbio}) including the loss of customers after the system fails ($F_n(t)V_n$, $F_{bio}(t;\bar{s})V_{bio}$, $F_{nn}(t)V_{nn}$, $F_{nbio}(t;\bar{s})V_{nbio}$) which is determined by the percentage of privacy sensitive customers (ρ), the percentage of convenience sensitive customers (δ), and the current market share or market value of customers (m), and (3) the expected losses once the system fails (F(t)L).*

From Appendix F and Proposition 1, there are several points worth noting. First, the condition for the additional implementation costs shows that the additional implementation costs of the new system have to be smaller than $F_n(t)(V_n + L_n) - F_{bio}(t;\bar{s})(V_{bio} + L_{bio}) - V_{net\_bio}$ (or $F_{nbio}(t;\bar{s})(V_{nbio} + L_{nbio}) - F_{nn}(t)(V_{nn} + L_{nn}) + V_{net\_nbio} - V_{net\_nn}$ in the two factor case). The threshold reflects the following. Although the probability of system failure could be smaller for the new system (depending on the provider's choice and the CDF defined earlier), the change in the customer base also plays an important role. The possible decrease in the probability of system failure is not enough to justify the spending for the new systems. Specifically, the implementation costs of the new system need to be balanced with the reduced losses as well as the net change of customer value. Obviously, if the new system can attract more customers and reduce the losses at the same time, even the implementation costs is relatively higher, the new system is still more preferable. This can be shown as in Figure 4.2. Figure 4.2 illustrates that when all other factors are fixed (m = 0.5, δ = 0.8, ρ = 0.8, $\varepsilon_2 = \varepsilon_3 = \varepsilon_4 = \varepsilon_5 = \varepsilon_6 = \varepsilon_7 = 0.8$, $L_{nn} = L_{nbio} = 0.8$, t = 200, $\lambda_1 = \lambda_2 = 100$, $b_1 = b_2 = 3$, $c_{bio} = 1$), the associate between the difference of the implementation costs under two non-repairable component as well as one non-repairable and one biometric component system and the total expected costs and losses. Obviously, when all other factors are fixed, the higher the implementation cost, the less preferable a system is (the total expected costs and losses (C) is larger).

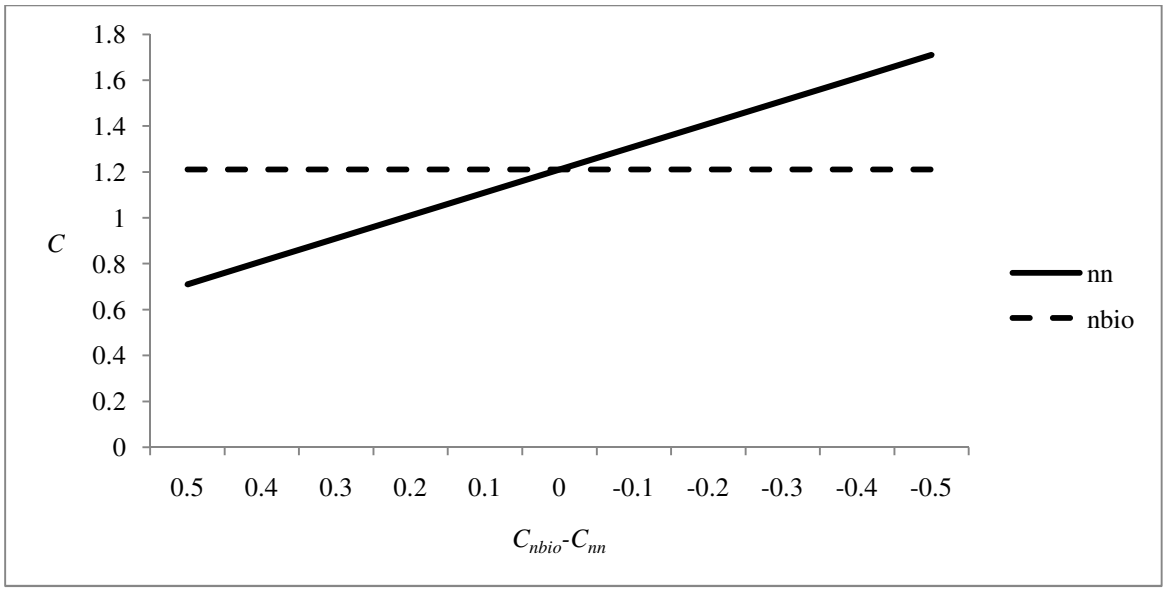

Figure 4.2 The Impact of Implementation Costs on Authentication System Decision

Second, in order to make the new system more preferable compared to the base case (i.e., compare two one-factor authentication systems), the percentage of privacy sensitive customers in the market the provider faces should be within $\frac{-Y-\sqrt{Y^2-4XZ}}{2X}$ and $\frac{-Y+\sqrt{Y^2-4XZ}}{2X}$ where $X = F_{bio}(t;\bar{s})(1-m)\varepsilon_3\varepsilon_4$, $Y = F_{bio}(t;\bar{s})(m\varepsilon_4 - m\delta\varepsilon_2\varepsilon_4) - (1-m)\varepsilon_3 - F_n(t)m\varepsilon_1$, $Z = c_{bio} - c_n + m\delta\varepsilon_2 + F_{bio}(t;\bar{s})L_{bio} - F_n(t)L_n$. If the percentage of privacy sensitive customers is too low, the additional implementation costs and expected losses cannot be justified by the improving of the security level. For example, we observe that many online service or product providers only choose to have the authentication system in the base case because the transaction amount is generally small and the transaction frequency is generally low. The customers only need to provide the

name and address to complete the transaction. In this case, a complicated authentication system is not necessary.

However, the condition also suggests that the percentage of privacy sensitive customers should not be too high. This result seems to be counter intuitive at first glance because if most of the customers care about whether the provided information is used properly, it seems that an authentication system with higher security level should better fit with the customers' preference. However, when we investigate the conditions in detail, it seems that if most of the customers are privacy sensitive, the provider might be able to attract new customers by shifting to the new authentication system but could lose more customers once the system fails. The loss of more customers could result from the loss of reputation and customers' expectations.

Different from the case of two one-factor authentication systems, the condition for the percentage of privacy sensitive customers when comparing two two-factor authentication systems is either smaller than $\frac{-Y-\sqrt{Y^2-4XZ}}{2X}$ or larger than $\frac{-Y+\sqrt{Y^2-4XZ}}{2X}$ where $X = F_{nbio}(t;\bar{s})(1-m)\varepsilon_9\varepsilon_{10} - F_{nn}(t)(1-m)\varepsilon_6\varepsilon_7$, $Y = F_{nbio}(t;\bar{s})(m\varepsilon_{10} - m\delta\varepsilon_8\varepsilon_{10}) - (1-m)(\varepsilon_6 - \varepsilon_9) - F_{nn}(t)(m\varepsilon_7 - m\delta\varepsilon_5\varepsilon_7)$, $Z = c_{nbio} - c_{nn} + m\delta(\varepsilon_5 - \varepsilon_8) + F_{nbio}(t;\bar{s})L_{nbio} - F_{nn}(t)L_{nn}$. On the one hand, when the majority of the customer base is not privacy sensitive, obviously, there is no need for a complicated system. On the other hand, if most of the customers are privacy sensitive, similarly, the one non-repairable and one biometric component system might attract more customers than the two non-repairable component system but could lose more once the system fails.

This can be illustrated as in Figure 4.3. Figure 4.3 shows that when all other factors are fixed (m = 0.5, δ = 0.8, $\varepsilon_2 = \varepsilon_4 = \varepsilon_5$ = 0.5, $\varepsilon_3 = \varepsilon_6$ = 0.5, ε_7 = 0.5, $L_{nn} = L_{nbio}$ = 0.8, $c_{nn} = c_{nbio}$ = 0.8, t = 200, $\lambda_1 = \lambda_2$ = 100, $b_1 = b_2$ = 3), the two non-repairable component system is more preferable when ρ is bigger than 80% or smaller than 1%. Therefore, we state our second proposition.

Proposition 2: Other things being equal, a securer (in terms of the probability of system failure) authentication system could attract more new customers but could also cause the loss of more customers once the system fails.

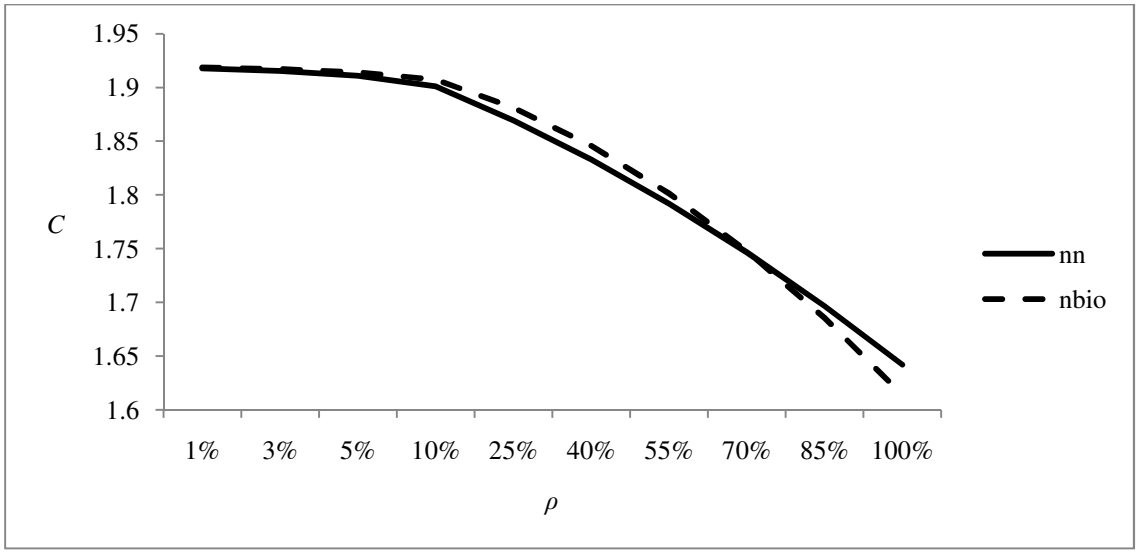

Figure 4.3 The Impact of the Percentage of Privacy Sensitive Customer on Authentication System Decision

