

ESSAYS ON INFORMATION SECURITY FROM AN ECONOMIC PERSPECTIVE:  
INFORMATION SECURITY DISCLOSURES, INVESTORS' PERCEPTIONS ON  
SECURITY INCIDENTS, AND TWO-FACTOR AUTHENTICATION SYSTEMS

Krannert Graduate School of Management

Purdue University

by

Ta-Wei Wang

October 2008

## TABLE OF CONTENTS

	Page
LIST OF TABLES .....	iv
LIST OF FIGURES .....	v
ABSTRACT .....	vi
CHAPTER 1. INTRODUCTION .....	1
CHAPTER 2. THE IMPACT OF INFORMATION SECURITY DISCLOSURES ON MARKET REACTIONS TO SECURITY BREACHES .....	5
2.1. Introduction .....	5
2.2. Literature Review .....	7
2.2.1. Information Security .....	7
2.2.2. Disclosures in Accounting .....	8
2.3. Research Framework and Hypotheses Development .....	10
2.4. Cross-Sectional Analysis .....	16
2.4.1. Sample Selection .....	16
2.4.2. Regression Models .....	18
2.4.3. Results .....	21
2.4.4. Robustness Tests .....	25
2.5. Text Mining .....	27
2.5.1. Classification Model .....	28
2.5.2. Comparison of the Disclosure Groups .....	32
2.6. Conclusions and Discussion .....	35
CHAPTER 3. INVESTORS' PERCEPTIONS ON SECURITY INCIDENTS AND PROFITABLE SHORT-TERM INVESTMENT OPPORTUNITIES .....	39
3.1. Introduction .....	39
3.2. Literature Review .....	41
3.2.1. Information Security .....	42
3.2.2. Trading Volume .....	42
3.2.3. Analysts' Forecasts .....	43
3.3. Theoretical Background and Hypothesis Development .....	44
3.4. Research Methodology .....	48
3.4.1. Identify Information Security Incidents .....	48
3.4.2. Estimate Abnormal Trading Volume .....	49
3.4.3. Analyze Analysts' Forecasts .....	50
3.4.4. Implied Volatility and Profitable Short-Term Investment Opportunities .....	52
3.5. Preliminary Empirical Results .....	54
3.6. Conclusion .....	57

CHAPTER 4. COST AND BENEFIT ANALYSIS OF TWO-FACTOR AUTHENTICATION SYSTEMS .....	59
4.1. Introduction .....	59
4.2. Literature Review .....	61
4.2.1. Authentication .....	61
4.2.2. Privacy .....	64
4.3. Model .....	64
4.3.1. Basic Settings .....	65
4.3.2. Probability of System Failure .....	66
4.3.3. Analysis .....	69
4.4. Managerial Implications.....	71
4.5. Conclusions .....	77
CHAPTER 5. CONCLUSIONS .....	79
BIBLIOGRAPHY .....	81
APPENDICES	
Appendix A. An Example of the Disclosures of Internal Control and Procedures.....	92
Appendix B. Examples of Risk Factors .....	94
Appendix C. Sample .....	96
Appendix D. Stock Price Reactions from Information Security Incidents .....	99
Appendix E. Cluster Analysis and Concept Links.....	100
Appendix F. Variable Definitions .....	102
Appendix G. Conditions that Make the New Authentication System More Preferable .....	104

## LIST OF TABLES

	Page
Table 2.1 Descriptive Statistics of Disclosures.....	18
Table 2.2 List of Variables.....	19
Table 2.3 Results for the Cross-Sectional Analysis.....	22
Table 2.4 Confusion Matrix of the Verifying Results .....	31
Table 2.5 Text Mining Results of Information Security Related Risk Factors .....	33
Table 3.1 Results for Equation (3-2).....	56

## LIST OF FIGURES

	Page
Figure 2.1 Timeline for Two Information Sets.....	12
Figure 2.2 Process Flow for the Classification Model.....	28
Figure 2.3 An Instance of Decision Tree.....	30
Figure 2.4 Examples of Concept Links.....	34
Figure 3.1 Trading Volume Change across Time.....	54
Figure 4.1 Types of Customers.....	66

## ABSTRACT

Information security has become a critical issue to most organizations. Given its importance, managers and researchers have strived to better assess the impact of information security threats and to better manage security risks. In this proposal, we attempt to better understand information security from three different perspectives that are discussed below.

The first essay investigates the relationship between the characteristics of information security related disclosures and the stock price reactions to security incidents through a cross-sectional analysis and text mining techniques. The results from the cross-sectional analysis demonstrate that the investors perceive security risk factors disclosed in financial reports as warnings to future incidents. Building on the findings from the cross-sectional analysis, the text mining results further show that the disclosures with action oriented terms are less likely to be inferred as warning to future incidents.

The second essay examines the investors' perceptions on the impact of security incidents on the breached firm's future performance. The preliminary results show that informed investors perceive security risks as part of a firm's daily operation risks and do not react negatively. This essay is still in progress. We plan to propose the use of implied volatility as a better measure that captures the informed investors' perception on the uncertainty of a firms' future performance. Last, we demonstrate possible profitable

short-term investment opportunities from breach announcements because of the information asymmetry among investors.

The third essay focuses on the decision of choosing two-factor authentication systems. By comparing the expected costs and losses of different authentication systems, this study provides suggestions on whether the two-factor authentication system is more preferable. The elements that managers need to consider are additional implementation costs, the value of customer switch, and expected losses. By following large firms' choice of authentication system and by setting the proper level of penalty and fines, this essay also suggests strategies for firms and regulators that make a new authentication system more preferable to the firms.

## CHAPTER 1. INTRODUCTION

Business nowadays relies heavily on information technology to perform daily operations. This increasing reliance on information technology raises the concerns about information security. Researchers and managers have strived to better understand and assess information security risks as well as the impact of information security incidents. Therefore, this proposal approaches the issues in information security from three different perspectives in order to provide insights about (1) the relationship between security disclosures and the impact of security incidents, (2) investors' perceptions on security incidents, and (3) the decision rules when determining authentication systems.

The first essay addresses the relationship between security disclosures and the market reactions to security incidents. Information security related disclosures in financial reports could formulate the expectation that the firm is either prepared for future incidents or sending out warnings about future incidents to avoid future lawsuits. The former could lower the impact of security incident on a firm's business value while the latter could make the impact larger. Given this lack of clarity of the association between security disclosures and the impact of security incidents on a firm's business value, the first essay attempts to understand how security disclosures affect market reactions to security breaches. To do so, the essay first quantitatively investigates the association between information security incidents and the corresponding stock price reactions, and

information security disclosures in annual reports through a cross-sectional analysis. Based on the association found in the cross-sectional analysis, this essay further qualitatively explores the contents within the disclosures that characterize the formulation of investors' perceptions using text mining techniques. The text mining section consists of two parts. The first part is the classification model. This model investigates whether different disclosure patterns lead to different possibility of future breach announcements. The association allows us to verify whether a certain disclosure pattern signals to future breaches (i.e., being perceived as warnings). The second part of the text mining section is the cluster analysis. In particular, different disclosure patterns are explored to provide insights about how the investors' perceptions are formed and how firms should appropriately disclose information security related risk factors.

The second essay investigates a more fundamental issue when understanding the impact of security incidents on a firm's business value. This issue is the investors' perceptions on the impact of security breaches. Investors' perceptions provide explanations to managers and researchers about what leads to the market reactions to security incidents. Also, understanding investors' perceptions could help general investors make better investment decisions by lowering information asymmetry among investors. By investigating the trading volume behavior after the breach announcement, this study is able to understand how the uninformed and informed investors' beliefs regarding the breached firm's future performance are revised. More importantly, how informed investors perceive the breach announcement? Therefore, the study then specifically investigates the informed investors' beliefs by using analysts' forecasts as the proxy. This study is still in progress. As a next step, this essay attempts to propose a

timely measure that reflects the informed investors' perceptions on the impact of security breach on the uncertainty of a firm's future performance. Specifically, this essay will investigate how the implied volatility in the option pricing model changes after the announcement of security incidents. Furthermore, the implication of implied volatility is verified with analysts' forecasts and the decision based on implied volatility is compared with that based on stock price reactions. The comparison results provide investment suggestions to investors. Last, this essay will demonstrate one investment strategy that could help investor take advantage of the information asymmetry among investors and make profit in the short-run.

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. However, it also accompanies with possible privacy concerns and inconvenience. This study 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 study provides suggestions on whether the new authentication system is more preferable.

This dissertation contributes to the field of information security in the following ways. Essay one and essay two provide two different perspectives when assessing and understanding the impact of security incidents. In particular, essay one emphasizes on how firms should disclose their concerns about information security. Since investors infer what the firm knows and what the firm's action is regarding information security from the disclosures, it is important for firms to convey their security policy and practices to the public appropriately. Essay two formally investigates how informed and

uninformed investors perceive the impact of security incident on a firm's future performance. More importantly, essay two proposes a new way for researchers and investors to understand the impact of security incidents on the uncertainty of a firm's future profit generating capability. 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 theoretical framework and both the quantitative and qualitative results are discussed in the subsections. Chapter 3 presents the second essay where the theoretical background and preliminary 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 proposal.

## CHAPTER 2. THE IMPACT OF INFORMATION SECURITY DISCLOSURES ON MARKET REACTIONS TO SECURITY BREACHES

### 2.1. Introduction

Information security related incidents often lead to a disruption in business. 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). The impact of such disruptions is also significant. CSI/FBI 2007 survey estimates that the total dollar amount of financial losses resulting from security breaches is approximately \$200,000 US dollars per firm (CSI/FBI 2007). Moreover, the number of security incidents reported by the attacked firms is fast growing (CERT 2007). Firms often convey concerns about such potential disruptions through financial report disclosures. Our paper focuses on disclosures related to information security.

Disclosures, in general, are relevant to issues involving information asymmetry between a firm and its investors. In the accounting literature, two different motivations are provided for disclosures. On the one hand, papers such as Dye (1985), Verrecchia (1983), and Verrecchia (2001), argue that a firm only discloses information that is positively correlated to its business value. On the other hand, papers such as Kasznik and Lev (1995), and Skinner (1994) present evidence that a firm discloses in order to reduce its legal and reputation costs from the disappointing information it expects. At the first glance, it is not clear which specific motivation would be applicable to

information security disclosures. If information security disclosures indicate preparedness for security incidents, consistent with the first motivation, the disclosures would have a positive impact on the valuation of the firm when an information security incident is observed. On the contrary, as with the second motivation, disclosure itself can also imply future litigation or reputation costs, which decrease future cash flows and also the valuation of the firm. Understanding which motivation is applicable should aid managers in deciding the extent of information security disclosures provided. If the first motivation holds, managers should encourage disclosure. However, if the second motivation holds, managers should be careful about how they convey their security practices to the public.

In light of this apparent lack of clarity, we seek to answer the following research questions: Do information security disclosures in financial reports mitigate or worsen stock price reactions when a firm faces information security incidents? What are the elements within these disclosures that have significant impact on stock prices and characterize these disclosures?

To answer these questions, we associate the information security incidents and stock price reactions to such incidents, with the disclosures in financial reports. For the disclosures, we employ two different sources. One is the voluntary disclosure of risk factors that firms include regarding their future performance and forward-looking statements. The other source is the internal control report, which is mandated by Sarbanes-Oxley Act (SOX) Section 404, describing the weaknesses of internal controls and financial systems. Using the data, we perform a cross-sectional analysis on the firm's stock price to various aspects of disclosures. Since how risk factors are disclosed

in financial reports and the readability of financial reports can affect investors' expectations (Katz 2001; Li 2006), we also analyze the contents of risk factor disclosures using text mining techniques. In particular, we first build a classification model to associate the breach announcement with the content of the disclosures. Then, we further explore the characteristics of the content and suggest ways to disclose security related risk factors. Thus, our paper provides a comprehensive investigation involving both quantitative and qualitative analyses.

The rest of the paper is organized as follows. We first review the literature on information security and disclosures. Building on the literature, the research framework and hypotheses are elaborated. Next, details of the cross-sectional analysis and the results are presented. In addition to the cross-sectional analysis, we further analyze the textual data of the disclosures. 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 information security. The other is the literature on disclosures in accounting.

### 2.2.1. Information Security

A majority of the information security literature focuses on technical issues but analytical and empirical studies in information security from an economic perspective are relatively limited. For instance, several studies have been done to address information

security investments analytically (e.g., Gordon and Loeb 2002; Gordon et al. 2003). Studies have also pointed out that information security breaches can result in material impacts of business operation, including physical and intangible impacts such as negative company image and loss of reputation (Glover et al. 2001; Warren and Hutchinson 2000). Further, several empirical studies investigate the impact of information security events on business value. Based on different methodologies and different datasets, some of the results show that there exist significant negative impacts (Alessandro et al. 2008; Cavusoglu et al. 2004; Ettredge and Richardson 2003; Garg et al. 2003), while others do not find such impact (Campbell et al. 2003; Hovav and D'Arcy 2003; Kannan et al. 2007). For example, Ettredge and Richardson (2003) investigate the impacts of the denial of service attacks which happened in February 2000 and attempt to determine which firm might suffer or benefit from similar incidents in the future. Their results demonstrate the existence of information transfer and show that the larger the firm, the larger the abnormal return. As another example, Kannan et al. (2007) also analyze short-term and long-term impacts of security announcements on market value and do not uncover a relationship between announcements and business value. Although our paper also considers security breach events, we focus on understanding the impact of information security disclosures.

### 2.2.2. Disclosures 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., Dye 1985; Verrecchia 1983). The disclosure decision also depends on whether such disclosure will provide information to competitors and depends on other mandatory disclosures (e.g., Darrough 1993; Eihorn 2005; Verrecchia 1983). Disclosure may also be used so as to reduce legal and reputation costs from bad news or when the firm faces earnings disappointments (Kasznik and Lev 1995; Skinner 1994). Specific to risk disclosures, one recent study by Jorgensen and Kirschenheiter (2003) has formally modeled managers' decisions on voluntarily disclosing a firm's risks. Furthermore, several empirical studies focus 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; Elliott and Jacobson 1994; King et al. 1990; Lang and Lundholm 1996; Lev and Penman 1990).