Third, as shown in Appendix F, the conditions for the percentage of convenience sensitive customers suggest are $\delta < \dfrac{c_n - c_{bio} + (1-m)\rho\varepsilon_3(1-F_{bio}(t;\bar{s})\rho\varepsilon_4)}{m\varepsilon_2(1-F_{bio}(t;\bar{s})\rho\varepsilon_4)}$

$-\frac{F_{bio}(t;\bar{s})(m\rho\varepsilon_4 + L_{bio}) - F_n(t)(V_n + L_n)}{m\varepsilon_2(1 - F_{bio}(t;\bar{s})\rho\varepsilon_4)}$ when comparing two one-factor authentication systems and $\delta < \frac{c_{nn} - c_{nbio} + (1-m)\rho(\varepsilon_9 - \varepsilon_6) - F_{nbio}(t;\bar{s})[(m + (1-m)\rho\varepsilon_9)\rho\varepsilon_{10} + L_{nbio}]}{m(\varepsilon_8 - \varepsilon_5) + F_{nn}(t)m\rho\varepsilon_5\varepsilon_7 - F_{nbio}(t;\bar{s})m\rho\varepsilon_8\varepsilon_{10}} +$

$\frac{F_{nn}(t)[(m + (1-m)\rho\varepsilon_6)\rho\varepsilon_7 + L_{nn}]}{m(\varepsilon_8 - \varepsilon_5) + F_{nn}(t)m\rho\varepsilon_5\varepsilon_7 - F_{nbio}(t;\bar{s})m\rho\varepsilon_8\varepsilon_{10}}$ when comparing two two-factor authentication systems. These conditions exist only when the expected costs and losses of the original system are larger than those for the new system before considering the impact of inconvenience. In other words, before we consider the impact of inconvenience, all the other expected costs and losses must be relatively smaller. That is, if convenience is the main concern when deciding switching to the new authentication system, the provider should first evaluate whether the new system could fulfill the needs of its potential customers, instead of the existing customers. Otherwise, the new system is not preferable. For example, Figure 4.4 illustrates that when all other factors are fixed (m = 0.5, ρ = 0.8, $\varepsilon_2 = \varepsilon_7 = 0.5$, $\varepsilon_3 = \varepsilon_4 = \varepsilon_5 = \varepsilon_6 = 0.8$, $L_{nn} = L_{nbio} = 0.8$, $c_{nn} = c_{nbio} = 0.8$, t = 200, $\lambda_1 = \lambda_2 = 100$, $b_1 = b_2 = 3$), the two non-repairable component system is more preferable when δ is smaller than 0.3. Accordingly,

Proposition 3: If the service or product provider operates in the market where convenience is the major issue, the provider should focus on whether the new system could satisfy the needs of potential customers before evaluating the impact of privacy when deciding shifting to the new authentication system.

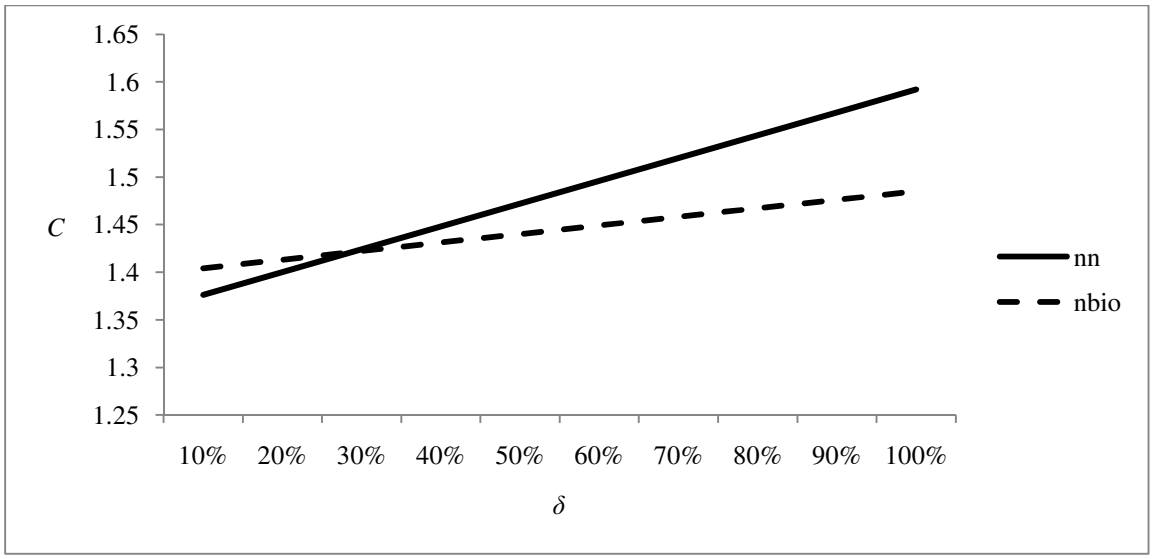

Figure 4.4 The Impact of the Percentage of Convenience Sensitive Customer on Authentication System Decision

Proposition 3 suggests that if the provider sells services or products where convenience is the major concern (no matter privacy concern is high or low), it should focus on the system that can lower the convenience concerns before worrying about the impact of privacy. If the inconvenience concerns cannot be lowered, the new system is not preferable and there is no need to consider the privacy issue.

Fourth, the current market share of the provider must be larger than $\frac{c_{bio}-c_n-\rho\varepsilon_3+F_{bio}(t;\bar{s})\rho^2\varepsilon_3\varepsilon_4+F_{bio}(t;\bar{s})L_{bio}-F_n(t)L_n}{\rho\varepsilon_3+\rho\varepsilon_1F_n(t)-\delta\varepsilon_2(1-F_{bio}(t;\bar{s})\rho\varepsilon_4-\rho\varepsilon_4F_{bio}(t;\bar{s})(1-\rho\varepsilon_3))}$ (or larger than $\frac{c_{nn}-c_{nbio}-\rho(\varepsilon_6-\varepsilon_9)+F_{nn}(t)(\rho^2\varepsilon_6\varepsilon_7+L_{nn})-F_{nbio}(t;\bar{s})(\rho^2\varepsilon_9\varepsilon_{10}+L_{nbio})}{\rho(\varepsilon_9-\varepsilon_6)+\delta(\varepsilon_8-\varepsilon_5)+F_{nn}(t)\rho\varepsilon_7(\rho\varepsilon_6-\delta\varepsilon_5-1)+F_{nbio}(t;\bar{s})\rho\varepsilon_{10}(1-\delta\varepsilon_8-\rho\varepsilon_9)}$ for two two-factor authentication systems) for the new authentication system (the two non-repairable component system) to be more preferable. The threshold for the market share that makes the new system more preferable increases as the additional implementation

costs increase. The market share (or the value of the existing customers) should be large enough because this value determines the net value change from the customers after shifting to the new authentication system which makes the new system more preferable. If the provider chooses a new system with the characteristics that are more expensive, the provider needs to have a larger market value of customers to balance and to justify the spending. This can be illustrated as in Figure 4.5. As shown in Figure 4.5, when holding other factors fixed ($c_{nn} = c_{nbio} = 0.8$, $\delta = 0.8$, $\rho = 0.8$, $\varepsilon_2 = \varepsilon_3 = \varepsilon_4 = \varepsilon_5 = 0.8$, $\varepsilon_6 = \varepsilon_7 = 0.5$, $L_{nn} = L_{nbio} = 0.8$, $t = 200$, $\lambda_1 = \lambda_2 = 100$, $b_1 = b_2 = 3$), as the market share getting larger, the system with two non-repairable component becomes more preferable (i.e., the cost is smaller).

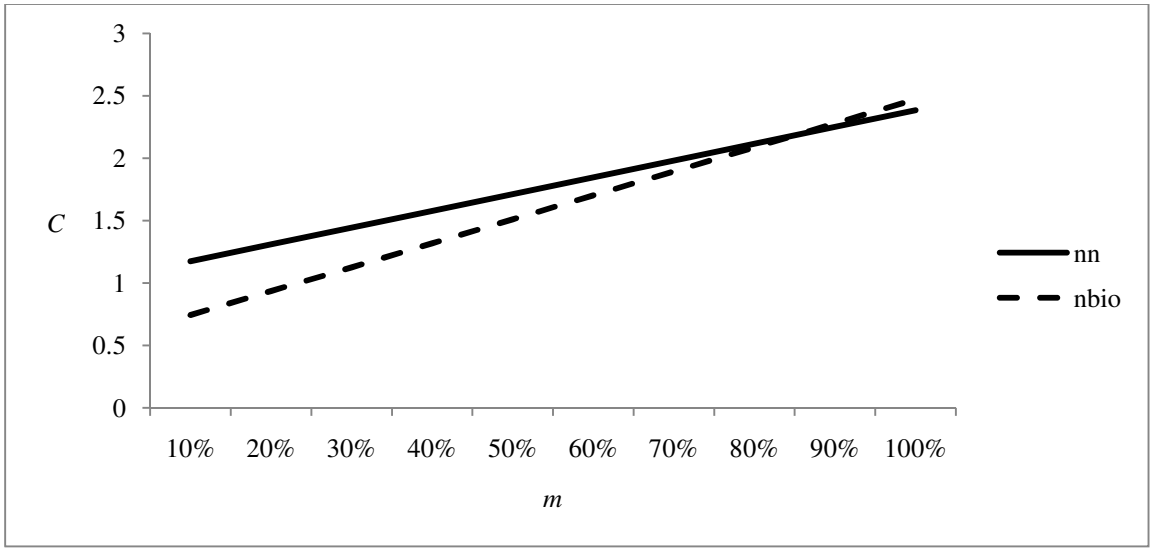

Figure 4.5 The Impact of Market Share on Authentication System Decision

However, in the real world cases, we do see the small market participants adopt the same new authentication system as the large market participants do. This seems to be contradicted with our result because the shift to a new authentication system is not

beneficial for small market participants. On the contrary, the conditions help explain this observation. These small market participants can in fact reduce the impact of the net change of customer value by adopting the same authentication system as the large market participants do especially when the majority of the customers are privacy sensitive. This is because the customers in this case do not have other alternatives of authentication systems among the providers. Therefore, the small market participants can justify the spending by the reduced outflow of customers toward other providers' new authentication system and the reduced probability of system failure especially when the shift of authentication system is mandatory. Specifically, when the shift to two- or multi-factor authentication system is mandatory, small market participants can adopt the same system (solution) as the large market participants do. For example, when financial institutions adopt new authentication systems in response to FFIEC, they tend to choose those adopted by large financial institutions. By doing so, they can not only ascertain their selection is acceptable by the regulator but also avoid possible losses from the switch in customers given similar institutions all adopt the same authentication system.

Proposition 4: Other things being equal, market participants with large market share can adopt the new authentication system by balancing the costs and expected losses with the net change of customer value while the small market participants can also adopt the same authentication system as the large market participants do in order to reduce the impact of the change of customer value caused by the shifting of authentication system of the larger market participants especially when the shift is mandatory.

Last, the expected losses resulting from the failure of the new authentication system should not exceed $c_n - c_{bio} - m\delta\varepsilon_2 + (1-m)\rho\varepsilon_3 - F_{bio}(t;\bar{s})V_{bio} + F_n(t)(V_n + L_n)$ (or $c_{nbio} - c_n + V_{net\_nbio} - V_{net\_nn} + F_{nbio}(t;\bar{s})(V_{nbio} + L_{nbio}) - F_{nn}(t)V_{nn}$) for two two-factor authentication systems) in order to make the new authentication system more preferable. If we plot the relationship between the expected losses and the total expected costs and losses (*C*), the figure will be similar to Figure 4.2. Although this result seems to be obvious, it has implication for public policies. In order to make the new system more preferable, one way is to relatively (comparing to the original system) lower the penalty and the compensation to customers associated with the new system once the new system fails. The other way is to relatively increase the penalty and the compensation to customers if the provider determines to keep the original authentication system. In other words, the providers could be penalized by implementing a less secure authentication system (in terms of the probability of system failure). By doing so, the relatively lowered penalty for the new system creates an environment where the new authentication is more attractable than the original one. The regulators could then force the provider to shift to the new system. Therefore,

Proposition 5: Other things being equal, by reducing the penalty associated with the new authentication system, the regulator is able to encourage the providers to adopt a more secure authentication system (in terms of the probability of system failure).

From the above propositions, we notice that the composition of customers and the change in customer base are important factors when determining authentication systems.

This observation leads us to propose that an online service or product provider's does not necessarily have to choose either one-factor or two-factor authentication systems. Instead, it could have both at the same time depending on the customers' preferences and the nature of the service or product category. Specifically, for different group of customers, the provider can implement different authentication systems in order to fit the preferences of different group of customers.

4.5. Conclusion

This study compares the expected costs and losses of different authentication methods. The results show the key factors and several insights for online service or product providers when shifting to a new authentication system. In order to make the new authentication system more preferable, the managers need to take into account the additional implementation costs, the current market share and the composition of customers. We show that small market participants can follow the same strategy adopted by the large market participants in order to lower the impact of the switch in customer especially when the shift is mandatory. Also, the conditions demonstrate that government can encourage the use of a more secure authentication system by adjusting the penalty a firm faces when the system fails. Finally, it might be appropriate for a firm to implement both one-factor and two- or multi-factor authentication systems depending on the customers' preferences.

The contribution of this study can be two folds. First, this study adds to the literature on authentication systems. To the best knowledge of the authors, the paper is the first paper attempting to understand the decision of authentication systems from an economic

setting instead of proposing technical solutions. More importantly, this study demonstrates that all kinds of authentication systems can be modeled into two broad categories: non-repairable and biometric. Although the parameters associated with different technology solutions vary, this generalization allows us to analyze the decision without concerning about the complexity of various authentication systems which can also be used for future studies about authentication systems. Second, for managers, this study provides suggestions when considering shifting to a new authentication system. As discussed in Section 4.4, all the elements need to be taken into account when determining whether the new system is worth engaging. More importantly, the rules we extract are general enough for managers to consider for different decisions regarding various authentication systems. This general rules can also be used even for multi-factor authentication systems the firm might adopt in the future.

There are several future extensions. First, as mentioned in the text, we choose to address our research question in a more static setting. There is still room for modeling competitors in a game theory setting and better capturing the effect of customer switching. Second, with the improvement of the technology and the standardization of the devices, the biometric authentication can have a totally different status, regardless of the accuracy, the costs and even the convenience. In the near future, it is interesting to discuss specifically on biometric systems in more detail and to consider two or more biometric components combined with each other. Finally, we can address the authentication issue from the users' perspectives and investigate how users perceive different systems and what the impacts on their adoption behavior are.

CHAPTER 5. CONCLUSIONS

This dissertation investigates three issues in information security to further our understanding and assessment of information security risks. In particular, this dissertation first focuses on how market participant can better interpret information security risk factors disclosed in financial reports. Then we examine investors' reactions to security incidents and propose a profitable short-term investment opportunity. Last, we investigate the authentication system decisions.

The first essay provides a comprehensive analysis to quantitatively and qualitatively investigate the association between security risk factors disclosed and subsequent security breach announcements. Specifically, we propose a classification model to link disclosure characteristics with breach announcements. The model shows that disclosures without action or process related terms are more likely to be associated with subsequent breach announcements and to be perceived as warnings to future incidents. To evaluate the usefulness of our model, we perform a cross-sectional analysis. The results demonstrate that the market values security risk factors disclosed at the time when financial reports are released but perceive such factors differently after breach announcements. Accordingly, our proposed model helps investors and debtors better interpret a firm's disclosed security risk factors and better assess the firm's future uncertainty regarding information security.

The second essay investigates how the textual contents of news articles about breach announcements affect the price and volume reactions to security incidents and profitable short-term investment opportunities by considering sophisticated investors reactions. The results demonstrate that general information about security breaches leads to different assessment of the impact of security incidents on a firm's future performance. However, specific breach announcements or articles about confidentiality type incidents could result in a consistently negative perception of the impact of security incidents. Interestingly, sophisticated investors, such as analysts and investment institutions, do not react negatively to security breaches immediately around the breach announcement day. Given the different perception of the impact of security incidents on a firm's future performance, we show that one could, on average, make 300% annual return by taking advantage of the information asymmetry among investors.

The third essay focuses on the decision of choosing authentication systems. By generalizing the probability of system failure and comparing the expected costs and losses of different systems, this essay demonstrates the key factors managers need to consider when determining a new authentication system. Overall, there are three key factors managers need to consider: (1) implementation costs, (2) the net benefit of customer switch due to the shift of authentication system, and (3) expected losses. This essay also demonstrates that the service or product provider can lower the impact of customer switch by following the large provider's action. Last, regulators can encourage the adoption of a more secure authentication by changing the penalty and fine a firm faces once the system fails.

BIBLIOGRAPHY

BIBLIOGRAPHY

Aboba, B., Blunk, L., Vollbrecht, J., Carlson, J., and Levkowetz, H. 2004. "Extensible authentication protocol (EPA)," *The Internet Engineering Task Force-Request for Comments*.

Abrahamson, E., and Park, C. 1994. "Concealment of negative organizational outcomes: an agency theory perspective," *Academy of Management Journal* (37:5), pp. 1302-1334.

Acquisti, A., Friedman, A., and Telang, R. 2008. "Is there a cost to privacy breaches? An event study," Working Paper, Carnegie Mellon University.

Ajinkya, B. B., and Gift, M. J. 1984. "Corporate managers' earnings forecasts and symmetrical adjustments of market expectations," *Journal of Accounting Research* (22:2), pp. 425-444.

Akçura, M. T., and Srinivasan, K. 2005. "Research note: customer intimacy and cross-selling strategy," *Management Science* (51:6), pp. 1007–1012.

Albert, T. C., Goes, P. B, and Gupta, A. 2004. "GIST: a model for design and management of content and interactivity of customer-centric web sites," *MIS Quarterly* (28:2), pp. 161-182.

Anderson, R. 2001. "Why information security is hard—an economic perspective," *Computer Security Applications Conference*, New Orleans, Louisiana.

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

Ayers, B. C., Jiang, J., and Yeung, P. E. 2006. "Discretionary accruals and earnings management: an analysis of pseudo earnings targets," *The Accounting Review* (81:3), pp. 617-652.

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

Bamber, L. 1987. "Unexpected earnings, firm size, and trading volume around quarterly earnings announcements," *The Accounting Review* (62), pp. 510-532.

Bamber, L., Barron, O. E., and Stober, T. L. 1997. "Trading volume and different aspects of disagreement coincident with earnings announcements," *The Accounting Review* (72), pp. 575-597.

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

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

Barth, M., Kasznik, R., and McNichols M. 2001. "Analyst coverage and intangible assets," *Journal of Accounting Research* (39), pp. 1-34.

Baum, K. 2006. "Identity theft, 2004," Bureau of Justice Statistics Bulletin, Retrieved January 25, 2009, from http://www.ojp.usdoj.gov/bjs/pub/pdf/it04.pdf.

Beaver, W. 1968. "The information content of annual earnings announcements," *Journal of Accounting Research* (6), pp. 67-92.

Beneish, M. D. 2001. "Earnings management: A perspective," *Managerial Finance* (27:12), pp. 3-17.

Berry, M. J. A., and Linoff, G. 1999. *Mastering data mining: the art and science of customer relationship management*. John Wiley & Sons, Inc., New York.

Bettman, J. R., and Weitz, B. A. 1983. "Attributions in the board room: causal reasoning in corporate annual reports," *Administrative Science Quarterly* (28:2) pp. 165-183.

Bhargav-Spantzel, A., Squicciarini, A., and Bertino, E. 2006a. "Establishing and protecting digital identity in federation systems," *Journal of Computer Security* (13:3), pp. 269–300.

Bhargav-Spantzel, A., Squicciarini, A., and Bertino, E. 2006b. "Privacy preserving multi-factor authentication with biometrics," *Conference on Computer and Communications Security Proceedings of the Second ACM Workshop on Digital Identity Management*, pp. 63-72.

Bhattacharya, N. 2001. "Investors' trade size and trading responses around earnings announcements: an empirical investigation," *The Accounting Review* (76:2), pp. 221-244.

Bhushan, R. 1989. "Firm characteristics and analyst following," *Journal of Accounting and Economics* (11), pp. 255-274.

BioID.com. 2004. *About FAR, FRR, and EER*. Retrieved July 8, 2006, from http://www.bioid.com/sdk/docs/About_EER.htm.

Bowen, R. M., Davis, A. K., and Matsumoto D. A. 2002. Do conference calls affect analysts' forecasts? *The Accounting Review* (77:2), pp. 285-316.

Bowen, P., Hash, J., and Wilson, M. 2006. *Information security handbook: a guide for managers*, NIST Special Publication 800-100.

Braghin, C. 2001. *Biometric authentication*. Department of Computer Science, University of Helsinki. Retrieved July 8, 2006, from http://www.avanti.ltol.org.