In this paper, we link both the above streams of research. To the best of our knowledge, Sohail (2006) and Balakrishnan et al. (2008) are the only two studies that have also linked these two streams. In Sohail's paper, he demonstrates that security disclosures themselves are positively related to stock price. His work solely focuses on disclosures but does not consider the relationship between the disclosures and subsequent information security incidents, which we consider. By including the incidents, we are able to better understand how disclosures formulate investors' expectations and, in turn, affect the business value. The other paper, Balakrishnan et al. (2008), focuses on the impact of SOX and investigates whether the timeliness of information induced by SOX

increases the quality of information disclosed to the market. It does so by analyzing 8-K reports (important events not covered by previous annual or quarterly reports such as material disposition of assets or bankruptcy) and drawing relationship between the disclosure of 8-K reports and stock market reactions. However, our paper has a different focus. We focus on the relationship among risk factors disclosed in financial reports (10-K or 20-F reports), information security incidents and stock price reactions to the incidents. Our paper is different from these two studies in that we not only analyze how the characteristics of information security incidents and disclosures in financial reports affect the valuation of a firm but also consider how investors react to disclosures and how firms can appropriately convey information security concerns or practices through disclosures.

### 2.3. Research Framework and Hypotheses Development

We develop our hypotheses based on the efficient market hypothesis (Fama 1970). According to it, a firm's business value at time  $t$ , denoted as  $V_t$ , can be expressed as the discounted value of expected future cash flows given all the available information until that time:

$$V_t = E \left\{ \sum_{i=t}^T \frac{x_i | \Phi_t}{\prod_{j=t}^i (1+r_j^t)} \right\} \quad (2-1)$$

In Equation (2-1),  $E$  is the expectation operator,  $T$  denotes the assumed terminal period which can be infinity,  $x_i | \Phi_t$  is the net cash flow in period  $i$  given the information  $\Phi_t$  available at time  $t$ , and  $r_j^t$  is the interest rate faced by the firm in period  $j$  at time  $t$ . Often, there is asymmetry in the information available to the firm and its investors. In

this paper, the asymmetry we deal with is with respect to information security risks/threats the firm faces. The security threats can be one of the following three types (Bowen et al. 2006; Gordon et al. 2006): (1) confidentiality, such as theft of source code or customer data, (2) integrity, such as a virus attack which deletes or alters files, or (3) availability, such as denial-of-service attacks. The threats can lead to both direct and indirect costs for the firm (Cavusoglu et al. 2004; Ettredge and Richardson 2003; Garg et al. 2003). The direct costs include the loss of productivity, the costs related to informing consumers, litigation costs, and etc. The indirect costs include the loss of future transactions with consumers (and partners) that may be unwilling to trust the firm (i.e., reputation costs). Therefore, as with any other type of risk, the investors' uncertainty regarding the risks can negatively affect the expectation of the future cash flow and also the valuation of the firm. Given the uncertainty, each firm decides whether to disclose the threats to its future cash flows to the investors (Jorgensen and Kirschenheiter 2003).

In the information security context, investors gain information ( $\Phi_t$  in Equation (2-1)) regarding the threats a firm faces (the timeline is provided in Figure 2.1) from two different sources. The first involves breach related information announced in the media and we denote it by  $\eta_{t+1}$ . The second involves information security disclosures submitted by the firm in financial reports and is represented by  $\varphi_t$ . Within the financial reports, information security related disclosures can occur in two different places. The first is the disclosure of internal control and procedures mandated by Sarbanes-Oxley Act (SOX) section 404 denoted by  $\varphi_{tl}$  (see Appendix A for an example). This disclosure is considered in the information security context because it points out threats to the *integrity*

of information used by the firms. The second is the list of risk factors or possible uncertainties regarding forward-looking statements that may adversely affect a firm's future performance including information security related risk factors represented by  $\varphi_{t2}$  (see Appendix B for examples). In general, our paper considers firms that are breached (i.e.,  $\eta_{t+1} \neq NULL$ ), and investigates how  $\varphi_t$  and  $\eta_{t+1}$  affect the change in a firm's business value, which is defined as  $\Delta V = V_{t+1} |_{\eta_{t+1} \neq NULL, \varphi_t} - V_{t+1} |_{\eta_{t+1} = NULL, \varphi_t}$ .

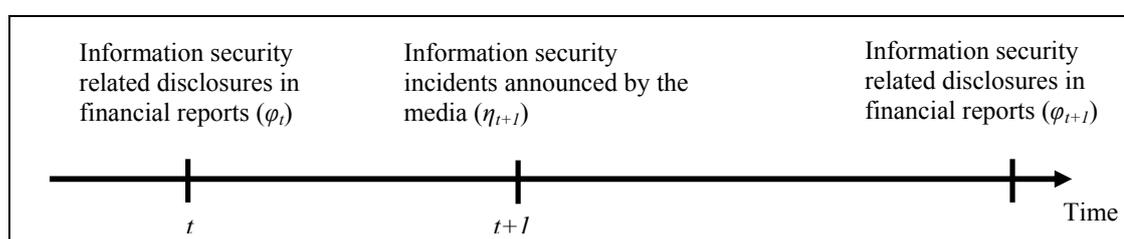


Figure 2.1 Timeline for Two Information Sets

In order to understand the impact of  $\varphi_t$ , we consider both the quantitative and qualitative nature of security disclosures. On the one hand, quantitatively, we count the number of elements within the internal control report for  $\varphi_{t1}$ , and the number of information security related risk factors mentioned by the firm in annual reports under the section of risk factors or the section of forward-looking statements for  $\varphi_{t2}$ . This measurement is consistent with the accounting literature (e.g., Francis et al. 1994; Lang and Lundholm 2000; Jo and Kim 2007). For our counting measurement, we posit that, since firms generally group several elements with similar consequences in one risk factor, investors also take these elements as a single factor and evaluate the impacts. On the other hand, qualitatively, we investigate the characteristics of security disclosures in the text mining section.

As the first hypothesis, we investigate the impact of  $\varphi_t$  on  $\Delta V$ . With information security disclosure, one particular concern is that it can expose the firm to the risks mentioned in the disclosure resulting in industrial espionage, loss of reputation and/or loss of competitive advantage (Gordon et al. 2005). Thus, the disclosure itself implies that the probability of incurring the costs is non-zero and, as a consequence, the future expected cash flows decreases. Despite the concern, we observe that firms disclose information security risk factors in their financial reports. The accounting literature (e.g., Kasznik and Lev 1995; Skinner 1994) argues that firms, in the cases where the future cash flows are expected to decrease due to disclosure, only disclose when the accompanied litigation and reputation costs from the threat are even larger. So, a breach ( $\eta_{t+1} \neq NULL$ ) signifies the realization of the probabilistic event where the litigation and reputation costs are incurred. This should drive investors to lower their expectation regarding future cash flows and, in turn, the business value. These imply that disclosure leads to  $V_{t+1} | \eta_{t+1} \neq NULL, \varphi_t \leq V_{t+1} | \eta_{t+1} = NULL, \varphi_t$  or simply  $\Delta V \leq 0$ . As  $\varphi_t$  increases, the realization of the probability of incurring the costs increases and hence, we hypothesize that  $\Delta V$  is negatively affected by  $\varphi_t$ . Formally:

*Hypothesis 1: For breached firms, as the number of internal control related items disclosed in the section of “Control and Procedures” ( $\varphi_{t1}$ ) and the number of disclosures of information security related risk factors ( $\varphi_{t2}$ ) increase, the impact of information security incidents on stock prices ( $\Delta V$ ) increases.*

Hypothesis 1 plays an important role in the paper. It not only leads to the cross-sectional analyses but also serves as the basis for exploring the contents within the disclosures in the text mining section.

Hypothesis 1 simply investigates the overall impact of disclosures. It does not distinguish between the natures of the disclosures, i.e., the relationship between  $\varphi_{t1}$  and  $\varphi_{t2}$ . Disclosures in Section 404 are mandated by the Sarbanes-Oxley Act whereas risks disclosed in the forward-looking statements are done so voluntarily. While the accounting literature has considered the mandatory and the voluntary disclosures to be independent of each other (e.g., see discussion in Eihorn 2005), there have been recent discussions regarding whether or not the two types of disclosures are correlated. Bagnoli and Watts (2007) analytically demonstrate that, when disclosures involve risks, the two types of disclosures are supplements, i.e., “the probability of voluntary risk disclosure is decreasing in the mandated amount of risk disclosures” (see Bagnoli and Watts 2007, p.904). Since we are dealing with information security risks, we expect the mandatory,  $\varphi_{t1}$ , and the voluntary disclosures,  $\varphi_{t2}$ , to also be supplements. In other words, the interaction between  $\varphi_{t1}$  and  $\varphi_{t2}$  should negatively affect market reactions to security incidents.

*Hypothesis 2: For breached firms, as the interaction between the number of internal control related items disclosed in the section of “Control and Procedures” ( $\varphi_{t1}$ ) and the number of disclosures of information security related risk factors ( $\varphi_{t2}$ ) increases, the impact of information security incidents on stock prices ( $\Delta V$ ) increases.*

An issue that Hypothesis 1, when hypothesizing about the impact of disclosure at the aggregate level, does not account for is the realization of the expectations. Prior literature has investigated the investors' reaction to the realization of the expectations. For example, Bagnoli et al. (2002), and Begley and Fischer (1998) study the investor reaction to whether a firm meets or misses the expected earnings report date. Similarly, Kasznik and McNichols (2002) study the reaction to realization, the so-called "meet or miss" earnings expectations. That is, whether the realization of an event meets investors' expectations built from disclosures can result in different stock price reactions. In our context, meeting expectations refers to the realization of the actual warning, i.e., information security incidents. Therefore, we suspect that the "match" between security related disclosures,  $\varphi_t$ , and incidents,  $\eta_{t+1}$ , is an important supplement to our argument in Hypothesis 1. Accordingly, in Hypothesis 3, we focus on the relationship between security related disclosures ( $\varphi_t$ ) and market reactions to security incidents ( $\Delta V$ ).

*Hypothesis 3: For breached firms, as the number of matched security related disclosure ( $\varphi_t$  matches  $\eta_{t+1}$ ) increases, the impact of information security incidents on stock prices ( $\Delta V$ ) increases.*

Yet another issue that has not been considered in Hypothesis 1 relates to the textual content of the disclosures. As shown by Katz (2001), how these risks are disclosed affects how the investors form expectations of the firm's future performance. Therefore, in order to fully investigate Hypothesis 1, we need to understand the qualitative contents of the disclosures. Accordingly, in addition to the quantitative analysis, we further explore the textual information of the disclosures. Particularly, we investigate the relationship between disclosure patterns and breach announcements in the text mining

section. In the following section, we test the above three hypotheses. Based on the results, in the section after that, we investigate the qualitative nature of the disclosures through text mining.

## 2.4. Cross-Sectional Analysis

In order to test our hypotheses, we first identify information security incidents. For the firms experiencing the incidents, we extract information security related disclosures from financial reports ( $\phi_{t1}$  and  $\phi_{t2}$ ), and the associated stock prices ( $\Delta V$ ). Based on the data collected, we investigate the relationship between stock price reactions and the disclosures in financial reports.

### 2.4.1. Sample Selection

To identify incidents, we search for news articles from 1997 to 2007 in the *Wall Street Journal*, *USA Today*, the *Washington Post*, and the *New York Times* via the Factiva database as well as in *CNet* and *ZDNet* with the following keywords: (1) security breach, (2) hacker, (3) cyber attack, (4) virus or worm, (5) computer break-in, (6) computer attack, (7) computer security, (8) network intrusion, (9) data theft, (10) identity theft, (11) phishing, (12) cyber fraud, and (13) denial of service. These keywords are similar to those used in prior studies (e.g., Campbell et al. 2003; Garg et al. 2003; Kannan et al. 2007). Only the samples with the following properties are retained in our dataset. First, the articles must enable us to identify a specific date of the security incident announcement. Second, only publicly traded firms are included in the analysis/sample. Third, only announcements from media are considered; we make sure that we do not

include any self-disclosed breaches on a firm's websites since those announcements may have a different impact than those from the media. Last, annual reports (10-K or 20-F filings) of the sample firms must be available one period prior to the event from EDGAR Online (<http://www.sec.gov/edgar.shtml>). The resulting sample consists of 112 firm-event observations. A list of the firms in our sample is provided in Appendix C. These breached firms are referred to as the *experimental group* in the rest of the paper.

For each incident, we collect the following data: (1) Information regarding the breached firm: the firm name, the industry identification code (SIC code), and CUSIP/PERM number for the firm's stock, (2) Security incident information: news source, date, and article. (3) Disclosures made in the financial report of the breached firm one period prior to the security incident: 10-K or 20-F filings depending on whether the firm is a foreign firm or not, elements from the section "Control and Procedures" ( $\varphi_{t1}$ ), and security related risk factors ( $\varphi_{t2}$ ) as well as other non-security related risk factors from the section of risk factors or forward-looking statement. As mentioned earlier, consistent with accounting literature (e.g., Francis et al 1994; Lang and Lundholm 2000; Jo and Kim 2007), we treat  $\varphi_{t1}$  and  $\varphi_{t2}$  as the counts of the number of risk factors disclosed. This measurement was evaluated by two independent raters and since the inter-rater reliability was high (Cohen's  $\kappa = 97.23\%$ ), the authors' coding results is used.<sup>1</sup>

The descriptive statistics regarding the disclosures, including the number of information security related risk factors and the total number of risk factors, are provided

---

<sup>1</sup> What we have done can be illustrated as follows. For instance, one risk factor disclosed by Amazon in year 2000 (see Appendix B) was "We face intense competition". The other was "System interruption and the lack of integration and redundancy in our systems may affect our sales". Thus, after looking into the content of the disclosures, we count one for information security related risk factors and two for the total risk factors in this case.

in Table 2.1. It can be easily seen that, on average, there is a greater number of security related disclosure and total number of risk factors disclosed per firm-event observation after SOX was introduced in 2002.

Table 2.1 Descriptive Statistics of Disclosures

Risk Factor Disclosures	Number of Security Related Risk Factors Disclosed		Number of Total Risk Factors Disclosed	
	before 2002	after 2002	before 2002	after 2002
Total	24	34	915	817
Average (stdev)	0.44 (1.014)	0.74 (1.063)	16.63 (9.358)	17.76 (9.562)
Max (min)	4 (0)	4 (0)	38 (0)	43 (0)

<sup>a</sup> SOX was enacted in 2002

#### 2.4.2. Regression Models

In order to test our hypotheses, we first focus exclusively on the primary model used for our *cross-sectional analysis*. We also validate our results through various robustness tests discussed later in another subsection.