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

Breslow, N. E. and Day, N. E. 1980. *Statistical methods in cancer research*. IARC Science Publication, Lyon.

Bromba Biometrics. 2006. *Biometric FAQ*. Retrieved July 9, 2006, from http://bromba.com/faq/biofaq.htm.

Brown, L. D. 1991. "Forecast selection when all forecasts are not equally recent," *International Journal of Forecasting* (7), pp. 349-356.

Brown, L. D. 1993. "Earnings forecasting research: its implications for capital markets research," *International Journal of Forecasting* (9), pp. 295-320.

Brown, L. D., and Han, J. C. Y. 2000. "Do stock prices fully reflect the implications of current earnings for future earnings for ar1 firms?" *Journal of Accounting Research* (38:1), pp. 149-164.

Brown, S., and Warner J. 1985. "Using daily stock returns: the case of event studies," *Journal of Financial Economics* (14), pp. 3-31.

Camenisch, J., and Lysyanskaya, A. 2001. "Efficient non-transferable anonymous multi-show credential system with optional anonymity revocation," in B. Pfitzmann, editor, *Advances in Cryptology* — EUROCRYPT 2001 (2045), pp. 93–118.

Cameron, A. C., and Trivedi, P. K. 2007. *Microeconometrics: methods and applications*, Cambridge University Press, New York.

Campbell, K., Gordon, L. A., Loeb, M. P., and Zhou, L. 2003. "The economic cost of publicly announced information security breaches: empirical evidences from the stock market," *Journal of Computer Security* (11), pp. 431-448.

Caudill, E. M., and Murphy, P. E. 2000. "Consumer online privacy: legal and ethical issues," *Journal of Public Policy and Marketing* (19:1), pp. 7-19.

Cavusoglu, H., Mishra, B., and Raghunathan, S. 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),* pp. 69-105.

Cecchini, M., Aytug, H., Koehler, G. J., and Pathak, P. 2007. "Detecting management fraud in public companies," Working Paper, University of South Carolina.

Chen, Y., and Iyer, G. 2002. "Consumer addressability and customized pricing," *Marketing Science* (21:2), pp. 197-208.

Christensen, B. J., and Prabhala, N. R. 1998. "The relation between implied and realized volatility," *Journal of Financial Economics* (50), pp. 125-150.

Core, J. E. 2001. "A review of the empirical disclosure literature: discussion," *Journal of Accounting and Economics* (31:1-3), pp. 441-456.

Cosslett, S. R. 1981. "Maximum likelihood estimator for choice based samples," *Econometrica* (49:5), pp. 1289-1316.

CSI/FBI. 2008. The CSI/FBI computer crime and security report in 2007, Retrieved May 29 2009, from http://i.cmpnet.com/v2.gocsi.com/pdf/CSIsurvey2008.pdf.

Davida, G. I., Frankel, Y., and Matt, B. J. 1998. "On enabling secure applications through off-line biometric identification," *Proceedings of the 1998 IEEE Symposium of Privacy and Security*, pp. 148–157.

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

Dhamija, R., and Tygar, J. D. 2005. "The battle against phishing: dynamic security skins," *Proceedings of the 2005 Symposium on Usable Privacy and Security (SOUPS '05)*, pp. 77–88.

Diffle, W., van Oorschot P. C., and Wiener, M. J. 1992. "Authentication and authenticated key exchanges," *Designs, Codes and Cryptography* (2:2), pp. 357-390.

Donkers, B., Frances, P., and Verhof, P. 2003. "Selective sampling for binary choice models," *Journal of Marketing Research* (40), pp. 492-497.

Dye, R. A. 1985. "Disclosure of non-proprietary information," *Journal of Accounting Research* (12:1), pp. 123-145.

Easley, D., and O'Hara, M. 1987. "Price, trade size, and information in securities markets," *Journal of Financial Economics* (19), pp. 69-90.

Elgers, P., and Murray, D. 1992. "The relative and complementary performance of analyst and security-price-based measures of expected earnings," *Journal of Accounting and Economics* (15), pp. 303-316.

Elliott, R., and Jacobson, P. 1994. "Costs and benefits of business information disclosure," *The Accounting Horizons* (8:4), pp. 80-96.

Ettredge, M. L., and Richardson, V. J. 2003. "Information transfer among Internet firms: the case of hacker attacks," *Journal of Information Systems* (17:2), pp. 71-82.

Fama, E., and French, K. 1992. "The cross-section of expected stock returns," *The Journal of Finance* (47:2), pp. 427–465.

Fan, W., Davidson, I., Zadrozny, B., and Yu, P. S. 2005. "An improved categorization of classifier's sensitivity on sample selection bias," *5th IEEE International Conference on Data Mining*, Houston, Texas, USA.

Fan, W., Wallace, L., Rich, S., and Zhang, Z. 2006. "Tapping the power of text mining," *Communication of the ACM* (49:9), pp. 77-82.

Feldman, R., and Sanger, J. 2006. *The text mining handbook: advanced approaches in analyzing unstructured data*, UK: Cambridge University Press.

FFIEC. 2005. *FFIEC releases guidance on authentication in internet banking environment*. Federal Financial Institutions Examination Council. Retrieved July 8, 2006, from http://www.ffiec.gov/press/pr101205.htm.

Field, L., Lowry, M., and Shu, S. 2005. "Does disclosure deter or trigger litigation?" *Journal of Accounting and Economics* (39), pp. 487-507.

Foxman, E. R., and Kilcoyne, P. 1993. "Information technology, marketing practice, and consumer privacy: ethical issues," *Journal of Public Policy and Marketing* (12:1), pp. 106-119.

Francis, R., Philbrick, D., and Schipper, K. 1994. "Shareholder litigation and corporate disclosure," *Journal of Accounting Research* (32:2), pp. 137-164.

Francis, J., Hanna, J. D., Philbrick, D. R. 1997. "Management communications with securities analysts," *Journal of Accounting and Economics* (24), pp. 363-394.

Francis, J., Schipper, K., and Vincent, L. 2002. "Expanded disclosures and the increased usefulness of earnings announcements," *The Accounting Review* (77:3), pp. 515-546.

Frankel, R., Kothari, S. P., and Weber, J. 2006. "Determinants of the informativeness of analyst research," *Journal of Accounting and Economics* (41), pp. 29-54.

Frey, L., and Fisher, D. 1999. "Modeling decision tree performance with the power law," *Proceedings of the 7th International Workshop on Artificial Intelligence and Statistics*, San Francisco, CA, pp. 59-65.

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

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

Ghose, A., and Chen, P. Y. 2003. "Personalization vs. privacy: firm policies, business profits and social welfare," Working Paper, GSIA, Carnegie Mellon University.

Glover, S., Liddle, S., and Prawitt, D. 2001. *Electronic commerce: security, risk management, and control*, NL: Prentice Hall.

Goodhue, D. L., and Straub, D. W. 1991. "Security concerns of system users: a study of perceptions of the adequacy of security," *Information & Management* (20:1), pp. 13-27.

Goodwin, C. 1991. "Privacy: recognition of a consumer right," *Journal of Public Policy and Marketing* (10:1), pp. 149-166.

Gordon, L. A., and Loeb, M. P. 2002. "The economics of information security investment," *ACM Transaction on Information and System Security* (5:4), pp. 438-457.

Gordon, L. A., Loeb, M. P., and Lucyshyn, W. 2003. "Sharing information on computer systems security: an economic analysis," *Journal of Accounting and Public Policy* (22:6), pp. 461-485.

Gordon, L. A., Loeb, M. P., Lucyshyn, W., and Sohail, T. 2006. "The impact of the Sarbanes-Oxley Act on the corporate disclosures of information security activities," *Journal of Accounting and Public Policy* (25), pp. 503-530.

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

Grossman, S. J. 1981. "The information role of warranties and private disclosure about product quality," *Journal of Law and Economics* (24:3), pp. 461-483.

Han, J., Altman, R., Kumar, V., Mannila, H., and Pregibon, D. 2002. "Emerging scientific applications in data mining," *Communication of the ACM* (45:8), pp. 54-58.

Hann, I. H., Hui, K. L., Lee, T. S., and Png, I. P. L. 2005. "Consumer privacy and marketing avoidance," Unpublished manuscript, Department of Information Systems, National University of Singapore.

Harvey, C. R., and Whaley, R. E. 1992. "Dividends and S&P 100 index option valuation," *Journal of Futures Markets* (12), pp. 123-137.

Hasbrouck, J. 1988. "Trades, quotes, inventories and information," *Journal of Financial Economics* (22), pp. 229-252.

Hasbrouck, J. 1991. "Measuring the information content of stock trades," *The Journal of Finance* (46), pp. 179-207.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design science in information systems research," *MIS Quarterly* (28:1), pp. 75-105.

Hoffman, D. L., Novak, T. P., and Peralta, M. 1999. "Building consumer trust online," *Communications of the ACM* (42:4), pp.80-85.

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

Hui, K., and Png, I. P. L. 2005. *The economics of privacy*. Forthcoming in handbook of information systems and economics, Elsevier.

Imbens, G. 1992. "An efficient method of moments estimator for discrete choice models with choice-based sampling," *Econometrica* (60:5), pp. 1187-1214.

Ivkovic, Z. and Jegadeesh, N. 2004. "The timing and value of forecast and recommendation revisions," *Journal of Financial Economics* (73), pp. 433-463.

Jain, A. K., Ross, A. R., and Prabhakar, S. 2004. "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology* (14:1), pp. 4-20.

John, G., and Langley, P. 1996. "Static versus dynamic sampling for data mining," *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, Portland, pp. 367-370.

Jorgensen, B. N., and Kirschenheiter M. T. 2003. "Discretionary risk disclosures," *The Accounting Review* (78:2), pp. 449-469.

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

Karpoff, J. M. 1986. "A theory of trading volume," *The Journal of Finance* (41:5), pp. 1069-1087.

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

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

Kim, O., and Verrecchia, R. 1991. "Trading volume and price reactions to public announcements," *Journal of Accounting Research* (29), pp. 302-321.

Kim, O., and Verrecchia, R. 1994. "Market liquidity and volume around earnings announcements," *Journal of Accounting and Economics* (17), pp. 41-67.

Kim, O., and Verrecchia, R. 1997. "Pre-announcement and event-period private information," Working paper, University of Pennsylvania, Philadelphia, PA.