The impact of economic events on business value can be measured by the stock price reactions in a short time period according to the theory of market efficiency (Fama 1970; MacKinlay 1997). To capture the impact of security incidents on stock price ( $\Delta V$ ), we apply the market model (which is described in detail in Appendix D) and obtained the cumulative abnormal return (CAR) through a two-day period (window) around the event date (the date of announcement, denote as day 0), i.e., -1~0, where -1 represents 1 day *before* the event date. To properly measure the impact of security incidents, samples with confounding events, such as earnings announcements, merger and acquisition, and

stock splits, are first eliminated so as to avoid other possible causes to the stock price reaction. Also, given the impact of consecutive events are not clear, we only include the first day of this type of event in our analysis. The resulting sample size is 101 firm-event observations for the experimental group. As mentioned earlier,  $\varphi_{t1}$  and  $\varphi_{t2}$  are evaluated by counting the number of disclosures.<sup>2</sup>

Table 2.2 List of Variables

CAR	Cumulative abnormal return (defined in Appendix A)
Size	Firm size which equals to the logarithm of net assets.
ConP	The number of elements a firm discloses in the section of the internal control report. There are three possible elements ( $ConP_1$ , $ConP_2$ , and $ConP_3$ ) which are explained below
$ConP_1$	Dummy variable for whether a firm discloses how it evaluates its internal controls and procedures. 1 if disclose, 0 otherwise.
$ConP_2$	Dummy variable for whether a firm discloses how it manages its internal controls and procedures. 1 if disclose, 0 otherwise.
$ConP_3$	Dummy variable for whether a firm discloses if there is a change in its internal controls and procedures. 1 if disclose, 0 otherwise.
Sec	Number of information security related risk factors disclosed in financial reports.
Nrisk	Total number of other non-security related risk factors disclosed in financial reports.
MSec	A subset of Sec. Number of matched disclosures.
PSec	A subset of Sec. Defined as MSec divided by Sec, i.e., the level of matched disclosures
$\varepsilon$	Residual term

We next specify the regression model to test Hypothesis 1. The variables used in this regression model as well as the others are listed in Table 2.2. Recall that Hypothesis 1 is about investigating the effect of the disclosures of internal control and procedures ( $\varphi_{t1}$ ) as well as information security risk factors ( $\varphi_{t2}$ ). Consistent with the general accounting literature (e.g., see discussion in Eihorn 2005), we also treat the voluntary disclosures to be independent from the mandatory disclosures. So, for

<sup>2</sup> We also perform the same set of the remaining analyses by replacing the number of disclosures with disclosure level. Specifically, we sort by the number of disclosures. Then the top 50% of the firms are named as high disclosers and the bottom 50% of the firms are named as low disclosers. Our results remain similar for high disclosers compared to low disclosers.

validating Hypothesis 1, we test the impact of the number of information security risks, *Sec*, reported in the forward-looking statements as well as the elements in the control report (*ConP*). We control for non-security related risks in the section of risk factors or forward-looking statements (*Nrisk*) as well as the firm size (*Size*), which is the logarithm of a firm's net assets). Firm size is controlled for because previous studies have shown that large firms are more able to endure shocks than small ones and they also invest more in security (Fama and French 1992; PriceWaterhouseCoopers 2002). Thus, for Hypothesis 1, we estimate the following:

$$CAR_i = \beta_0 + \beta_1 Size_i + \beta_2 ConP_i + \beta_3 Sec_i + \beta_4 Nrisk_i + \varepsilon_i \quad (2-2)$$

Hypothesis 1 expects  $\beta_2$  and  $\beta_3$  to be negative in the above equation.

There are three different elements in the internal control reports which are captured by three different binary variables, i.e., whether the firm discloses information about how a firm evaluates its internal controls and procedures (*ConP<sub>1</sub>*), whether the firm discloses information about how a firm manages its internal controls and procedures (*ConP<sub>2</sub>*), and whether the firm discloses information about whether it changes its internal controls and procedures (*ConP<sub>3</sub>*). These three variables equal one if the firm discloses, and zero otherwise. We also test if Hypothesis 1 holds when considering these elements separately.

$$CAR_i = \beta_0 + \beta_1 Size_i + \beta_2 ConP_{1i} + \beta_3 ConP_{2i} + \beta_4 ConP_{3i} + \beta_5 Sec_i + \beta_6 Nrisk_i + \varepsilon_i \quad (2-3)$$

From Hypothesis 1, we expect  $\beta_2$ ,  $\beta_3$ ,  $\beta_4$ , and  $\beta_5$  in Equation (2-3) to be negative for the experimental group.

For Hypothesis 2, we further test the interaction between the above two disclosures separately through Equation (2-4), and (2-5).

$$CAR_i = \beta_0 + \beta_1 Size_i + \beta_2 ConP_i + \beta_3 Sec_i + \beta_4 Nrisk_i + \beta_5 ConP_i \times Sec_i + \varepsilon_i \quad (2-4)$$

$$CAR_i = \beta_0 + \beta_1 Size_i + \beta_2 ConP_{1i} + \beta_3 ConP_{2i} + \beta_4 ConP_{3i} + \beta_5 Sec_i + \beta_6 Nrisk_i + \beta_7 ConP_{1i} \times Sec_i + \beta_8 ConP_{2i} \times Sec_i + \beta_9 ConP_{3i} \times Sec_i + \varepsilon_i \quad (2-5)$$

From Hypothesis 2, we expect  $\beta_5$  in Equation (2-4),  $\beta_7$ ,  $\beta_8$ , and  $\beta_9$  in Equation (2-5) to be negative for the experimental group.

Next, to validate Hypothesis 3, we investigate the impact on CAR when the disclosure coincides with the incident type announced in the media, i.e.,  $\varphi_t$  matches  $\eta_{t+1}$ . One can consider the absolute number of matched disclosures and the level of matched disclosures for both the elements in the internal control report and security related risk factors. However, since the number of matched disclosures for the elements in the internal control report is zero for all the observations, we do not consider the term in our model. Let  $MSec$  represent the number of matched information security risk factors and  $PSec$  be the percentage of matched security risk factors, i.e.,  $MSec$  divided by  $Sec$ . Here again, we control for non-security related risks disclosed in the financial reports and the size of the firm.

$$CAR_i = \beta_0 + \beta_1 Size_i + \beta_2 MSec_i + \beta_3 Nrisk_i + \varepsilon_i \quad (2-6)$$

$$CAR_i = \beta_0 + \beta_1 Size_i + \beta_2 PSec_i + \beta_3 Nrisk_i + \varepsilon_i \quad (2-7)$$

Based on Hypothesis 3, we expect  $\beta_2$  in Equations (2-6) and (2-7) to be negative. The following section reports the results from testing our hypotheses.

### 2.4.3. Results

The relationships between the disclosures of internal control and procedures as well as the disclosures of information security risk factors and CAR are given in Table 2.3.

For Hypothesis 1, consider the coefficient estimates shown in the columns “Equation (2-2)” and “Equation (2-3)” of Table 2.3. Notice that the impact of the disclosures of internal control and procedures is not significant (i.e., the coefficients for *ConP* and *ConP*<sub>1</sub> to *ConP*<sub>3</sub> are not significant). However, the number of security related risk disclosures (*Sec*) negatively affects CAR (significant at 1% level). Thus, Hypothesis 1 is partially supported. This finding also appears to support our hypothesis that the investors expect the realization of litigation costs and reputation costs when a breach occurs and lower their expectation regarding the future cash flows. In other words, our result appears to indicate that the second motivation in Introduction is valid.

Table 2.3 Results for the Cross-Sectional Analysis

Variables	Equation (2-2)	Equation (2-3)	Equation (2-4)	Equation (2-5)	Equation (2-6)	Equation (2-7)
Intercept	-0.03	-0.03	-0.03	-0.03	-0.03	-0.05
<i>Size</i>	0.00	0.00	0.00	0.00	0.00	0.00
<i>ConP</i>	0.00		0.00			
<i>ConP</i> <sub>1</sub>		0.01		0.01		
<i>ConP</i> <sub>2</sub>		0.00		0.00		
<i>ConP</i> <sub>3</sub>		0.00		0.00		
<i>Sec</i>	<b>-0.02***</b>	<b>-0.02***</b>	<b>-0.02***</b>	<b>-0.02***</b>		
<i>ConP</i> × <i>Sec</i>			0.00			
<i>ConP</i> <sub>1</sub> × <i>Sec</i>				0.00		
<i>ConP</i> <sub>2</sub> × <i>Sec</i>				0.00		
<i>ConP</i> <sub>3</sub> × <i>Sec</i>				0.00		
<i>MSec</i>					<b>-0.04***</b>	
<i>PSec</i>						<b>-0.03*</b>
<i>Nrisk</i>	0.00	0.00	0.00	0.00	0.00	0.00
N	101	101	101	101	101	101
Adj. R <sup>2</sup>	0.10	0.10	0.09	0.10	0.15	0.15
* significant at 10% ** significant at 5% ***significant at 1%						

For Hypothesis 2, consider the regression coefficients in columns “Equation (2-4)” and “Equation (2-5)” in Table 2.3. Notice that the coefficients are similar to the ones from Equations (2-2) and (2-3). Also, observe that the interaction terms are not

significant, indicating that the decision to voluntarily disclose risk factors does not depend on the amount of information disclosed mandatorily. Therefore, the argument about the supplemental relationship between mandatory and voluntary disclosures is not applicable here. We further investigate why the interaction effects are not significant. By reading through the text, it appears that the disclosures of internal control and procedures are relatively standardized across firms. They are primarily focused on regulation compliance and establishing if there is a weakness in the internal control and not so much about the threats. Perhaps, as a result, these disclosures do not provide useful information to investors from the information security standpoint. It may also explain our observation.

For Hypothesis 3, consider the regression coefficients presented in columns “Equation (2-6)” and “Equation (2-7)” in Table 2.3. Our finding supports Hypothesis 3, i.e., the number of matched disclosures and the level of match (*MSec* and *PSec*) both significantly affect stock price reactions. The next question is: How much more is the impact of the match? In order to address this, we compare the coefficient estimate of *MSec* with that of *Sec*. We first estimate Equation (2-1) by replacing *Sec* with *MSec*. The coefficients for this new estimation are found to be similar to those from estimating Equation (2-6). It implies that, to assess the additional impact of the match, we can simply compute the difference between the coefficient estimate of *MSec* from Equation (2-6) in Table 2.3 and that of *Sec* from Equation (2-2) in Table 2.3. We find the coefficient of *MSec* to be more negative (statistically significant at 5% level) than the one of *Sec*. We claim this result by calculating the magnitude of the difference as well as the variance of the difference. Since the coefficient of *Sec* indicates the impact of

realization of the reputation/litigation costs, the additional negative impact suggests that the match has an externality effect which is negative. Perhaps, the realization of one of the disclosed risk factors (the match) makes investors nervous about the realization of the other threats also.

In summary, the results above demonstrate a negative association between security disclosures and market reactions to security incidents which reflects the following: (1) investors lower their expectation about a firm's future profitability by taking into account the firm's disclosure about the uncertainty of information security events. (2) Since the disclosures of internal control and procedures are relatively standardized and have less "real" information, these disclosures do not provide useful information for investors from the information security standpoint. (3) The firm's stock price reactions are larger when the risks disclosed in the reports are realized. Our findings first show that managers need to disclose threats or security practices in financial reports with caution. Especially when investors perceive that the specific warning will occur, the impact is more severe. Our findings also demonstrate that auditors should also take into account these risk disclosures when assessing a firm's business risks and control risks since these disclosures might convey crucial information about a firm's future profitability or address a firm's ongoing concern. However, are there appropriate ways to convey security concerns in financial reports? We address this question in the text mining section and provide more specific suggestions for disclosure policies.

#### 2.4.4. Robustness Tests

In order to validate the robustness of our hypothesis tests, we also estimate other regression models. First, we validate if our results are indeed the consequences of breaches suffered by the firms (the experimental group). One common way to check it is by verifying if our results also hold for other firms without any reported incidents (see, for example, Shadish et al. 2002). If our results also hold for the other firms, then we have captured some spurious relationship between security disclosures and market reactions to security incidents. In order to verify this, we determine, for every firm in the experimental group, one of its publicly-traded competitors that does not have any breach announcements. We gather this information from Yahoo! Finance and the Hoover's Database. If several competitors can be selected, we choose the one with similar market capitalization and with financial reports available. A list of our *control firms* is also provided in Appendix C. We perform the same analyses discussed previously by using the control firms but do 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. This result is also valuable in some other regards. It points to the existence of systematic difference between the two sets of firms, which is exploited in the text mining section. More specifically, the text mining section employs disclosures from these control firms and the experimental group as inputs for the analyses.

When testing the hypotheses in the previous section, we had estimated CARs using the -1 to 0 window. We find the results were consistent for other short-term windows

also. For example, we tried the windows, -1 to 1 and -7 to 1, among others. Similarly, we tried different estimation periods for the parameters in the market model such as 255 before the incidents and 180 days before the incident, but again we do not find any significant deviation.

Lastly, we also validate our results after taking into account other firm-specific and event-specific characteristics. The firm-specific characteristics we considered are industry and attack history. Lev and Pennman (1990) have shown that firms in the same industry might have similar disclosure policies. Therefore, we also accounted for this industry effect as a robustness check. However, our results are not sensitive to the industry effect. Furthermore, some industries might rely more on information technology than others such as the high-tech firms. We also control for these firms and our results remain the same. Another firm-specific characteristic is attack history. We take into account a firm's history of attack since a frequently attacked firm might have different disclosure policies and/or stock price reactions to security incidents. However, even after including the attack history into our model, our results remain similar.

The event-specific characteristic we investigated is the incident type. For this, we used two raters to code the event types. The inter-reliability rating was high (Cohen's  $\kappa = 92.83\%$ ). Based on the coding, there are 43, 31 and 44 incidents of confidentiality, integrity, and availability type incidents, respectively (6 incidents are classified into both the integrity type and the availability type). Here again, our results largely remain similar after controlling for incident types.

## 2.5. Text Mining

The analysis thus far has focused only on the quantitative attributes of the disclosures. We now focus on mining the textual data to further understand how information security risk factors inform investors to build their expectations. Text mining, in general, has proven to be a useful tool in such scenarios to extract information through finding nontrivial patterns and trends (Feldman and Sanger 2006; Tan 1999). 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 (Cecchini et al. 2007; Fan et al. 2006; Han et al. 2002; Masand et al. 1992; Young and Hayes 1985). In our information security context, we apply text mining techniques to the contents of risk factor disclosures so as to identify and categorize the elements of the risk factors that affect investors' perceptions on whether the disclosure is a warning to future incident. If a certain disclosure pattern leads to the perception that the managers are providing warnings, we should observe that a certain disclosure pattern associates with the breach announcement with a higher probability. This section builds on the analyses of the previous section in many regards. First, building on the results, we provide qualitative explanations to the negative association between stock price reactions and disclosures. Second, we focus only on the disclosures of information security risk factors since other factors including the internal control and procedures are not significant. Third, it is clear from the robustness test that one can account for the association among disclosures, perceptions, and stock price reactions simply by considering the experimental and the control groups.

### 2.5.1. Classification Model

While we know that the disclosures have an impact on market reactions to security incidents, we would like to know if textual content in these disclosures leads to varying impacts. The resulting association allows us to understand differences in the disclosure practices and uncover the underlying terms within disclosures that result in different perceptions. To accomplish this, we build a classification model by adopting a three-step procedure portrayed in Figure 2.2 and detailed below:

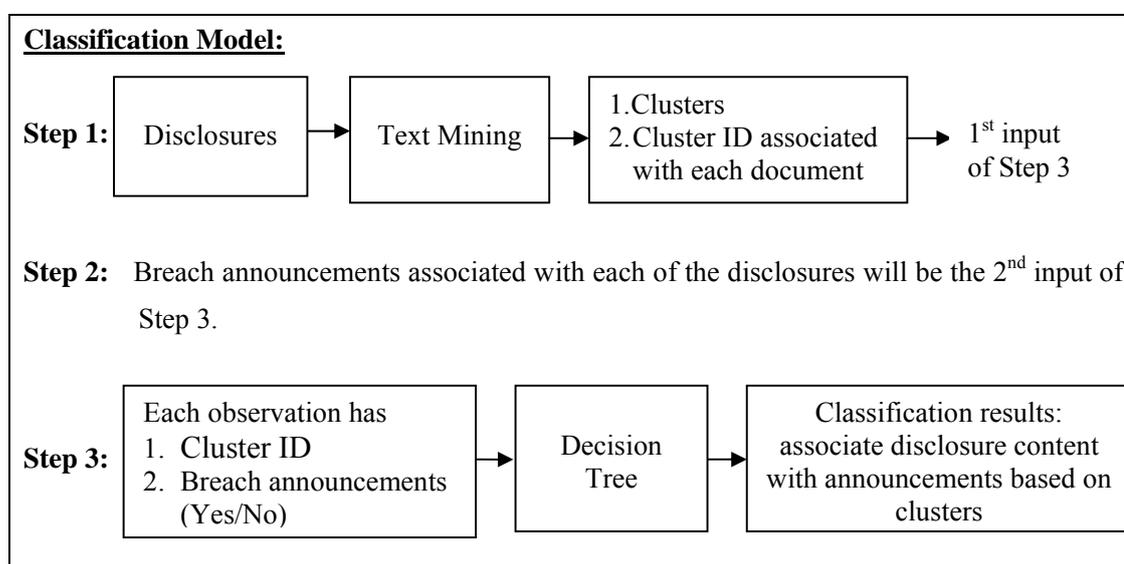


Figure 2.2 Process Flow for the Classification Model

In order to perform the analysis, we use disclosures from both the experimental and the control groups. We expand it by sampling 50 publicly-traded firms across different industries with different sizes that are not in the current sample and without breach announcements to add noise into our model. We were able to collect 23 new disclosures

from this sample. Note that, even without these additions, our results are largely similar.

Based on the data set, in the first step, we cluster the textual data in the expanded dataset involving 96 documents. We identified four unique clusters and each document is associated with a cluster ID from its associated cluster.

In the second step, we associate each disclosure with an indicator showing that whether the corresponding firm has breach announcement or not. If the disclosure is from the breached firm, the indicator shows “yes”, and shows “no” otherwise.

In order to perform the classification task, this new dataset (96 documents) is partitioned into three parts: training (80%), validation and testing (20%). Furthermore, when setting up the classifier (breach announcement), we set the prior probability of the classifier as the proportion of the number of related documents in the whole dataset. The classification model is trained, validated, and tested using a decision tree. We chose to use a decision tree due to its inherent transparency and interpretability which help users follow the path of the tree and understand the classification rules step by step (e.g., Baesens et al. 2003; Brandān et al. 2005; Kim et al. 2001; Zhang and Zhu 2006; Zhou and Jiang 2004). We tested other classification models (for example, neural networks) and obtained similar results. After the decision tree model is trained, we find that the resulting tree has two leaves from the root (see Figure 2.3 for an instance).

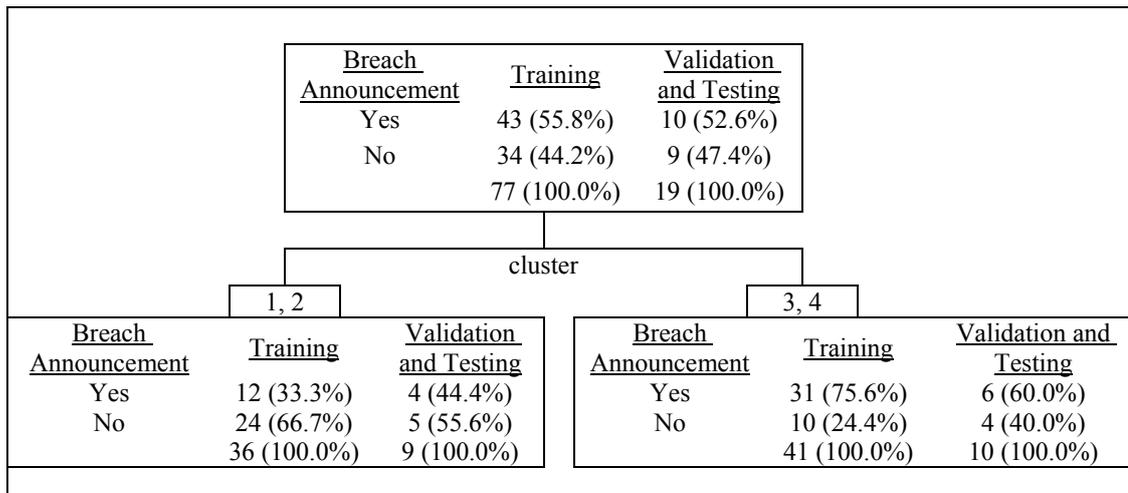


Figure 2.3 An Instance of Decision Tree

As shown in Figure 2.3, 77 and 19 documents are used for training, and validation and testing respectively. Furthermore, documents associated with cluster 1 and cluster 2 are classified into the left sub-tree and about 67% of them in the training dataset are associated with “no breach announcement”. Documents related to cluster 3 and cluster 4 are classified into the right sub-tree and about 76% of them in the training dataset are associated with “breach announcement”. However, since there are only 19 documents in the validation and testing dataset, this result needs to be further verified. In order to further verify our model, we use a commonly adopted procedure called 10-fold cross validation (e.g., Kohavi 1995; Weiss and Kapouleas 1989). The results from our 10-fold cross validation are given in Table 2.4. These results demonstrate that the overall accuracy rate for this model is about 71.79% (39.74%+32.05%).

Table 2.4 Confusion Matrix of the Verifying Results

Frequency Percentage Row Percentage Column Percentage	Predict			
	Breach Announcement	No Breach Announcement	Total	
Actual	Breach Announcement	31	13	44
		39.74	16.67	56.41
		70.45	29.55	
	No Breach Announcement	9	25	34
		11.54	32.05	43.59
		26.47	73.53	
Total	40	38	78	
	51.28	48.72	100.00	

This model demonstrates that there indeed exist textual differences between disclosures which lead to different perceptions. Also, it shows that there are two sets of clusters (cluster 1, cluster 2) and (cluster 3, cluster 4) that lead to different perceptions. Two interesting aspects of this model are worth noting. First, the high accuracy rate of the model suggests that a manager might be able to predict the impact of the disclosures on perceptions based on the contents disclosed. An even more interesting aspect is that 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 result in different reactions. Consequently, we further investigate the qualitative characteristics of the disclosures from these two sets of clusters labeled as Disclosure Group A (cluster 1 and cluster 2) and Disclosure Group B (cluster 3 and cluster 4) in the following section.

### 2.5.2. Comparison of the Disclosure Groups

In this section, we explore how the textual contents of disclosures from Disclosure Group A are different from those from Disclosure Group B. We perform the comparison at the aggregate level instead of focusing on clusters 1, 2, 3 and 4 separately. This is because some of these clusters have very few data points and are not amenable to any meaningful analysis. By comparing the textual contents of these two groups, we may be able to more closely link the characteristics of the disclosures with investors' perceptions.

We pool together all the disclosures from one group (Group A or Group B) of firms and identify clusters of textual content that commonly occur in that group (the process of identifying the clusters is a standard one and is detailed in Appendix E). Table 2.5 displays the clusters resulting from such a procedure for each Disclosure Group. Observe that each Disclosure Group has many clusters and each row in the table represents one cluster. Within each cluster, there are five terms. 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. Observe that, in Table 2.5, the top three clusters account for 50-100% of all disclosures. The discussion below will largely focus on these major clusters.

Table 2.5 Text Mining Results of Information Security Related Risk Factors

Cluster ID	Terms	Freq.	Percentage	RMS Std.
<b>Disclosure Group A (from cluster 1 and 2)</b>				
1	<b>+implement</b> <sup>a</sup> , <b>+protect</b> , <b>+require</b> , resource, +transaction	10	36%	0.1917
2	+affect, +breach, computer, +result, +security	10	36%	0.2319
3	compensate, +depend, <b>+interrupt</b> , +result, <b>+system</b>	8	29%	0.1919
<b>Disclosure Group B (from cluster 3 and 4)</b>				
1	+breach, confidential information, network, public, secure	13	22%	0.1561
2	+event, <b>+failure</b> , hardware, <b>+site</b> , web	12	21%	0.1502
3	+experience, +disaster, +disruption, +facility, <b>+failure</b>	7	12%	0.1562
4	adverse, +business, +customer, +product, software	6	10%	0.1759
5	+attack, +damage, denial, +disruption, vulnerable	6	10%	0.1432
6	capacity, data capacity, internet, +place, traffic	5	9%	0.0821
7	+activity, +breach, +incur, +relate, +report	5	9%	0.1473
8	+disaster, +employee, +loss, +risk, <b>+system</b>	4	7%	0.1439
<sup>a</sup> For readers' convenience, we highlight the examples discussed in the text as bolded and italicized.				

Recall that Disclosure Group A corresponds to the *no breach announcement group* while Disclosure Group B is related to *breach announcement group*. Since it appears from our classification model that the textual content of the disclosure is a pretty good predictor of the breach announcement, we associate here the clusters identified in Table 2.5 with the announcement. We assess the similarity between the clusters from the two groups by matching the terms. We obtain from two independent coders the measurements of matches between each cluster in Group A with every cluster in Group B. The reliability of the measurements is high (Cohen's  $\kappa = 80.00\%$ ). Based on codes, the clusters with IDs 2 and 3 in Table 2.5 under Disclosure Group A are similar to the clusters with IDs 1 to 3 in Disclosure Group B (the matches were between 60% and 100%). However, we find that the cluster ID 1 in Disclosure Group A has a very low similarity measure with other clusters in Group B. It possibly implies that the lack of terms about operations and actions such as “implement”, “protect”, and “require” in Group B lead to a negative interpretation of the disclosure.

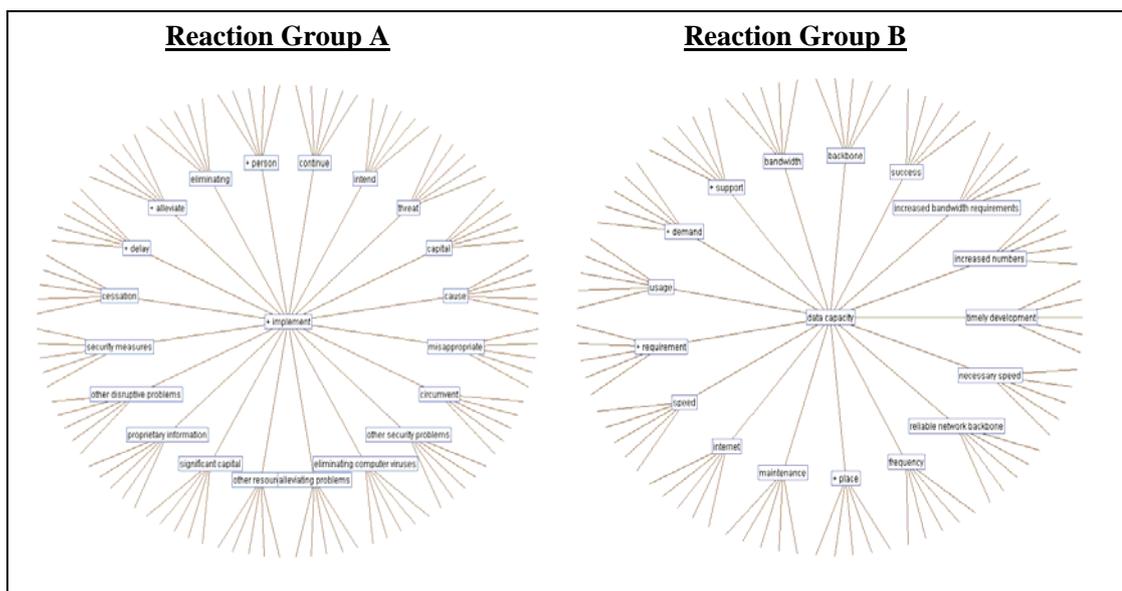


Figure 2.4 Examples of Concept Links

We use concept links (defined in Appendix E) between the terms in the cluster as another informal way to provide context of the terms and verify our observation regarding the action- or operation-related terms. An example concept link is shown in Figure 2.4. In that, for example, the terms “implement” and “alleviate” are related, implying that these two terms often co-occur. We checked the concept links for all the terms in clusters for both the Disclosure Group A and the Disclosure Group B. For Disclosure Group A, most of the terms with concept links are observed to be action terms such as “implement” (see Figure 2.4 for an example). However, for Disclosure Group B, most of the terms with concept links are more general with the phrases such as “breach”, “data capacity” and “infrastructure” (see Figure 2.4 for an example).