King, R., Pownall, G., and Waymire, G. 1990. "Expectations adjustment via timely management forecasts: review, synthesis, and suggestions for future research," *Journal of Accounting Literature* (9), pp. 113-144.

King, G., and Zheng, L. 2001. "Logistic regression in rare events data," *Political Analysis* (9:2), pp. 137-163.

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, pp. 781-787.

Kohl's. 2007. Annual report for the year ended February 3, 2007. Retrieved November 30, 2008 from http://www.kohlscorporation.com/InvestorRelations/pdfs/10k.pdf.

Kotulic, A. G., and Clark, J. G. 2004. "Why there aren't more information security research studies," *Information & Management* (41), pp. 597-607.

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

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

Lancaster, T., and Imbens, G. 1991. "Choice based sampling: inference and optimality," Working Paper, Department of Economics, Brown University.

Lang, M. H., Lins, K. V., and Miller, D. P. 2004. "Concentrated control, analyst following, and valuation: Do analysts matter most when investors are protected least?" *Journal of Accounting Research* (42:3), pp. 589-623.

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

Lang, M. H., and Lundholm, R. J. 1996. "Corporate disclosure policy and analyst behavior," *The Accounting Review* (71:4), pp. 467-492.

Landsman, W., and Maydew, E. 2002. "Has the information content of quarterly earnings announcements declined in the past three decades?" *Journal of Accounting Research* (40:3), pp. 797-807.

Leuz, C., and Verrecchia, R. E. 2000. "The economic consequences of increased disclosure," *Journal of Accounting Research* (38:3), pp. 91-124.

Lev, B., and Pennman, S. H. 1990. "Voluntary forecast disclosure, nondisclosure, and stock prices," *Journal of Accounting Research* (28:1), pp. 49-76.

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

Li, F. 2008. "Annual report readability, current earnings, and earnings persistence," *Journal of Accounting and Economics* (45:2-3), pp. 221-247.

Liebl, A. 1993. **"**Authentication in distributed systems: a bibliography," *ACM SIGOPS Operating Systems Review* (27:4), pp. 31-41.

Lohmeyer, D. F., McCroy, J., and Pogreb, S. 2002. "Managing information security," *The McKinsey Quarterly*, Retrieved May 29 2009, from http://news.cnet.com/ Managing-information-security/2009-1017_3-933185.html.

Long, W. J., Griffith, J. L., Selker, H. P., and D'agostino, R. B. 1993. "A comparison of logistic regression to decision-tree induction in a medical domain," *Computers and Biomedical Research* (26), pp. 74-97.

Manski, C. F., and McFadden, D. 1981. "Alternative estimators and sample designs for discrete choice analysis," C. F. Manski, D. McFadden, eds. *Structural analysis of discrete data with econometric applications*. MIT Press, Cambridge, MA.

Masand, G., Linoff, G., and Waltz, D. 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, pp. 59-65.

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

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

Milgrom, P. R. 1981. "Good news and bad news: representation theorems and applications," *Bell Journal of Economics* (12:2), pp. 380-391.

Morgan J., Daugherty, R., Hilchie, A., and Carey, B. 2003. "Sample size and modeling accuracy with decision tree based data mining tools," *Academy of Information and Management Science Journal* (6:2), pp. 77-92.

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

Nowak, G., and Phelps, J. 1992. "Understanding privacy concerns," *Journal of Direct Marketing* (6:4), pp. 28-39.

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

O'Gorman, L. 2003. "Comparing passwords, tokens, and biometrics for user authentication," *Proceedings of the IEEE* (91:12), pp. 2021-2040.

OptionMetrics. 2006. *Ivy DB file and data reference manual*, NY: OptionMetric LLC.

Panko, R. R. 2003. *Corporate computer and network security*. NJ: Prentice-Hall.

Penno, M. 1997. "Information quality and voluntary disclosure," *The Accounting Review* (72:2), pp. 275-284.

Perrig, A., Stankovic, J., and Wagner, D. 2004. "Security in wireless sensor networks," *Communications of the ACM* (47:6), pp. 53-57.

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

Pinczowski, D., Ekbom, A., Baron, J., Yuen, J., and Adami, H. 1994. "Risk factors for colorectal cancer in patients with ulcerative colitis: a case-control study," *Gastroenterology* (107:1), pp. 117-120.

Provost, F., Jensen, D., and Oates, T. 1999. "Efficient progressive sampling," *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining*, San Diego, CA, pp. 23-32.

Rejman-Greene, M. 2005. "Privacy issues in the application of biometrics: an European perspective," in Wayman, J. L., Jain, A. K., Maltoni, D., and Maio, D. editors, *Biometric Systems: Technology, Design and Performance Evaluation*, pp. 335-359, NY: Sprinter.

Ross, A. A., Nandakumar, K., and Jain, A. K. 2006. *Handbook of multibiometrics*. NY: Sprinter.

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

Rudolfer, S. M., Paliouras, G., and Peers, I. S. 1999. "A comparison of logistic regression to decision tree induction in the diagnosis of carpal tunnel syndrome," *Computers and Biomedical Research* (32), pp. 391-414.

Sandoval, G., and Wolverton, T. 2000. *Leading web sites under attack*. Retrieved April 17, 2007, from http://news.com.com/Leading+Web+sites+under+attack/2100-1017_3-236683.html.

SAS Institute Inc. 2004. *Getting started with SAS® 9.1 Text Miner*. Cary, NC: SAS Institute Inc.

SAS Institute Inc. 2008. *SAS/STAT® 9.2 user's guide*. Cary, NC: SAS Institute Inc.

Schechter, S. E., Smith, M. D. 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.

Shadish, W. R., Cook, T. D., and Campbell, D. T. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. NY: Houghton Mifflin Company.

Sheikh, A. 1989. "Stock splits, volatility increases and implied volatility," *The Journal of Finance* (44), pp. 1361-1372.

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

Siponen, M., and Iivari, J. 2006. "Six design theories for IS security policies and guidelines," *Journal of the AIS* (7:7), pp. 445-472.

Sivakumar, K. N., and Waymire, G. 1993. "The information content of earnings in a discretionary reporting environment: evidence from NYSE industrials, 1905-1910," *Journal of Accounting Research* (31), pp. 62-91.

Skinner, D. J. 1994. "Why firms voluntarily disclose bad news," *Journal of Accounting Research* (32:1), pp. 38-60.

Smyth, P. 2000. "Model selection for probabilistic clustering using crossvalidated likelihood," *Statistics and Computing* (10), pp. 63-72.

Sohail, T. 2006. *To tell or not to tell: market value of voluntary disclosures of information security activities*. Unpublished doctoral dissertation, University of Maryland, Maryland.

Sordo, M., and Zeng, Q. 2005. On sample size and classification accuracy: a performance comparison. Biological and Medical Data Analysis, Springer.

Sorenson, O., and Stuart, T. 2001. "Syndication networks and the spatial distribution of venture capital investment," *The American Journal of Sociology* (106:6), pp. 1546-1588.

Steinberg, G. D., Carter, B. S., Beaty, T. H., Childs, B. P., and Walsh, C. 2006. "Family history and the risk of prostate cancer," *The Prostate* (17:4), pp. 337-347.

Steyerberg, E., Roobol, M., Kattan, M., van der Kwast, T., de Koning, H., and Schröder F. 2007. "Prediction of indolent prostate cancer: validation and updating of a prognostic nomogram," *The Journal of Urology* (177:1), pp. 107-112.

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

Stigler, G. J. 1980. "An introduction to privacy in economics and politics," *Journal of Legal Studies* (9:4), pp. 623-644.

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

Stocken, P. 2000. "Credibility of voluntary disclosure," *RAND Journal of Economics* (31:2), pp. 359-374.

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

Sutcu, Y., Sencar, H. T., and Memon, N. 2005. "Authenticaiton/protocols: a secure biometric authentication scheme based on robust hashing," *Proceedings of the 7th Workshop on Multimedia and Security (MM&Sec '05)*, pp. 111-116.

Tan, A. H. 1999. "Text mining: the state of the art and the challenges," *Proceedings of the PAKDD'99 Workshop on Knowledge discovery from Advanced Databases*, Beijing.

Tanaka, H., Matsuura, K., and Sudoh, O. 2005. "Vulnerability and information security investment: an empirical analysis of e-local government in Japan," *Journal of Accounting and Public Policy* (24:1), pp. 37-59.

Tang, Z., Hu, J. Y., and Smith, M. D. 2008. "Gaining trust through online privacy protection: self-regulation, mandatory standards, or caveat emptor," *Journal of Management Information Systems* (24:4), pp. 153-173.

Tardo, J. J., and Alagappan, K. 1991. "SPX: global authentication using public key certificates," *Proceedings of IEEE Symposium on Research in Security and Privacy*, pp. 232-244.

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

Varian, H. R. 1985. "Price discrimination and social welfare," *American Economic Review* (75:4), pp. 870-875.

Verrecchia, R. E. 1983. "Discretionary disclosure," *Journal of Accounting and Economics* (5:3), pp. 179-194.

Verrecchia, R. E. 2001. "Essays on disclosures," *Journal of Accounting and Economics* (32:1-3), pp. 97-180.

Wang, T. W., Rees, J., and Kannan, K. 2008. "Reading disclosures with new eyes: bridging the gap between information security disclosures and incidents," Workshop on Economics and Information Security (WEIS 2008), New Hampshire.

Warren, M. J., and Hutchinson, W. E. 2000. "Cyber attacks against supply chain management systems," *International Journal of Physical Distribution and Logistics Management* (30), pp. 710-716.

Webber, R. 2001. *EDP auditing—conceptual foundations and practice*, NY: McGraw-Hill.

WeiBull.com. 2003. "Analysis reference: reliability, availability, and optimization," ReliaSoft's eTextbook.

Weiss, S. M., and Kapouleas, L. 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, Michigan, pp. 781-787.

Westin, A. 1967. *Privacy and freedom*. NY: Atheneum.

Wildstrom, S. H. 2005. "New weapons to stop identity thieves," *BusinessWeek* (May), p. 24.

Winter, 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), pp. 196-202.

Woo, T. Y. C., and Lam, S. S. 1992. "Authentication for distributed systems," *Computer* (25:1), pp. 39-52.

Young, S. R., and Hayes, P. J. 1985. "Automatic classification and summarization of banking telexes," *Proceedings of the 2nd IEEE Conference on AI Applications*, Miami Beach, FL, pp. 402-409.

Yun, Y. W. 2002. "The '123' of biometric technology," *Synthesis Journal*, pp. 83-96.

Zadrozny, B. 2004. "Learning and evaluating classifiers under sample selection bias," *Proceedings of the 21st International Conference on Machine Learning*, Banff, Canada, pp. 903-910.

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

Zhou, Z., and Jiang, Y. 2004. "NeC4.5: Neural Ensemble Based C4.5," *IEEE Transactions on Knowledge and Data Engineering*, (16:6), pp. 770-773.

APPENDICES

Appendix A. An Example of Information Security Risk Factors

"System Interruption and the Lack of Integration and Redundancy in Our Systems May Affect Our Sales

Customer access to our Web sites directly affects the volume of goods we sell and thus affects our net sales. We experience occasional system interruptions that make our Web sites unavailable or prevent us from efficiently fulfilling orders, which may reduce our net sales and the attractiveness of our products and services. To prevent system interruptions, we continually need to: add additional software and hardware; upgrade our systems and network infrastructure to accommodate both increased traffic on our Web sites and increased sales volume; and integrate our systems.

Our computer and communications systems and operations could be damaged or interrupted by fire, flood, power loss, telecommunications failure, break-ins, earthquake and similar events. We do not have backup systems or a formal disaster recovery plan, and we may have inadequate insurance coverage or insurance limits to compensate us for losses from a major interruption. Computer viruses, physical or electronic break-ins and similar disruptions could cause system interruptions, delays and loss of critical data and could prevent us from providing services and accepting and fulfilling customer orders. If this were to occur, it could damage our reputation."

Excerpt from Amazon's annual report for year 2000, retrieved on Apr.23, 2007
Source:
http://www.sec.gov/Archives/edgar/data/1018724/000103221001500087/0001032210-01-500087.txt

Appendix B. Stock Price and Trading Volume Reactions to Security Incidents

In our study, the market model is used to capture the impact of security incidents.

$$R_{it} = \beta_0 + \beta_1 R_{mt} + \varepsilon_{it} \quad \text{Eq. B.1}$$

where R_{it} denotes company i's return at period t which equals to $(p_t - p_{t-1}) / p_t$. Dividends and stock splits are excluded here because (1) they are rare events and (2) we have already considered confounding events. Thus, stock return of a certain company equals to the change in stock price or the capital gain. R_{mt} stands for the corresponding market return at period t and is estimated by the CRSP equally weighted index. The CRSP equally weighted index is the average of the returns of all trading stocks in NYSE, AMEX and NASDAQ. β_0 and β_1 are the parameters and estimated in a 255-day periods ending at 45 days before the estimation window we choose by ordinary least square (OLS) method. We calculate the abnormal return (AR) from the market model:

$$AR_{it} = R_{it} - \hat{\beta}_0 - \hat{\beta}_1 R_{mt} \quad \text{Eq. B.2}$$

As shown by Eq. B.2, abnormal return is the return that cannot be captured by the market as a whole or the ex post return over the event window minus the normal return. The total effect of an economic event on stock price is reflected in mean cumulative abnormal return, which is the summation of abnormal returns for company-event observations in 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 trading day for the window we choose. Cumulative abnormal return (CAR, $\sum_{t_0}^{t_1} AR_{it}$) for each observation is used for the cross-sectional analysis.

The Fama-French three-factor model (Fama and French 1992) is

$$R_{it} = \alpha + \beta_i R_{mt} + s_i SMB_t + h_i HML_t + \varepsilon_{it} \quad \text{Eq. B.3}$$

where R_{it} is company i's return of the common stock at period t, R_{mt} is the return of a market index at period t, SMB_t is the average return on small market-capitalization portfolios minus the average of three large market-capitalization portfolios, HML_t is the average return on two high book-to-market equity portfolios minus the average on two low book-to-market equity portfolios. See Fama and French (1992) for a detailed explanation. β_i, s_i, and h_i are the parameters and estimated in a 255-day periods ending at 45 days before the estimation window we choose by ordinary least square (OLS) method.. The abnormal return (AR) is calculated as

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

Based on the abnormal return, the mean cumulative abnormal return and cumulative abnormal return can be calculated as described above.

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

$$V_{it} = \alpha + \beta\, V_{mt} + \varepsilon_{it} \quad \text{Eq. B.5}$$

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 above 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 C. Cluster Analysis and Concept Links

The cluster analysis is performed as follows using SAS® 9.1 Text Miner. First, text parsing decomposes the sentences into terms and creates a frequency matrix as a quantitative representation of the input documents. When decomposing the documents, we choose to rule out definite as well as indefinite articles, conjunctions, auxiliaries, prepositions, pronouns and interjections since these terms do not help provide meaningful results in our context. This matrix also shows the weight for the terms. The weight for term i in document j (w_{ij}) is the multiplication of the frequency weight (L_{ij}) and the term weight (G_i). In our study, the frequency weight is 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) is calculated as $1 + \sum_j \left(p_{ij} \log_2 \left(p_{ij} \right) / \log_2 n \right)$, where $p_{ij} = f_{ij} / g_{ij}$, g_i is the number of times term i appears in the dataset, and n is the number of documents in the dataset. These two methods put more weights on words that show in few documents and generally give the best results (SAS Institute Inc 2004). For dimension reduction, we use the single value decomposition (SVD) method. SVD generates the dimensions that best represent the original frequency matrix. The singular value decomposition of a frequency matrix (A) is to factorize the matrix into matrices of orthonormal columns and a diagonal matrix of singular values, i.e., $A = U\Sigma V^T$. Then the original documents are projected to matrix U (SAS Institute Inc 2004). Through matrix factorization and projection, SVD forms the dimension-reduced matrix. In our analysis, we set the maximum reduced dimensions to be one hundred (as default) and test three different levels of reduced dimensions (high, medium and low resolutions) as a robustness check. The resulting SVD dimensions are further used for cluster analysis. We then divide our data into disjoint groups using expectation maximization clustering by setting the maximum clusters to be forty (as default). The expectation maximization method is an iterative process that estimates the parameters in the mixture model probability density function which approximates that data distribution by fitting k cluster density function to a dataset. The mixture model probability density function evaluated at point x equals $\sum_{h=1}^{k} \omega_h f_h$, where μ_h, Σ_h are the mean vector and covariance matrix for cluster h under Gaussian probability distribution. For each observation x at iteration j, whether x belongs to a cluster h equals to $\left(\omega_h^j f_h \left(x | \mu_h^j, \Sigma_h^j \right) \right) / \left(\sum_i \omega_i^j f_i \left(x | \mu_i^j, \Sigma_i^j \right) \right)$ (SAS Institute Inc 2004). The iteration terminates if the likelihood value of two iterations is less than $\varepsilon > 0$ or a maximum $\left(x | \mu_h, \Sigma_h \right)$ of five iterations are reached (SAS Institute Inc 2004).

The concept links are determined based on the following criteria when all three of them are met: (1) Both terms occur in at least n documents, where n equals Max (4, A, B). A is the largest value of the number of documents that a term appears in divided by 100 and B is the 1000th largest value of the number of documents that a term appears in for concept links (SAS Institute Inc 2004), (2) Term 2 occurs when term 1 occurs at least 5% of the time (SAS Institute Inc 2004), and (3) The relationship between terms is highly significant (the chi-square statistic is greater than 12) (SAS Institute Inc 2004).

Appendix D. Implied Volatility

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

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

$p = Ke^{-rT}N(-d_2) - Se^{-qT}N(-d_1)$ Eq. 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 2006 for a detailed explanation), q is the continuously-compounded dividend yield (see Ivy DB Reference Manual 2006 for a detailed explanation), and σ is the historical volatility which equals the standard deviation of historic price change per share). In Eq. C.1 and Eq. C.2, d_1 equals $\left[\ln(S/K)+\left(r-q+1/2\sigma^2\right)T\right]/\sigma\sqrt{T}$ and d_2 equals $d_1-\sigma\sqrt{T}/2$.

Different from the historical volatility in Eq. C.1 and Eq. 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.

Appendix E. Variable Definitions

Variable	Definition
m	The online service or product provider's current market share which is defined between zero and one. It can be interpreted as the total value the provider can get from the customers comparing to other providers.
α	The percentage of information a customer needs to provide in order to complete the transaction which is defined between zero and one.
L	The compensation paid to customers or the legal penalty or fine when system fails.
ρ	Proportion of privacy sensitive customers which is defined between zero and one.
δ	Proportion of convenience sensitive customers which is defined between zero and one.
$F_n(t)$	The probability of system failure (CDF) of one non-repairable component across time t.
λ	Mean-time-to-failure
b	Change of failure rate across time
$F_{nn}(t)$	The probability of system failure (CDF) of two non-repairable component across time t.
ψ	False acceptance rate (FAR) of a biometric system which is determined by the selected threshold.
φ	False rejection rate (FRR) of a biometric system which is determined by the selected threshold.
$\bar{s}$	The threshold for the biometric system
$F_{bio}(t; \bar{s})$	The probability of system failure (CDF) of biometric system across time t.
w_{FRR}	The weight for FRR when choosing biometric systems
w_{FAR}	The weight for FAR when choosing biometric systems
$F_{nbio}(t; \bar{s})$	The probability of system failure (CDF) of one non-repairable component and one biometric component across time t.
C	The expected costs and losses
c	Implementation costs of the system
V	The loss of the value of customers as the system fails
ε	The percentage change of customers. We use ten different percentages for our analysis. ε_1 (ε_4, ε_7, ε_{10}) represents the percentage of customer a provider could lose when system fails under the base case (the biometric system, two non-repairable component system, one non-repairable and one biometric component system). ε_2 (ε_5, ε_8) represents the percentage of convenient sensitive customer a provider could lose when shifting to the biometric system (two non-repairable component system, one non-repairable and one biometric component system). ε_3 (ε_6, ε_9) represents the percentage of privacy sensitive customer a provider could attract when shifting to the biometric system (two non-repairable component system, one non-repairable and one biometric component system).