Summarizing the results from our text mining section, we identify two groups based on how the disclosures affect the perceptions. The high accuracy rate for our

classification model indicates that a manager can assess the potential impact of disclosures. Moreover, we build on the classification model and investigate the textual content of the disclosure. Specifically, we find that when disclosures involve action terms or terms about processes, the disclosures are less likely to lead to negative perceptions.

## 2.6. Conclusions and Discussion

We have often observed that firms disclose information security risks in the financial reports. However, as mentioned in the Introduction, it was not ex ante clear whether the disclosures indicate a positive (e.g., preparedness for such threats) or a negative (e.g., indicates potential litigation/reputation costs) signal. In order to clarify this issue, our paper investigates the relationship among information security related risks disclosed in the financial reports, investors' perceptions on disclosures, and the stock market reactions to information security incidents. The investigation is executed in two stages. First, using commonly accepted measures, we quantitatively study the impact of disclosures. For this analysis, we consider firms that suffer an information security incident between 1997 and 2007. For each firm, we use two different sources for information security related disclosures. One is the disclosures of internal control and procedures mandated by SOX. The other one is the disclosures of risk factors. The following are the key findings from our quantitative analyses. We find that, for a firm suffering a security breach, the more the number of disclosures of information security risk factors, the larger the impact of information security incidents. Also, after further investigating the argument of expectation formulation in detail, we find that the impact of information

security incidents on stock price reactions depends on whether the incidents match the content of the disclosures. These results indicate that the disclosures actually create the perception of warnings of future incidents.

Following that, we extend the quantitative analysis and perform textual analysis to assess the qualitative impact of disclosures. We first develop a classification model and demonstrate that the textual content of disclosure is a good predictor of investors' perceptions on the information conveyed through disclosures. Building on this, we consider the characteristics of disclosure that lead to different inference of the disclosures. We argue that firms, which disclose more actionable information when they provide information security risk factors, can alter the information investors could infer from the disclosures.

The results and analyses shed light to a manager on how they can convey security practices to their customers and investors more effectively. We observe that standardized disclosures of information security related issues (mandated by Section 404 of SOX) provide relatively little information from an information security standpoint. Also, 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.

One unique aspect about our paper is that it is based on various streams to provide actionable insights to the managers. Our hypothesis section builds on various theories on disclosure from the accounting literature. Consistent with that literature, we employ a quantitative measure to study the impact of disclosure. We test our hypotheses using the event study methodology. We also supplement our quantitative analysis with some

qualitative analyses. By analyzing the characteristics of the disclosed documents, we find certain interesting patterns in how disclosures could have different meanings to investors.

Our paper is not without its limitations. One of the major limitations of our study is sample size. Although we attempt to capture as large of a sample as possible, it is still problematic to collect a larger dataset based on our filtering processes and our research questions. A larger dataset allows us to get different perspectives of the text mining results from different industries. A larger dataset also makes the classification model more reliable. 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. INVESTORS' PERCEPTIONS ON SECURITY INCIDENTS AND PROFITABLE SHORT-TERM INVESTMENT OPPORTUNITIES

### 3.1. Introduction

Information security incidents could result in a severe impact on a firm's operation and reputation (e.g., Glover et al. 2001; Warren and Hutchinson 2000). Also, these incidents could increase the volatility of a firm's business value. This increased volatility could result in an increase in the firm's cost of capital and harm the firm financially (e.g., Froot et al. 1992; Bushee and Noe 2000; Allayannis et al. 2005). There are two possible pieces of information that could cause the volatility: fundamental information and non-fundamental (i.e., noisy) information about the firm's future performance (e.g., Black 1986; Bushee and Noe 2000; Venkatachalam 2000). In the former case, the firm can focus on improving its fundamental and disclose its practices to the public in order to lower the volatility (e.g., Bushee and Noe 2000; Lang and Lundholm 1993). In the context of information security incidents, this means that if security incidents indeed affect a firm's long-term profit generating ability, the firm should focus on security policy and investments to change this fundamental and convey to the investors.

However, in the latter case, some investors trade on the noise as if the noise is information. The price of the stock now contains information about the informed

traders' information and uninformed traders' noise (e.g., Black 1986). Therefore, the stock price in the latter case could lead some investors to believe that the noisy price truly reflect the firm's future performance and make investment decisions wrongfully. In the context of information security, this means the breach does not harm the firm in the long-run. However, because of the noise, investors might perceive that there exists a permanent impact. Consequently, by understanding what leads to the reaction to security incidents provides guidance on whether firms should pay more attention to security investments and disclosures, and the association between the impact of security incidents and a firm's cost of capital. Also, the understanding of informed investors' perceptions can help general investors make better investment decisions when facing the announcements of security breaches since the existence of information asymmetry among investors (i.e., noise versus information) could demonstrate opportunities for profitable short-term investments (e.g., Black 1986).

Accordingly, this study focuses on the information asymmetry among investors and attempts to address the following questions: What are the perceptions of uninformed and informed investors on the impact of information security incidents on a firm's future performance? Does the impact of security incidents on a firm's business value mainly result from the noise or from fundamental information? More importantly, what can we learn from informed investors' perceptions? Is there a better measure that can help researchers and investors to capture the impact of security incidents on the uncertainty of a firm's future performance from the informed investors' perspective in a timely manner? Do information security incidents create short-term profitable investment opportunities due to the information asymmetry among investors?

In order to approach our research questions, we draw upon the literature in economics, finance, and accounting to understand the perceptions of informed and uninformed investors' reactions from the observed trading volume behavior after security incidents. We further investigate the informed investors' perceptions, and the association between the number of informed investors and the trading volume behavior. This association helps us verify the underlying causes to the impact of security incidents on a firm's business value. Last, we propose the use of implied volatility in the option market to better capture the impact of security incidents and the possible short-term profitable investment opportunities.

The rest of the paper is organized as follows. Literature related to this study is discussed in Section 2. In Section 3, the theoretical framework and our hypotheses are elaborated. The methodology is discussed in Section 4 while the empirical analysis and the preliminary results are shown in Section 5. We conclude with plans for future analyses and contributions in Section 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 abnormal trading volume corresponding to information announcements. The third stream of literature is about analysts' forecasts.

### 3.2.1. Information Security

Several studies focus on information security from an economic perspective, such as information security investments (Gordon and Loeb 2002; Gordon et al. 2003), and the impact of information security breaches on business operation, including physical and intangible impacts (Glover et al. 2001; Warren and Hutchinson 2000). Also, studies have investigated the impact of information security breach announcements on business value based on different methodologies and different datasets. Some of the results show that there exist significant negative impacts (Alessandro et al. 2008; Cavusoglu et al. 2004; Ettredge and Richardson 2003; Garg et al. 2003), while others do not find such impact (Campbell et al. 2003; Hovav and D'Arcy 2003; Kannan et al. 2007). Different from the literature, this paper investigates a different but more fundamental issue—investors' perceptions of security incidents and information asymmetry among investors.

### 3.2.2. Trading Volume

The discussion of trading volume can be traced back to Beaver (1968). Beaver (1968) found that earnings announcement generates not only abnormal price changes but also high trading volume. Price changes reflect the change in market's average beliefs aggregately while trading volume is the sum of all individual investors' trades (e.g., Kim and Verrecchia 1991; Bamber 1987; Bamber and Cheon 1995). The association between the inconsistent of beliefs and trading volume demonstrates that a subset of investors have the advantage in processing the information (Morse 1981; Kim and Verrecchia 1994, 1997; Bamber et al. 1997; Easley and O'Hara 1987; Hasbrouck 1988,

1991; Bhattacharya 2001). Therefore, trading volume could reflect that individual investors have different belief revisions after information announcements (Karpoff 1986; Kim and Verrecchia 1991; Bamber and Cheon 1995). In this paper, we apply this concept in the context of the announcements of information security incidents to investigate the different beliefs among investors.

### 3.2.3. Analysts' Forecasts

In order to capture the reactions of informed investors to security incidents, this study also builds upon the literature on analysts' forecasts. 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; Francis et al. 1997; Roulstone 2003). Their forecasts have been widely investigated such as how analysts formulate their expectations about firms' earnings (e.g., Kross et al. 1990; Brown 1993). In the literature, the role played by analysts in the market can be used as proxies of informed traders because of their information processing capabilities and communication with the firms (e.g., Francis et al. 2002; Roulstone 2003; Easley et al. 1998). Analysts' 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, analysts' forecasts can be a good proxy and reference point of a firm's future performance. Accordingly, in this paper, analysts are treated as

a proxy of informed investors while their forecasts are served as the reference point of the impact of security incidents on a firm's future performance.

### 3.3. Theoretical Background and Hypothesis Development

The literature on trading volume behavior is based on the argument as follows (e.g., 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). The trading volume is the sum of all individual investors' trades. Therefore, the trading volume keeps the differences between investors' reactions to announcements which are otherwise cancelled out in the aggregation process when determining prices. That is, when there is a disclosure or announcement, some investors might interpret the information as favorable information while others might consider it as unfavorable. This counterbalanced belief is averaging out in the price but is kept in the trading volume behavior after the announcement of the information. Therefore, the trading volume reflects the difference in interpretation to the announcements (i.e., belief revisions) and could demonstrate that some investors have a superior capability of processing information. In the context of information security incidents, the above concept can be applied as follows. When information security breaches are announced, the uninformed investors, according to the literature on trading volume, may not notice the announcements in a timely manner or may be unable to infer the impact of the incidents on the breached firms' future performance. As a result, they generally follow the firm's stock price reactions since the price is the aggregation of information in the market (e.g., Kim and Verrecchia 1991; Bamber and Cheon 1995). However, the informed investors

can better interpret the impact of security breaches on a firm's value from the announcements. As discussed above, when some investors are more capable of processing information than others and assessing the impact of security incidents, it is expected to observe an increase in trading volume. Accordingly, similar to prior literature (e.g., Beaver 1968; Morse 1981; Bamber 1986), we state our first hypothesis as the following.

*Hypothesis 1: Trading volume increases significantly when the firm faces breach announcements.*

As a next step, we are interested in assessing how different the beliefs of the uninformed and the informed investors? Given the uninformed investors make their investment decisions based on a firm's stock price reaction after breach announcements, as discussed above, it is expected that the uninformed investors react negatively to security incidents (e.g., Alessandro et al. 2008; Cavusoglu et al. 2004; Ettredge and Richardson 2003; Garg et al. 2003). Hypothesis 1 suggests that different belief revisions exist between the informed and the uninformed investors after security incidents while the above paragraph argues that the uninformed investors follows the price and react negatively to security incidents. As a result, Hypothesis 1 and the above argument about uninformed investors suggest that the informed investors do not react negatively to security incidents because of the following reason. If the informed traders also react negatively (no matter more negatively or less negatively than the uninformed traders), the expectations of the informed and uninformed traders about the impact of security incidents are all aimed toward the same direction (i.e., negative) which will lead to a significant negative price reaction but small trading volume. Therefore,

*Hypothesis 2: Informed investors do not react negatively to information security incidents after the announcements of breaches.*

Building on the above hypotheses, it seems that the observation of negative impacts on a firm's business value might primarily result from the uninformed traders' trading strategy. Specifically, if the majority of the investors of a firm are uninformed traders, we expect to observe a negative stock price reaction but with small trading volume. Therefore, we hypothesize that the number of informed investors is negatively associated with the trading volume after security announcements. This hypothesis serves as a verification of whether the observed negative impacts are mainly from noise. If this hypothesis is true, then firms need to take proper action to lower the information asymmetry between the firm and outside investors. Also, this hypothesis leads to the following analysis when measuring the impact of security incidents on the uncertainty of a firm's future performance from the informed investors' perspective.

*Hypothesis 3: The number of informed investors is negatively associated with the trading volume after security announcements.*

Given the above hypotheses, the next question is: is there a measure that we can use to capture the impact of information security incidents on the uncertainty of a firm's future performance from the informed investors' perspective? Therefore, we attempt to propose a less-noisy but still timely measure for the impact of security incidents on the uncertainty of a firm's future performance from the option market because of the following reason. As shown in the literature, informed investors are more likely to trade in the option market because of the relative lower transaction costs and general financial leverage (e.g., Black 1975; Mayhew, Sarin, and Shastri 1995). Furthermore, when

informed traders have private information about the volatility of a firm, the informed traders can only trade for this information in the option market (Back 1993; Cherian 1993). In the context of information security, the announcements of security breaches convey information about the uncertainty about a firm's future performance. As the informed traders can better interpret the announcements than others, this advantage in information will be reflected through the option market. Therefore, we argue that the change in implied volatility calculated based on the Black-Scholes model (Black and Scholes 1973, shown in the methodology section) better reflect the impact of security incidents on the uncertainty of a firm's future performance from the informed investors' perspective given that implied volatility can be a good forecast for volatility in different contexts (e.g., Harvey and Whaley 1992; Sheikh, 1989; Christensen and Prabhala 1998). As hypothesized (Hypothesis 2), the informed investors do not react negatively to security incidents. That is, from the informed investors' perception, the uncertainty of a firm's future performance should not increase after breach announcements. Also, given that the analysts do not react negatively, it means that the stock price will restore to the normal state from the temporary drop which generally refers to a decrease in implied volatility (e.g., Dumas et al. 1998; Black 1986). Therefore, in terms of implied volatility, we should expect to see a decrease in implied volatility after the announcement of security breaches. Accordingly, we state our fourth hypothesis:

*Hypothesis 4: Implied volatility of a firm's options decreases after the announcements of security incidents.*

From the above discussion, we notice that the informed investors, comparing to uninformed investors, are more capable of processing information and are able to

interpret the impact of breach announcement on a firm's future performance when making investment decisions. Furthermore, from Hypothesis 2 and 4, it is expected that the negative reaction of stock price after security incident is only temporary. As long as it is temporary, the stock price should restore to the normal state subsequently. This temporary drop of stock price thus provides a short-term investment opportunity. Consequently, we state our fifth hypothesis:

*Hypothesis 5: The breach announcements provide profitable short-term investment opportunities.*

### 3.4. Research Methodology

In order to test our hypotheses, we first identify information security incidents. Based on the data we collected, we investigate the uninformed and informed investors' perceptions on security incidents.

#### 3.4.1. Identify Information Security Incidents

We identify the breached firms by searching for news articles from 1997-2007 in *Factiva* database, *CNet* and *ZDNet*. The keywords used in our search include (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, which are 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 limit our search in the major news media, such as *the Wall Street Journal*, *USA Today*, *the Washington Post*, and *the New York Times*. For our

analysis purpose, we only include the news articles about publicly traded firm with specific event date after ruling out the observations with confounding events, such as earnings announcements, new product release and merger and acquisition (see Appendix C for our sample). For our analysis, we also exclude the consecutive-attack observations except the first day, such as the series of DoS attack in 2000, and the observations without trading data and analyst forecast data. The resulting sample size is 84 for the remaining analyses.

### 3.4.2. Estimate Abnormal Trading Volume

For Hypothesis 1, based on the observations we collected, we use two measures commonly used in the literature (e.g., Atiase and Bamber 1994; Bamber and Cheon 1995; Kross et al., 1994) to investigate the trading volume by controlling the market effect and the cross-sectional effect separately. The first measure controls for the market effect. In particular, we use *Eventus*<sup>®</sup> to estimate the cumulative abnormal daily trading volume percentage ( $CAV_{it}$ ) for firm  $i$  at time  $t$  through Equation (3-1).

$$V_{it} = \alpha + \beta V_{mt} + \varepsilon_{it} \quad (3-1)$$

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).  $\alpha$  and  $\beta$  are the parameters and  $\varepsilon$  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. According to Hypothesis 1, we expect to observe that the trading volume increases significantly at the announcement day.

Since the increasing trend could result from cross-sectional heterogeneity (i.e., the firm-specific effect), the second measure for Hypothesis 1 controls for this cross-sectional effect and investigates the trading volume behavior after breach announcements. In particular, we calculate the abnormal trading volume by the average trading volume of firm  $i$  after the announcement in our two-day window divided by the average trading volume of firm  $i$  30 days before the announcement. This measure allows us to examine whether the trading volume is different from the normal behavior of each firm. From Hypothesis 1, it is expected that this ratio should be significantly larger than 100%. We also use this measure when testing Hypothesis 3.

### 3.4.3. Analyze Analysts' Forecasts

As discussed, for Hypothesis 2, we use analysts' forecasts as a proxy for the reactions of informed investors to security incidents. Therefore, for each of the breached firm identified, we collect analysts' forecasts of diluted earnings per share (EPS) excluding extraordinary items on *I/B/E/S* database for the corresponding quarter before and after the incidents. In order to build the association between forecasts revision and

security incidents, we also collect the date when analysts make the forecasts from *I/B/E/S* database. Furthermore, since the actual quarterly performance can help us explain and verify the actual impact comparing to analysts' forecasts, the actual quarterly performance are also collected.

For the forecasts before the incidents, we calculate the median of analysts' forecasts made within one year before the quarter when incidents occur to get the consensus of analysts' forecasts for the breached firm. This consensus is used as the reference point for the firm's performance for that quarter *without* security incidents. For instance, Amazon.com was hit by DoS attacks in February 2000. The corresponding quarter for this attack is the quarter ended on March 30, 2000. Therefore, we collect the analysts' forecasts made between April 1999 and the attack announcement date (February 8, 2000, for example) for Amazon's performance ended on March 30, 2000. Then we calculate the median of all the forecasts to form the consensus of Amazon's performance on March 30, 2000. We choose 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 forecasts after the incidents, we search for any forecast revision immediately after the incidents. To be conservative, we search all the possible forecast revisions within two weeks after the incidents. We pick the two-week period is because the longer the time frame, the more other events could affect the forecast and cannot be associated with the incident. If there is any revision, it is attributed to the incidents after controlling for all other announcements such as merger and acquisition announcements by searching for news articles on *LexisNexis* and the firm's website. According to

Hypothesis 2, we expect that the consensus of analysts' forecasts does not negatively and significantly change after security incidents.

To test Hypothesis 3, we form the following regression model to show the association between trading volume and the number of informed investors.

$$CAV_{it} = \alpha + \beta_1 Size_{it} + \beta_2 Age_{it} + \beta_3 Price\_Reaction_{it} + \beta_4 Confidentiality_{it} + \beta_5 Integrity_{it} + \beta_6 NAnalysts_{it} + \varepsilon_{it} \quad (3-2)$$

where *Size* is the logarithm of total assets for firm *i* at the quarter when the security incident occurs, *Age* is the number of months for firm *i* being listed till breach announcements, *Price\_Reaction* is the contemporaneous price change at the time of breach announcements in the two-day window under investigation. The above three control variables are commonly used in the trading volume literature (e.g., Bamber and Cheon 1995; Kross et al. 1994). In the context of information security, we also control for incident types where *Confidentiality* (*Integrity*) is the dummy variable representing confidentiality (integrity) type incidents when equals 1. When both equal 0, it represents availability type incidents. The last variable is the number of analyst following (*NAnalysts*) for the quarter when the security incident occurs which is the proxy of the number of informed traders. Based on Hypothesis 3, it is expected that  $\beta_6$  is significantly negative.

#### 3.4.4. Implied Volatility and Profitable Short-Term Investment Opportunities

Next, as discussed, we propose the use of implied volatility from the option market to assess the impact of security incidents on the uncertainty of a firm's future performance. The implied volatility is calculated based on the Black-Scholes option

pricing model through the database *OptionMetrics* (Ivy DB Reference Manual 2006):

$$c = Se^{-qT}N(d_1) - Ke^{-rT}N(d_2) \quad (3-3)$$

$$p = Ke^{-rT}N(-d_2) - Se^{-qT}N(-d_1) \quad (3-4)$$

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  $\sigma$  is the historical volatility which equals the standard deviation of historic price change per share). In Equation (3-3) and (3-4),  $d_1$  equals  $[\ln(S/K) + (r - q + 1/2 \sigma^2)T]/\sigma\sqrt{T}$  and  $d_2$  equals  $d_1 - \sigma\sqrt{T}/2$ .

Different from the historical volatility in Equation (3-3) and (3-4), implied volatility is the volatility in the Black-Scholes model calculated based on the option price and the stock price of the firm on that day. In order to do so, we obtain all the call option and put option data for the breached firm identified from the database *OptionMetrics*. For each firm, we select the options that have the expiration date close to the end of the quarter when the incidents occur. This time period allows us to compare the results to the analysts' forecasts. Then, we calculate the average change in implied volatility after the breach announcement in the two-day window. According to Hypothesis 4, we expect to see a significant negative change in implied volatility.

Last, for Hypothesis 5, we calculate the return if we buy the stock on the breach announcement date using the closing price and sell the stock after one, two and three trading days using the closing price. We choose this one-day, two-day, and three-day

period because the longer the time period, the higher the possibility that there are other events affecting the stock price. According to Hypothesis 5, we should expect to see a positive return based on this trading strategy.

### 3.5. Preliminary Empirical Results

We first investigate the price change and trading volume change on the day of breach announcement. For price change, we do not observe a significant negative stock price reaction to breach announcements in the two-day window which confirms our belief that the market participants could have different perceptions on such announcements. For trading volume change, we first consider the measure that controls for the market effect in Equation (3-1). We plot how trading volume changes across time after controlling for the market effect in Figure 3.1. The peak at day 0 (significant at 10% level) demonstrates that the breach announcements indeed induce more trading volume.

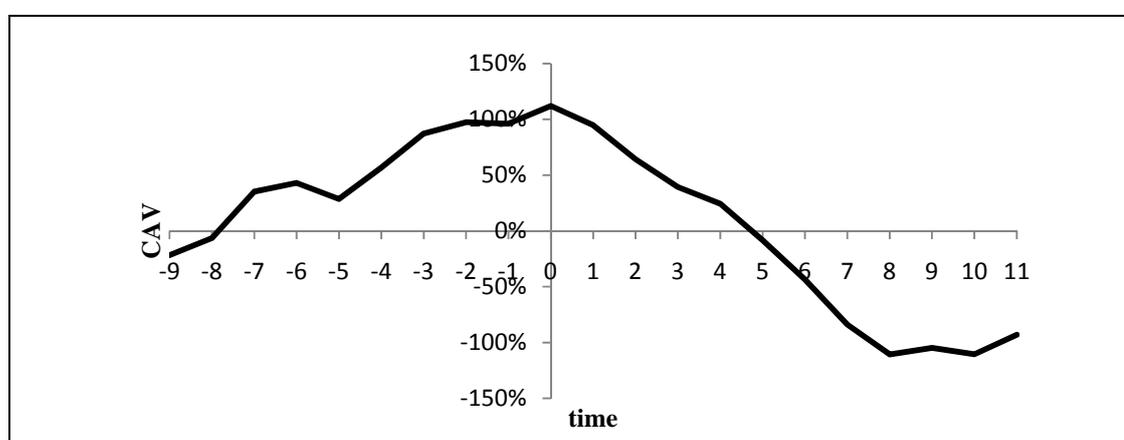


Figure 3.1 Trading Volume Change across Time

However, this increasing could also result from the firm-specific effect. Therefore, we next use the second measure to examine whether the trading volume is larger than

average for firm  $i$  after breach announcements. The results show that, on average, the trading volume is 13.62% more than usual after breach announcements and significant at 10% level. The above results support our first hypothesis that investors have different beliefs of security incidents.

Next, as discussed, analysts' forecasts are used as the proxy for informed traders' reactions. About 33% of our sample can be matched to some analysts' forecasts revision after the breach announcement. However, interestingly, none of these forecast revisions can be associated directly to security incidents. The finding suggests that the informed investors might not perceive that information security breaches will affect a firm's future performance. This finding explains why we do not observe a significant negative stock price reaction to security incidents because not all the investors react negatively. This observation is further verified when we compare the breached firm's subsequent actual quarterly performance with the analysts' forecasts. The comparing results demonstrate that, without other future events, all the firms' performance is greater or equal to the analysts' forecasts. That is, other things being equal, security incidents are not believed to affect the breached firms' future performance. The finding leads us to believe that, 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. This finding supports the second hypothesis that informed investors do not react negatively to information security incidents after the announcements of breaches.

The third hypothesis is tested using Equation (3-2). The results for Equation (3-2) are given in Table 3.1. Similar to the literature (e.g., Bamber and Cheon 1995; Kross et

al. 1994), the contemporaneous stock price change can affect the level of trading volume. Also, as expected in Hypothesis 3, the number of informed traders can negatively affect trading volume. This result suggests that if the breached firm faces a majority of noisy traders, there will be a negative stock price reaction after security incidents but this is not the case for the firms with more informed investors.

Table 3.1 Results for Equation (3-2)

Variables	Coefficient
Intercept	<b>1.03</b> <sup>***</sup>
<i>Size</i>	0.01
<i>Age</i>	0.00
<i>Price Reaction</i>	<b>2.46</b> <sup>**</sup>
<i>Confidentiality</i>	-0.05
<i>Integrity</i>	-0.02
<i>NAnalysts</i>	<b>-0.01</b> <sup>**</sup>
* significant at 10% level, ** significant at 5% level, *** significant at 1% level	

This argument is further verified by examining the stock price reaction only for the firms with fewer informed investors. This is defined as the observations corresponding to the smallest half of the number of analyst following. Specifically, the data is sorted by the number of analyst following and we select half of the observations from the smallest then investigate the stock price reaction. Interestingly, the stock price reaction in the same two-day window now becomes significantly negative at 10% level. Therefore, the impact of security incidents on business value does not fully depend on the breach announcement but instead depend on how the investors interpret the announcements. If the investors are not capable of incorporating the news article into their decision information set, the breached firm could still be harmed in the short-run

even the incident does not affect the firm's future profitability. Accordingly, firms can reduce this unnecessary reaction by attracting more informed investors and lower the information asymmetry between the firm and outside investors through providing a more transparent information environment.

### 3.6. Conclusion

This study is still in progress. The preliminary results suggest the existence of different beliefs among investors. The uninformed investors react negatively to security incidents but the informed investors appear to treat security incidents as part of the overall business operation risks, i.e., they do not believe that the incidents affect the firm's future performance. This perception difference is consistent with various observations. First, we observe that the breached firms' subsequent actual quarterly performance is not affected by the incidents and is in sync with analysts' forecasts. Second, we notice a significant increase in the trading volume of the firm's stock on the day of its breach. Last, interestingly, the composition of a firm's investors alters the impact of security incidents from nearly zero to negative as the portion of noisy trader increases.

Based on the preliminary results and our Hypothesis 4 and 5, this study will explore the use of implied volatility to measure the change of the expectation about the uncertainty of the breached firm's future performance. We then compare this result with that for Hypothesis 2 and verify whether these two results both reflect the perception of informed investors. Also, by comparing the investment decision made based on CAR and on implied volatility, we could propose a new measure that better reflect the impact

of security incidents from the informed investors' perspective. Last, we test Hypothesis 5 by forming actual investment strategy and show that there exists profitable (on average) short-term investment opportunity after the announcement of information security incidents.

This study adds to the literature of information security by investigating a more fundamental problem—investors' perceptions which is the key element when understanding and estimating the impact of security incidents on a firm's business value. Furthermore, this study has implications for managers and investors. For managers, our results suggest that allocating more resources to information security investment is not an effective way to lower the temporary 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 lower the noise in the market and lower the temporary impact of 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 sources other than the stock price itself which could also result in profitable investment decisions.

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

### 4.1. Introduction

Identity theft refers to a situation in which someone wrongfully obtains and uses another person's personal data in some way that involves fraud or deception (Office of Justice Programs 2004). About 3.2 million households were victims of identity theft and 30% of them suffered a loss between \$500 and \$2,499 (Office of Justice Programs 2004). In the online world, identity theft has become a more serious issue given it is much more difficult to prove you are the one you claim to be. This problem may not only hinder the development of e-commerce but also increase concerns when retrieving or exchanging highly confidential information, such as personal medical history and electronic health records (EHR).

As the concerns about identity theft have increased its popularity, people start to argue whether the current authentication system can effectively distinguish imposters from genuine users. For example, Federal Financial Institutions Examination Council (FFIEC) released guidance on authentication in Internet banking environment on October 12, 2005 (FFIEC 2005). This guidance asks all the regulated agencies, by the end of 2006, to conduct risk-based assessments and develop security measures to reliably authenticate customers remotely accessing their online financial services, which may be

two-factor or multi-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 later) for any online service or product provider. Although it seems to be more secure, however, multi-factor authentication could also be accompanied by concerns about the use of additional information collected by the firm. Also, it might need additional implementation costs, such as software, hardware, and training (Wildstrom 2005). For customers, the new interfaces, new devices, and longer authentication processes could result in incontinence of the authentication process and prolong the time needed to complete the transaction. All the above issues could at the same time adversely lower the customers' acceptance level of the two-factor authentication system provided by a certain service or product provider. Accordingly, the new authentication system could be more preferable depending on several inter-related factors. However, the relationship between these inter-related factors as well as the impact of these factors on a firm's decision of adopting a new authentication system are not clear.