Appendix F. Conditions that Make the New Authentication System More Preferable

Panel A. Shift to biometric system

implementation costs

$$c_{bio} < c_n - V_{net\_bio} - F_{bio}(t;\bar{s})(V_{bio} + L_{bio}) + F_n(t)(V_n + L_n)$$

if $c_n - V_{net\_bio} - F_{bio}(t;\bar{s})(V_{bio} + L_{bio}) + F_n(t)(V_n + L_n) > 0$

percentage of privacy sensitive customers:

$$\frac{-Y - \sqrt{Y^2 - 4XZ}}{2X} < \rho < \frac{-Y + \sqrt{Y^2 - 4XZ}}{2X}$$

$$X = F_{bio}(t;\bar{s})(1-m)\varepsilon_3\varepsilon_4$$

$$Y = F_{bio}(t;\bar{s})(m\varepsilon_4 - m\delta\varepsilon_2\varepsilon_4) - (1-m)\varepsilon_3 - F_n(t)m\varepsilon_1$$

$$Z = c_{bio} - c_n + m\delta\varepsilon_2 + F_{bio}(t;\bar{s})L_{bio} - F_n(t)L_n$$

if $-Y + \sqrt{Y^2 - 4XZ} > 0$ and $-Y - \sqrt{Y^2 - 4XZ} > 0$

percentage of convenience sensitive customers:

$$\delta < \frac{c_n - c_{bio} + (1-m)\rho\varepsilon_3(1 - F_{bio}(t;\bar{s})\rho\varepsilon_4) - F_{bio}(t;\bar{s})(m\rho\varepsilon_4 + L_{bio}) + F_n(t)(V_n + L_n)}{m\varepsilon_2(1 - F_{bio}(t;\bar{s})\rho\varepsilon_4)}$$

if $c_n - c_{bio} + (1-m)\rho\varepsilon_3(1 - F_{bio}(t;\bar{s})\rho\varepsilon_4) - F_{bio}(t;\bar{s})(m\rho\varepsilon_4 + L_{bio}) + F_n(t)(V_n + L_n) > 0$

market share:

$$m > \frac{c_{bio} - c_n - \rho\varepsilon_3 + F_{bio}(t;\bar{s})\rho^2\varepsilon_3\varepsilon_4 + F_{bio}(t;\bar{s})L_{bio} - F_n(t)L_n}{\rho\varepsilon_3 + \rho\varepsilon_1 F_n(t) - \delta\varepsilon_2(1 - F_{bio}(t;\bar{s})\rho\varepsilon_4 - \rho\varepsilon_4 F_{bio}(t;\bar{s})(1 - \rho\varepsilon_3))}$$

if both the denominator and nominator are positive or negative

expected losses:

$$F_{bio}(t;\bar{s})L_{bio} < c_n - c_{bio} - m\delta\varepsilon_2 + (1-m)\rho\varepsilon_3 - F_{bio}(t;\bar{s})V_{bio} + F_n(t)(V_n + L_n)$$

if $c_n - c_{bio} - m\delta\varepsilon_2 + (1-m)\rho\varepsilon_3 - F_{bio}(t;\bar{s})V_{bio} + F_n(t)(V_n + L_n) > 0$

Panel B. Shift to two non-repairable component authentication system

implementation costs: $c_{nn} < c_n - V_{net\_nn} - F_{nn}(t)(V_{nn} + L_{nn}) + F_n(t)(V_n + L_n)$ if $c_n - V_{net\_nn} - F_{nn}(t)(V_{nn} + L_{nn}) + F_n(t)(V_n + L_n) > 0$
percentage of privacy sensitive customers: $\frac{-Y - \sqrt{Y^2 - 4XZ}}{2X} < \rho < \frac{-Y + \sqrt{Y^2 - 4XZ}}{2X}$ $X = F_{nn}(t)(1-m)\varepsilon_6\varepsilon_7$ $Y = F_{nn}(t)(m\varepsilon_7 - m\delta\varepsilon_5\varepsilon_7) - (1-m)\varepsilon_6 - F_n(t)m\varepsilon_1$ $Z = c_{nn} - c_n + m\delta\varepsilon_5 + F_{nn}(t)L_{nn} - F_n(t)L_n$ if $-Y + \sqrt{Y^2 - 4XZ} > 0$ and $-Y - \sqrt{Y^2 - 4XZ} > 0$
percentage of convenience sensitive customers: $\delta < \frac{c_n - c_{nn} + (1-m)\rho\varepsilon_6(1 - F_{nn}(t)\rho\varepsilon_7) - F_{nn}(t)(m\rho\varepsilon_7 + L_{nn}) + F_n(t)(V_n + L_n)}{m\varepsilon_5(1 - F_{nn}(t)\rho\varepsilon_7)}$ if $c_n - c_{nn} + (1-m)\rho\varepsilon_6(1 - F_{nn}(t)\rho\varepsilon_7) - F_{nn}(t)(m\rho\varepsilon_7 + L_{nn}) + F_n(t)(V_n + L_n) > 0$
market share: $m > \frac{c_{nn} - c_n - \rho\varepsilon_6 + F_{nn}(t)\rho^2\varepsilon_6\varepsilon_7 + F_{nn}(t)L_{nn} - F_n(t)L_n}{\rho\varepsilon_6 + \rho\varepsilon_1 F_n(t) - \delta\varepsilon_5(1 - F_{nn}(t)\rho\varepsilon_7 - \rho\varepsilon_7 F_{nn}(t)(1 - \rho\varepsilon_6))}$ if both the denominator and nominator are positive or negative
expected losses: $F_{nn}(t)L_{nn} < c_n - c_{nn} - m\delta\varepsilon_5 + (1-m)\rho\varepsilon_6 - F_{nn}(t)V_{nn} + F_n(t)(V_n + L_n)$ if $c_n - c_{nn} - m\delta\varepsilon_5 + (1-m)\rho\varepsilon_6 - F_{nn}(t)V_{nn} + F_n(t)(V_n + L_n) > 0$

Panel C. Shift to one non-repairable component and one biometric authentication system

implementation costs: $c_{nbio} < c_n - V_{net\_nbio} - F_{nbio}(t;\bar{s})(V_{nbio} + L_{nbio}) + F_n(t)(V_n + L_n)$ if $c_n - V_{net\_nbio} - F_{nbio}(t;\bar{s})(V_{nbio} + L_{nbio}) + F_n(t)(V_n + L_n) > 0$
percentage of privacy sensitive customers: $\frac{-Y - \sqrt{Y^2 - 4XZ}}{2X} < \rho < \frac{-Y + \sqrt{Y^2 - 4XZ}}{2X}$ $X = F_{nbio}(t;\bar{s})(1-m)\varepsilon_9\varepsilon_{10}$ $Y = F_{nbio}(t;\bar{s})(m\varepsilon_{10} - m\delta\varepsilon_8\varepsilon_{10}) - (1-m)\varepsilon_9 - F_n(t)m\varepsilon_1$ $Z = c_{nbio} - c_n + m\delta\varepsilon_8 + F_{nbio}(t;\bar{s})L_{bio} - F_n(t)L_n$ if $-Y + \sqrt{Y^2 - 4XZ} > 0$ and $-Y - \sqrt{Y^2 - 4XZ} > 0$
percentage of convenience sensitive customers: $\delta < \frac{c_n - c_{nbio} + (1-m)\rho\varepsilon_9(1 - F_{nbio}(t;\bar{s})\rho\varepsilon_{10}) - F_{nbio}(t;\bar{s})(m\rho\varepsilon_{10} + L_{nbio}) + F_n(t)(V_n + L_n)}{m\varepsilon_8(1 - F_{nbio}(t;\bar{s})\rho\varepsilon_{10})}$ if $c_n - c_{nbio} + (1-m)\rho\varepsilon_9(1 - F_{nbio}(t;\bar{s})\rho\varepsilon_{10}) - F_{nbio}(t;\bar{s})(m\rho\varepsilon_{10} + L_{nbio}) + F_n(t)(V_n + L_n) > 0$
market share: $m > \frac{c_{nbio} - c_n - \rho\varepsilon_9 + F_{nbio}(t;\bar{s})\rho^2\varepsilon_9\varepsilon_{10} + F_{nbio}(t;\bar{s})L_{nbio} - F_n(t)L_n}{\rho\varepsilon_9 + \rho\varepsilon_1 F_n(t) - \delta\varepsilon_8(1 - F_{nbio}(t;\bar{s})\rho\varepsilon_{10} - \rho\varepsilon_{10}F_{nbio}(t;\bar{s})(1 - \rho\varepsilon_9))}$ if both the denominator and nominator are positive or negative
expected losses: $F_{nbio}(t;\bar{s})L_{nbio} < c_n - c_{nbio} - m\delta\varepsilon_8 + (1-m)\rho\varepsilon_9 - F_{nbio}(t;\bar{s})V_{nbio} + F_n(t)(V_n + L_n)$ if $c_n - c_{nbio} - m\delta\varepsilon_8 + (1-m)\rho\varepsilon_9 - F_{nbio}(t;\bar{s})V_{nbio} + F_n(t)(V_n + L_n) > 0$

Panel D. Compare two non-repairable component system to one non-repairable component and one biometric authentication system (conditions when two non-repairable component system is more preferable)

implementation costs:

$$c_{nn} < c_{nbio} - V_{net\_nn} + V_{net\_nbio} - F_{nn}(t)(V_{nn} + L_{nn}) + F_{nbio}(t;\bar{s})(V_{nbio} + L_{nbio})$$

if $c_{nbio} - V_{net\_nn} + V_{net\_nbio} - F_{nn}(t)(V_{nn} + L_{nn}) + F_{nbio}(t;\bar{s})(V_{nbio} + L_{nbio}) > 0$

percentage of privacy sensitive customers:

$$\rho < \frac{-Y - \sqrt{Y^2 - 4XZ}}{2X} \text{ or } \rho > \frac{-Y + \sqrt{Y^2 - 4XZ}}{2X}$$

$$X = F_{nbio}(t;\bar{s})(1-m)\varepsilon_9\varepsilon_{10} - F_{nn}(t)(1-m)\varepsilon_6\varepsilon_7$$

$$Y = F_{nbio}(t;\bar{s})(m\varepsilon_{10} - m\delta\varepsilon_8\varepsilon_{10}) - (1-m)(\varepsilon_6 - \varepsilon_9) - F_{nn}(t)(m\varepsilon_7 - m\delta\varepsilon_5\varepsilon_7)$$

$$Z = c_{nbio} - c_{nn} + m\delta(\varepsilon_5 - \varepsilon_8) + F_{nbio}(t;\bar{s})L_{nbio} - F_{nn}(t)L_{nn}$$

if $-Y + \sqrt{Y^2 - 4XZ} > 0$ and $-Y - \sqrt{Y^2 - 4XZ} > 0$

percentage of convenience sensitive customers:

$$\delta < \frac{c_{nn} - c_{nbio} + (1-m)\rho(\varepsilon_9 - \varepsilon_6) - F_{nbio}(t;\bar{s})[(m + (1-m)\rho\varepsilon_9)\rho\varepsilon_{10} + L_{nbio}]}{m(\varepsilon_8 - \varepsilon_5) + F_{nn}(t)m\rho\varepsilon_5\varepsilon_7 - F_{nbio}(t;\bar{s})m\rho\varepsilon_8\varepsilon_{10}} + \frac{F_{nn}(t)[(m + (1-m)\rho\varepsilon_6)\rho\varepsilon_7 + L_{nn}]}{m(\varepsilon_8 - \varepsilon_5) + F_{nn}(t)m\rho\varepsilon_5\varepsilon_7 - F_{nbio}(t;\bar{s})m\rho\varepsilon_8\varepsilon_{10}}$$

if $c_{nn} - c_{nbio} + (1-m)\rho(\varepsilon_9 - \varepsilon_6) - F_{nbio}(t;\bar{s})[(m + (1-m)\rho\varepsilon_9)\rho\varepsilon_{10} + L_{nbio}] + F_{nn}(t)[(m + (1-m)\rho\varepsilon_6)\rho\varepsilon_7 + L_{nn}] > 0$

market share:

$$m > \frac{c_{nn} - c_{nbio} - \rho(\varepsilon_6 - \varepsilon_9) + F_{nn}(t)(\rho^2\varepsilon_6\varepsilon_7 + L_{nn}) - F_{nbio}(t;\bar{s})(\rho^2\varepsilon_9\varepsilon_{10} + L_{nbio})}{\rho(\varepsilon_9 - \varepsilon_6) + \delta(\varepsilon_8 - \varepsilon_5) + F_{nn}(t)\rho\varepsilon_7(\rho\varepsilon_6 - \delta\varepsilon_5 - 1) + F_{nbio}(t;\bar{s})\rho\varepsilon_{10}(1 - \delta\varepsilon_8 - \rho\varepsilon_9)}$$

if both the denominator and nominator are positive or negative

expected losses:

$$F_{nn}(t)L_{nn} < c_{nbio} - c_n + V_{net\_nbio} - V_{net\_nn} + F_{nbio}(t;\bar{s})(V_{nbio} + L_{nbio}) - F_{nn}(t)V_{nn}$$

if $c_{nbio} - c_n + V_{net\_nbio} - V_{net\_nn} + F_{nbio}(t;\bar{s})(V_{nbio} + L_{nbio}) - F_{nn}(t)V_{nn} > 0$

VITA

VITA

EDUCATION

Doctor of Philosophy (MIS, minor in Accounting), August 2009
Purdue University

Master of Business Administration (Accounting), 2002
National Taiwan University, Taiwan
Thesis Title: An XBRL Taxonomy for Taiwan's Commercial and Industrial Companies

Bachelor of Business Administration (Accounting), 2000
National Taiwan University, Taiwan

WORK EXPERIENCE

2005-present, Graduate Assistant, Krannert Graduate School of Management, Purdue University
2004-2005, Teaching Assistant, College of Management, National Taiwan University
2002-2004, Second lieutenant, Taiwan Air Force
2000, Assistant Auditor, Cheng-Yeh CPA Firm, Taiwan

PROFESSIONAL CERTIFICATION

- Certified Public Accountant, Taiwan, 2000
- Certified Internal Auditor, 2002

RESEARCH AWARDS AND HONORS

- Bilsland Dissertation Fellowship, Purdue University, 2008-2009
- Purdue Research Foundation Summer Research Grant, 2008
- Ross Fellowship, Purdue University, 2005-2006
- ICIS Doctoral Consortium, Paris, France, 2008
- AMCIS Doctoral Consortium, Toronto, Canada, 2008
- Institute for Information Infrastructure Protection (I3P) scholarship for WEIS 2008
- Taiwan Ministry of Education scholarship for WEIS 2008
- Taiwan Ministry of Education scholarship for 2007 Annual Meeting of the Academy of Management
- Best master thesis award, Dragon League Academic Competition, National Taiwan University, 2002

TEACHING AWARDS AND HONORS

- Graduate Student Award for Outstanding Teaching, Purdue University, 2009
- The Krannert Dean's Certificate for Distinguished Teaching, Purdue University, 2008
- The Krannert Dean's Certificate for Outstanding Teaching, Purdue University, 2006

COURSES TAUGHT

Instructor

- MGMT 382, Introduction to Management Information Systems, Purdue University, Spring 2008, overall instructor score 4.4/5.0
- MGMT 482, Introduction to Management Information Systems, Purdue University, Fall 2006, overall instructor score 4.0/5.0
  (MGMT 382/482 is a required course for all undergraduate management majors)

Teaching Assistant

- MGMT 382, Introduction to Management Information Systems, Purdue University, Spring 2007 to Spring 2008
- MGMT 482, Introduction to Management Information Systems, Purdue University, Fall 2006
- Freshman Economics, National Taiwan University, Fall 2004 to Spring 2005
- Intermediate Accounting, National Taiwan University, Fall 2004 to Spring 2005
- Cost and Managerial Accounting, National Taiwan University, Fall 2004 to Spring 2005
- Managerial Accounting, Division of Continuing Education and Professional Development, National Taiwan University, Fall 2004 to Spring 2005
- Accounting Information Systems, National Taiwan University, Fall 2000 to Spring 2002
- Computer Security and Auditing, National Taiwan University, Fall 2000 to Spring 2002

DISSERTATION

Essays on Information Security from an Economic Perspective

PUBLICATIONS

- "Knowledge management systems and organizational knowledge processing challenges: A field experiment," with Jungpil Hahn, in *Decision Support Systems*, Available online 12 March 2009
- "The recognition of the interest compensation of redeemable bonds—a review of alternative methods," with H. J. Lin and C. Y. Tseng, in *Journal of Contemporary Accounting*, Vol. 88 pp. 63-76, 2002.
- "XBRL—what facilitates accounting reports transferred electronically," with K. T. Chen, C. C. Chow, and P. J. Chen, in *Accounting Research Monthly*, Vol. 187, 188, 189, 2001.

PAPERS UNDER REVIEW

- "Cost and benefit analysis for two-factor authentication systems," with Kemal Altinkemer, under first round review at *Decision Support Systems*, April 2009.

WORKING PAPER

- "The association between information security risk factors and breach announcements: a design science approach" with Jackie Rees and Karthik Kannan, July, 2009.
- "Investors' reactions to information security incidents and profitable short-term investment opportunities," with Karthik Kannan and Jackie Rees, July, 2009.
- "Does web disclosure matter?" with Susan Watts and Mark Bagnoli, July, 2009.

TECHNICAL REPORTS

- "Research on the compensation scheme and bonus plan of Taiwan Postal Service," with H. J. Lin, T. L. Li, and L. F. Tang, 2000.
- "Research on the improvement of operation for the Experimental Forest of College of Bio-Resources and Agriculture, National Taiwan University," with H. J. Lin, and C. S. Tsai, 1999.

CONFERENCE/WORKSHOP PRESENTATIONS

- "Investors' perceptions of information security incidents and profitable short-term investment opportunities", with Karthik Kannan and Jackie Rees, 10th Annual Information Security Symposium Poster Session, Purdue University, March 2009.
- "Does web reporting matter?" with Susan Watts and Mark Bagnoli, BKD Accounting Workshop, Krannert Graduate School of Accounting, Purdue University, January 23, 2009.
- "Investors' perceptions of information security incidents," with Karthik Kannan and Jackie Rees, the 20th Workshop on Information Systems and Economics (WISE 2008), Paris, France, December 2008.
- "Reading the disclosures with new eyes: bridging the gap between information security disclosures and incidents," with Jackie Rees and Karthik Kannan, the 7th Workshop on the Economics of Information Security (WEIS 2008), Hanover, NH, June 2008.
- "Reading the disclosures with new eyes: bridging the gap between information security disclosures and incidents," with Jackie Rees and Karthik Kannan, 9th Annual Information Security Symposium Poster Session, Purdue University, March 2008.
- "Text mining: major factors of information security risks disclosed in financial reports," with Jackie Rees, INFORMS 2007, Seattle, WA, November 2007.
- "Knowledge management systems and organizational knowledge processing challenges: A field experiment," with Jungpil Hahn, 2007 Annual Meeting of the Academy of Management, Philadelphia, PA, August 2007.
- "Do information security disclosures in financial reports mitigate the impact of information security incidents?" discussed at 2007 Big Ten Information Systems Symposium, Purdue University, 2007.
- "An XBRL taxonomy for Taiwan's commercial and industrial companies," with Kuo-Tay Chen, 2002 Conference of Accounting Theory and Practice, Taipei, Taiwan, November 2002.
- "Republic of China's taxonomy," with Kuo-Tay Chen, Business Reporting and Data Assurance Conference, Bryant College, RI, April 2002.

AD HOC REVIEWER

- 12th IBIMA International Conference
- Pacific Asia Conference on Information Systems (PACIS) 2009
- 2009 AAA Annual Meeting
- Electronic Commerce Research and Applications
- 2009 Midwest AAA Meeting
- Information Systems Management
- International Conference on Information Systems (ICIS) 2008
- The Seventh Workshop on the Economics of Information Security (WEIS 2008)
- 2007, 2008 & 2009 Annual Meeting of the Academy of Management
- Americas Conference on Information Systems (AMCIS) 2007 & 2008

SERVICE

- 2009 AAA Annual Meeting XBRL Session Moderator
- AACSB accreditation pursuing committee, College of Management, National Taiwan University, 2004-2005
- Voluntary Income Tax Assistant (VITA), 1999

PROFESSIONAL MEMBERSHIPS

- Association for Information Systems (AIS)
- The Institute for Operation Research and the Management Sciences (INFORMS)
- American Accounting Association (AAA)

REFERENCES

- Professor Jackie Rees (Chair)
  Krannert Graduate School of Management
  Purdue University
  West Lafayette, IN 47907
  (765) 494-0320
  jrees@purdue.edu

- Professor Karthik Kannan (Co-Chair)
  Krannert Graduate School of Management
  Purdue University
  West Lafayette, IN 47907
  (765) 494-3414
  kkarthik@purdue.edu

- Professor Kemal Altinkemer
  Krannert Graduate School of Management
  Purdue University
  West Lafayette, IN 47907
  (765) 494-9009
  kemal@purdue.edu

- Professor Susan Watts
  Krannert Graduate School of Management
  Purdue University
  West Lafayette, IN 47907
  (765) 494-4504
  swatts@purdue.edu