Therefore, this paper attempts to address the following questions by using a static method as a first attempt to understand the decision of choosing authentication systems. 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 systems? What are the conditions that make the new authentication system more preferable? This study first generalizes all kinds of authentication systems into two broad types. Based on the definition of system failure under these two broad types of

authentication systems, we are able to compare the conditions that make the new authentication system more preferable. The 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 2. In Section 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. We conclude with contribution, and possible avenues for future research in Section 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

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). It has long been discussed from the technical perspective. For instance, Woo and Lam (1992) and Diffie 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; Gong 1992; Aboba et al. 2004) or ways to implement or improve authentication methods (e.g. Beng et al. 2004; Sutcu et al. 2005; Bhargav-Spantzel et al. 2006). However, studies about authentication from the economic perspective are often embedded in the discussion of other issues. For example, Anderson (2001) discussed the role of authentication in

information security from an economic perspective while authentication has also been discussed in internal control and EDP auditing literature (Webber 1997). This study, thus, adds to the literature and focuses on the decision of authentication systems from an economic perspective.

In this paper, we focus on identity authentication, i.e., the process of verifying a person's identity. In general, the information (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.

Different from factor 2 and factor 3 above, biometric authentication system measures an individual's physical features based on the data stored, and then determine the identity of the user. Biometric systems use "scores" to show the similarity between a pattern and a biometric template (BioID.com 2004; Braghin 2001; Bromba biometric 2006; Ross et al. 2006; Jain et al. 2004). 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 threshold pre-determined by the user, 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% (FindBiometric.com 2006; 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%

(FindBiometric.com 2006; Braghin 2001; Yun 2002; Panko 2003; Jain et al. 2004). Under the current state of technology solutions, different biometric traits have different accuracy rates and implementation costs given. 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 (Bromba biometric 2006; Panko 2003; Jain et al. 2004). This study formally models the probability of system failure for the system using the information someone has and someone knows, and build on the biometric literature to calculate the probability of system failure for biometric authentication systems. Specifically, this study generalizes the authentication systems into two broad categories based on the calculation of the probability of system failure.

To implement the authentication system, it is necessary to obtain users' personal identifiable information, such as names, addresses, and even purchasing history of an identifiable individual (Nowak and Phelps 1995). 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., Perrig et al. 2004; Bhargav-Spantzel et al. 2006; Dhamija and Tygar 2005; Camenisch and Lysyanskaya 2001; Davida et al. 1999). 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 once the system fails. The above two impacts in opposite direction could affect a firm's decision on implementing a new authentication system.

#### 4.2.2. Privacy

This study, thus, also relates to, though not directly, the literature on privacy from the economic perspective. Privacy is defined as the individual's ability to control the collection and use of personal information (Stigler 1980; Westin 1967; 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), how business use personal information for promotions and cross market information (e.g., Hann et al. 2005; Akçura and Srinivasan 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 (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 loss because of the system failure. Also, the 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. Last, 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. This provider currently has a market share of  $m$  in the service or product category it provides, where  $0 < m < 1$  (see Appendix F for variable definitions).  $m$  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 customer is required to provide a certain level ( $\alpha$ ,  $0 < \alpha \leq 1$ ) of personal information, such as name, address, and phone number. Once the system fails (defined later), the product or service provider might need to compensate 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 (Tang et al. 2008).

The customers are categorized along two dimensions: privacy and convenience. The first dimension is about privacy sensitivity. In the market the provider faces, a proportion of customers ( $\rho$ ,  $0 \leq \rho \leq 1$ ) are privacy sensitive. This portion of customers has more concerns about the information collected from them. Therefore, after the provider shifts to another authentication system or has been breached, some of these customers might choose to purchase the service or product from other providers because of the privacy concerns. The other dimension is about convenience. A proportion of customers ( $\delta$ ,  $0 \leq \delta \leq 1$ ) emphasizes more on the convenience of the transaction. After

the provider switches to another authentication system, a certain portion of these customers might not keep purchasing from this provider because the possible inconvenience caused by the new system. This categorization is illustrated through Figure 4.1.

Privacy Sensitivity	High	$\rho(1-\delta)$	$\rho\delta$
	Low	$(1-\rho)(1-\delta)$	$(1-\rho)\delta$
		Low	High
Convenience Sensitivity			

Figure 4.1 Types of Customers

In this paper, system failure is defined as any situation in which non-genuine users being able to access to the information or genuine users being unable to access to the information because of the failure of the software or hardware, compatibility issue of the software or hardware, or the successful action of the hackers. Based on the definition, we discuss the probability of system failure for different authentication systems.

#### 4.3.2. Probability of System Failure

As discussed in the literature review, there are three types of information people used for authentication systems. Since how biometric authentication system works is differently than others, we categorize all the authentication systems into two general types. The first type uses information someone has or someone knows. The other type uses biometric information. When the information used for authentication is the information someone knows or someone has, the authentication system can be seen as a

non-repairable system with one component. The reason is that the longer the time we use a system, the larger the probability 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  $\lambda$  is the mean-time-to-failure and  $b$  is the change of failure rate. The subscript  $n$  denotes one non-repairable component. From our discussion about the relationship between time and failure probability, it is expected that the change of failure rate increases with time. Therefore, we assume  $b$  is larger than 2 for the remainder of our analysis.

However, this probability only accounts for half of the probability of system failure. Specifically, when a hacker enters the correct password, the system should grant access and the system functions correctly. Therefore, we also need to take the hackers successful action into account. Also, 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. Therefore, the successful rate of the hackers' actions under different authentication methods should also be an increasing function of time and denote as  $H(t)$ . Based on our definition of system failure, the 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  $H(t)$ , i.e.,  $(1 - e^{-(t/\lambda)^b}) + H(t) - (1 - e^{-(t/\lambda)^b})H(t)$ .

Similarly, if there are two independent non-repairable components, based on our definition of system failure, the CDF of system failure across time  $t$  (denote as  $F_m(t)$ ) is

assessed by both  $1 - e^{-(t/\lambda_1)^{b_1} - (t/\lambda_2)^{b_2}}$  and  $H(t)$ . Again, the subscript  $nn$  represents two non-repairable components. The two components could also be dependent. However, our main proposition in the following section remains similar with two dependent components. Therefore, in the following analysis, we only discuss the case when the two components are independent.

The other information can be used for authentication systems is biometric information. From the literature, in the biometric system, there is always a probability of false acceptance (FAR,  $\psi$ ) and false rejection (FRR,  $\varphi$ ) 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 weight that matches its needs which is out of the scope of this study. Once the characteristics are determined (e.g., threshold, FAR, FRR), the probability of system failure given the pre-determined threshold ( $\bar{s}$ ) across time  $t$  (denote as  $F_{bio}(t; \bar{s})$ ) is calculated by both  $1 - (1 - w_{FRR}\varphi - w_{FAR}\psi)^t$  and  $H(t)$ , where  $w_{FRR}$  and  $w_{FAR}$  are the weights pre-determined by the provider at the time when it selects the system. Again, the subscript  $bio$  represents the biometric system.

Similarly, if the provider selects an authentication system that uses both biometric and non-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  $H(t)$ . Here, we do not have any assumptions

regarding the mean-time-to-failure, the threshold, the weights, FAR, and FRR. All these parameters could vary based on the authentication system the provider chooses.

#### 4.3.3. Analysis

We start our analysis with the base case: one non-repairable component authentication systems. 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 expressed as the addition of the implementation costs ( $c$ ), the change in customer base when system fails as defined earlier, and the expected losses. 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 ( $0 \leq \varepsilon_l \leq 1$ ) of  $\rho$  (see Appendix F for definition of  $\varepsilon_l$ ). The expected loss is the value the provider needs to compensate its customers and settle possible lawsuits and penalty ( $L$ ) once the system fails. Formally,

$$C_n = c_n + F_n(t)(V_n + L_n) \quad (4-1)$$

where 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 one non-repairable component authentication system, the associated expected costs and losses consist of four components. The first component is still the implementation costs. The

second component reflects the net change of the customer base when the provider shifts to the new system. 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 methods. The loss of existing customers equals the current market share ( $m$ ) times a certain percentage ( $0 \leq \varepsilon_2 \leq 1$ ) of  $\delta$  and the benefit of attracting new customers equals the potential customer ( $1 - m$ ) times a certain percentage ( $0 \leq \varepsilon_3 \leq 1$ ) of  $\rho$ . The last term is still the loss of customers and the expected losses once the system fails similar to the base case. Accordingly,

$$C_{bio} = c_{bio} + V_{net\_bio} + F_{bio}(t; \bar{s})(V_{bio} + L_{bio}) \quad (4-2)$$

where the subscript *bio* represents the biometric system and 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.

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 authentication system, the associated expected costs and losses still consists of four major components which are given in Equation (4-3) and Equation (4-4) respectively.

$$C_{nn} = c_{nn} + V_{net\_nn} + F_{nn}(t)(V_{nn} + L_{nn}) \quad (4-3)$$

$$C_{nbio} = c_{nbio} + V_{net\_nbio} + F_{nbio}(t; \bar{s})(V_{nbio} + L_{nbio}) \quad (4-4)$$

where the subscript *nn* (*nbio*) represents the two non-repairable component authentication system (the combination of one non-repairable component and one biometric

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.

By subtracting Equation (4-1) from Equation (4-2), (4-3), and (4-4), we determine the factors and the conditions that make the shifting worthwhile as shown in Panel A through Panel C in Appendix G. 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 compare one-factor with another one-factor system and to compare two-factor with another two-factor authentication system.

On the one hand, the results given in Appendix G 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. On the other hand, we also compare two different types of two-factor authentication systems. In particular, we subtract Equation (4-4) from Equation (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 G Panel D. These conditions are discussed in the next section.

#### 4.4. Managerial Implications

From the conditions given in Appendix G, the conditions that could make the new authentication system more preferable than the base case are essentially similar and can be boiled down to the factors stated in Proposition 1.

*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 implementation costs, (2) the net change of the value of its customers including the loss of customers after system failure which is determined by the percentage of privacy sensitive customers ( $\rho$ ), the percentage of convenience sensitive customers ( $\delta$ ), and the current market share or market value of customers ( $m$ ), and (3) the expected losses ( $F(t)L$ ).*

From Appendix G and proposition 1, there are several points worth noting. First, the condition for the implementation costs shows that the additional implementation costs of the new system compared to the base case have to be smaller than a certain threshold in order to make the new system more preferable. This is similar when we compare two two-factor authentication systems. The threshold reflects the following conditions. Although the probability of system failure could be smaller for the new system based on the system the provider chooses 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 needs 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 reduced the losses at the same time, the threshold of the implementation costs can be higher which still make the new system more preferable.

Second, in order to make the new system more preferable compared to the base case, the percentage of privacy sensitive customers in the market the provider faces should not be too low or too high. If the percentage of privacy sensitive customers is too low, the costs and expected losses cannot be justified by the improving of 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 their provided information is used properly, it seems that an authentication with higher security level should fit better with the customers' preference. One possible explanation of the results is 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 might lose more customers once the system fails. The loss of more customers could result from the loss of reputation and customers' expectations.

However, different from case when we compare two one-factor authentication systems, the conditions in Appendix G Panel D says that the majority of the customer base should be privacy sensitive or non privacy sensitive in order to make the two non-repairable component system more preferable. 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,

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. Therefore, we state our second proposition.

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

Third, the condition for the percentage of convenience sensitive customers suggests the following. This condition exists 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 smaller than those for the base case. That is, if privacy 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. Otherwise, the new system is not preferable to the base case.

*Proposition 3: If the service or product provider operates in the market where privacy 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 inconvenience when deciding shifting to the new authentication system*

Proposition 3 suggests that if the provider sells services or products involving confidential information, it should focus on the system that can lower the privacy concerns before worrying about the impact of inconvenience. If the privacy concerns

cannot be lowered, the new system is not preferable and there is no need to consider the inconvenience factor.

Fourth, the current market share of the provider must be large enough for the new authentication system to be more preferable. The threshold for the market share increases as the additional implementation costs increase. The market share (or the value of the 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. 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 which seems to be contradicted with our result. 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. 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. For example, when financial institutions adopt new authentication systems in response to FFEIC, 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.*

Last, the expected losses resulting from the new authentication system should not exceed the threshold in order to make the new authentication system more preferable. 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 lower the penalty and the compensation to customers associated with the new system once the new system fails comparing to the original system. 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.

*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).*

The above propositions also lead 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 since customer type and the change in customer base are important factors when determining authentication systems. Therefore, for different group of customers, the provider can implement different authentication systems in order to fit the preference of different group of customers.

#### 4.5. Conclusions

By comparing the expected costs and losses of different authentication methods, we show the key factors and several insights online service or product providers need to consider when shifting to a new authentication system. The factors are (1) the additional implementation costs, (2) the net change in customer value, and (3) the expected losses. The net change in customer value is determined by the market share and the composition of customers. A service or product provider needs to select the authentication system based on the current state of market share and the customers' preferences. We show that small market share providers can follow the same strategy adopted by the large share provider in order to lower the impact of the switch in customer especially when the shift is mandatory. Also, we demonstrate that government can encourage the shift by adjusting the penalty a firm faces once the system fails.

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. This study demonstrates that all kinds of authentication systems can be modeled into two broad categories: non-repairable and biometric. This categorization can be used for future studies about authentication systems. Also, this study provides suggestions to managers when considering shifting to a new authentication system. All the elements discussed in the study 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 system 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 consider two or more biometric components combined with each other. Third, 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 proposal investigates three different issues in information security: information security related disclosures, investors' perceptions on information security breaches, and two-factor authentication systems.

The first essay provides a comprehensive analysis to quantitatively and qualitatively investigate the association between security disclosures and the market reactions to security breaches. The results of the cross-sectional analysis demonstrate that the investors perceive these security risk factors disclosed in financial reports as warnings to future incidents and punish the firm once the firm faces security incidents. In order to provide insights about how firms should disclose information security related risk factors to the public, we explore the contents of the disclosures using text mining techniques. We first build a classification model to link disclosure patterns with breach announcements. The model shows that a certain disclosure pattern is more likely to be associated with subsequent breach announcements and to be perceived as warning to future incidents. After exploring the disclosure patterns, the cluster analysis shows that disclosures with action oriented terms are less likely to be inferred as warning to future incidents.

The second essay investigates the investors' perceptions on security breaches. The preliminary results demonstrate that there exist different beliefs about the impact among

informed and uninformed investors. Informed investors believe that the security incident is part of the risk a firm must face in daily operations and do not react negatively. However, the uninformed investors solely follow the price and make their investment decisions from a negative reaction perspective. This on-going study will further propose a measure that helps managers and investors capture informed investors' perceptions on the uncertainty of a firm's future performance. Furthermore, because of the information asymmetry among investors, this study will demonstrate one short-term profitable investment strategy.

The third essay focuses on the decision of choosing authentication systems. By 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 loss. The net benefit of customer switch needs to take into account the current market share and the customers' preferences. 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

- Aboba, B., Blunk, L., Vollbrecht, J., Carlson, J., and Levkowitz, H. 2004. "Extensible authentication protocol (EPA)," *The Internet Engineering Task Force-Request for Comments*.
- 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.
- Allayannis, G., Rountree, B., and Weston, J. P. 2005. "Earnings volatility, cash flow volatility, and firm value," Working Paper, University of Virginia.
- Alessandro, A., Friedman, A., and Telang, R. 2008. "Is there a cost to privacy breaches? An event study," Working Paper, Carnegie Mellon University.
- 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.
- Back, K. 1993. "Asymmetric information and options," *Review of Financial Studies* (6), pp. 435-472.
- 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.

- Balakrishnan, K., Ghose, A., and Ipeiritis, P. 2008. "The impact of information disclosure on stock market returns: the Sarbanes-Oxley Act and the role of media as an information," Working Paper, New York University.
- Bagnoli, M., and Watts, S. G. 2007. "Financial reporting and supplemental voluntary disclosures," *Journal of Accounting Research* (45:5), pp. 885-913.
- Bagnoli, M., Kross, W., and Watts, S. G. 2002. "The information in management's expected earnings report date: a day late, a penny short," *Journal of Accounting Research* (40:5), pp. 1275-1296.
- Bamber, L. 1986. "The information content of annual earnings releases: a trading volume approach," *Journal of Accounting Research* (24), pp. 40-56.
- 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.
- Beaver, W. 1968. "The information content of annual earnings announcements," *Journal of Accounting Research* (6), pp. 67-92.
- Begley, J., and Fischer, P. 1998. "Is there information in an earnings announcement delay?" *Review of Accounting Studies* (3), pp. 347-363.
- Beneish, M. D. 2001. "Earnings management: A perspective," *Managerial Finance* (27:12), pp. 3-17.
- Bhargav-Spantzel, A., Squicciarini, A., and Bertino, E. 2006. "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. 2006. "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](http://www.bioid.com/sdk/docs/About_EER.htm).
- Black, F., 1975. "Fact and fantasy in use of options," *Financial Analysts Journal* (31), pp. 36-41.
- Black, F. 1986. "Noise," *The Journal of Finance* (41:3), pp. 529-543.
- Black, F., and Scholes, M. S. 1973. "The pricing of options and corporate liabilities," *Journal of Political Economy* (81:3), pp. 637-654.
- 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.itol.org>.
- Brandăn, 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.
- 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.
- Bushee, B. J., and Noe, C. F. 2000. "Corporate disclosure practices, institutional investors, and stock return volatility," *Journal of Accounting Research* (38), pp. 171-202.
- 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.

- 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.
- CERT. 2007. *CERT/CC Statistics 1988-2006*. Retrieved Apr. 9 2007, from [http://www.cert.org/stats/cert\\_stats.html](http://www.cert.org/stats/cert_stats.html).
- Chen, Y., and Iyer, G. 2002. "Consumer addressability and customized pricing," *Marketing Science* (21:2), pp. 197-208.
- Cherian, J. 1993. *Option pricing, self-fulfilling prophecies, implied volatilities, and strategic interaction*. Unpublished Ph.D. dissertation, Cornell University.
- Christensen, B. J., and Prabhala, N. R. 1998. "The relation between implied and realized volatility," *Journal of Financial Economics* (50), pp. 125-150.
- CSI/FBI. 2007. *The CSI/FBI computer crime and security report in 2006*, Retrieved Apr. 9 2007, from <http://abovesecurity.com/doc/CommuniquesPDF/FBISurvey2006>.
- Darrough, M. N. 1993. "Disclosure policy and competition Cournot vs. Bertrand," *The Accounting Review* (68:3), pp. 534-561.
- 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.

- Dumas, B., Fleming, J., and Whaley, R. E. 1998. "Implied volatility functions: empirical tests," *The Journal of Finance* (53:6), pp. 2059-2106.
- 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.
- Easley, D., O'Hara, M., and Paperman, J. 1998. "Financial analysts and information based trade," *Journal of Financial Markets* (1:2), pp. 175-201.
- Eihorn, E. 2005. "The nature of the interaction between mandatory and voluntary disclosures," *Journal of Accounting Research* (43:4), pp. 593-621.
- 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. 1970. "The behavior of stock market prices," *The Journal of Finance* (25), pp. 383-417.
- Fama, E., and French, K. 1992. "The cross-section of expected stock returns," *The Journal of Finance* (47:2), pp. 427-465.
- 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>.
- FindBiometrics.com. 2006. *Convenience vs security: how well do biometrics work*. Retrieved July 8, 2006, from <http://www.findbiometrics.com/Pages/feature%20articles/convenience.html>.
- 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.
- Froot, K., Scharfstein, D., and Stein, J. 1993. "Risk management: coordinating corporate investment and financing policies," *The Journal of Finance* (48), pp. 1624-1658.
- 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.
- 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 Richardson, R. 2005. *10<sup>th</sup> annual CSI/FBI computer crime and security survey*. Computer Security Institute, pp. 1-26.
- 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.
- 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.
- 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.
- 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.
- Jo, H., and Kim, Y. 2007. "Disclosure frequency and earnings management," *Journal of Financial Economics* (84:2), pp. 561-590.
- 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.
- Kaszniak, 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.

- Kasznik, R., and McNichols, M. F. 2002. "Does meeting earnings expectations matter? Evidence from analyst forecast revisions and share prices," *Journal of Accounting Research* (40:3), pp. 727-759.
- Katz, S.B. 2001. "Language and persuasion in biotechnology communication with the public: How not to say what you're not going to say and not say it," *AgBioForum* (4:2), pp. 93-97.
- 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.
- Kohavi, R. 1995. "A study of cross-validation and bootstrap for accuracy estimation and model selection," *Proceedings of the 14<sup>th</sup> International Joint Conference on Artificial Intelligence*, Montréal, Québec, Canada, pp. 781-787.
- Kross, W., Ha, G., and Heflin, F. 1994. "A test of risk clientele effects via an examination of trading volume response to earnings announcements," *Journal of Accounting and Economics* (18), pp. 67-87.
- Kross, W., Ro, B., and Schroeder, D. 1990. "Earnings expectations: The analysts information advantage," *The Accounting Review* (65), pp. 461-476.
- 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.
- Lang, M. H., and Lundholm, R. J. 2000. "Voluntary disclosure and equity offerings: reducing information asymmetry or hyping the stock?" *Contemporary Accounting Research* (17:4), pp. 623-662.

- 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.
- Lev, B., and Pennman, S. H. 1990. "Voluntary forecast disclosure, nondisclosure, and stock prices," *Journal of Accounting Research* (28:1), pp. 49-76.
- Li, F. 2006. "Annual report readability, current earnings, and earnings persistence," Working Paper, University of Michigan.
- Liebl, A. 1993. "Authentication in distributed systems: a bibliography," *ACM SIGOPS Operating Systems Review* (27:4), pp. 31-41.
- MacKinlay, A. C. 1997. "Event studies in economics and finance," *Journal of Economics Literature* (35:1), pp. 13-39.
- Masand, G., Linoff, G., and Waltz, D. 1992. "Classifying news stories using memory based reasoning," *Proceedings of the 15<sup>th</sup> 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.
- Mayhew, S., Sarin, A., and Shastri, K. 1995. "The allocation of informed trading across related markets: an analysis of the impact of changes in equity-option margin requirements," *The Journal of Finance* (55), pp. 1635-1654.
- 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.
- 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.
- Office of Justice Programs. 2004. "Identity theft," U.S. Department of Justice.

- 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.
- PriceWaterhouseCoopers. 2002. *Information Security Breaches Survey 2002 – A Technical Report*. Prepared by PriceWaterhouseCoopers for the Department of Trade and Industry.
- 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: Springer.
- Ross, A. A., Nandakumar, K., and Jain, A. K. 2006. *Handbook of multibiometrics*. NY: Springer.
- Roulstone, D. T. 2003. "Analyst following and market liquidity," *Contemporary Accounting Research* (20:3), pp.551-578.
- 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](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.
- 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.

- Skinner, D. J. 1994. "Why firms voluntarily disclose bad news," *Journal of Accounting Research* (32:1), pp. 38-60.
- 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.
- Stigler, G. J. 1980. "An introduction to privacy in economics and politics," *Journal of Legal Studies* (9:4), pp. 623-644.
- Stocken, P. 2000. "Credibility of voluntary disclosure," *RAND Journal of Economics* (31:2), pp. 359-374.
- Sutcu, Y., Sencar, H. T., and Memon, N. 2005. "Authenticaiton/protocols: a secure biometric authentication scheme based on robust hashing," *Proceedings of the 7<sup>th</sup> 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.
- 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.
- Thoma, J., and Segal, A. 2006. "Identity theft: the new way to rob a bank," *CNN.com* (May).
- Varian, H. R. 1985. "Price discrimination and social welfare," *American Economic Review* (75:4), pp. 870-875.
- Venkatachalam, M. 2000. "Discussion of corporate disclosure practices, institutional investors, and stock return volatility," *Journal of Accounting Research* (38), pp. 203-207.
- 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 11<sup>th</sup> 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.
- 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 2<sup>nd</sup> 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.
- 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 the Disclosures of Internal Control and Procedures

### “Evaluation of Disclosure Controls and Procedures

The Company’s management, with the participation of the Company’s principal executive officer and principal financial officer, has evaluated the effectiveness of the Company’s disclosure controls and procedures (as such term is defined in Rules 13a-15(e) and 15d-15(e) under the Securities Exchange Act of 1934, as amended (the “Exchange Act”) as of the end of the period covered by this report. Based on such evaluation, the Company’s principal executive officer and principal financial officer have concluded that, as of the end of such period, the Company’s disclosure controls and procedures are effective in recording, processing, summarizing and reporting, on a timely basis, information required to be disclosed by the Company in the reports that it files or submits under the Exchange Act.

### Management’s Report on Internal Control Over Financial Reporting

The Company’s management is responsible for establishing and maintaining adequate internal control over financial reporting as defined in Rules 13a-15(f) and 15d-15(f) under the Exchange Act. Under the supervision and with the participation of the Company’s management, including its principal executive officer and principal financial officer, the Company conducted an evaluation of the effectiveness of its internal control over financial reporting based on criteria established in the framework in Internal Control—Integrated Framework issued by the Committee of Sponsoring Organizations of the Treadway Commission. Based on this evaluation, the Company’s management concluded that its internal control over financial reporting was effective as of December 31, 2005.

Because of its inherent limitations, internal control over financial reporting may not prevent or detect misstatements. Also, projections of any evaluation of effectiveness to future periods are subject to the risks that controls may become inadequate because of changes in conditions, or that the degree of compliance with the policies or procedures may deteriorate.

The Company's independent registered public accounting firm has audited management's assessment of the effectiveness of the Company's internal control over financial reporting as of December 31, 2005 as stated in their report which appears on page 58.

#### Changes in Internal Control Over Financial Reporting

There have not been any changes in the Company's internal control over financial reporting (as such term is defined in Rules 13a-15(f) and 15d-15(f) under the Exchange Act) during the most recent fiscal quarter that have materially affected, or are reasonably likely to materially affect, the Company's internal control over financial reporting.”

Excerpt from Yahoo's annual report for year 2005, retrieved on Apr.23, 2007

Source: [http://www.sec.gov/Archives/edgar/data/1011006/000110465906014033/a06-3183\\_110k.htm](http://www.sec.gov/Archives/edgar/data/1011006/000110465906014033/a06-3183_110k.htm)

## Appendix B. Examples of Risk Factors

### “We Face Intense Competition

The e-commerce market segments in which we compete are relatively new, rapidly evolving and intensely competitive. In addition, the market segments in which we participate are intensely competitive and we have many competitors in different industries, including the Internet and retail industries.

Many of our current and potential competitors have longer operating histories, larger customer bases, greater brand recognition and significantly greater financial, marketing and other resources than we have. They may be able to secure merchandise from vendors on more favorable terms and may be able to adopt more aggressive pricing or inventory policies. They also may be able to devote more resources to technology development and marketing than us.

As these e-commerce market segments continue to grow, other companies may enter into business combinations or alliances that strengthen their competitive positions. We also expect that competition in the e-commerce market segments will intensify. As various Internet market segments obtain large, loyal customer bases, participants in those segments may use their market power to expand into the markets in which we operate. In addition, new and expanded Web technologies may increase the competitive pressures on online retailers. The nature of the Internet as an electronic marketplace facilitates competitive entry and comparison shopping and renders it inherently more competitive than conventional retailing formats. This increased competition may reduce our operating profits, or diminish our market segment share.”

“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 C. Sample

Company	Control Company	Event Date	Type of Incident
Aastrom Bioscience	Baxter International	2000/2/18	phonyinfo <sup>i</sup>
About.com		2000/2/10	siteattack <sup>a</sup>
Akamai tech	Blue Coat System	2004/6/16	siteattack <sup>a</sup>
Amazon.com	Barnes and Noble	2000/2/8	DoS <sup>a</sup>
Amazon.com	Barnes and Noble	2000/2/9	DoS <sup>a</sup>
Amazon.com	Barnes and Noble	2000/2/10	DoS <sup>a</sup>
American express	Capital One	2003/2/19	hackinfo <sup>c</sup>
AOL	EarthLink	2000/6/18	break in <sup>c</sup>
AOL Times Warner	Walt Disney	2002/1/3	holediscovery <sup>c</sup>
AOL Times Warner	Walt Disney	2006/8/22	breach <sup>c</sup>
AT&T	Sprint	1999/6/10	worms <sup>i</sup>
AT&T	Sprint	2006/8/24	onlinetheft <sup>c</sup>
AT&T	Sprint	2006/8/30	onlinetheft <sup>c</sup>
Bank of America	US Bancorp	1999/11/30	virus <sup>i</sup>
Bank of America	US Bancorp	2003/2/6	worms <sup>ia</sup>
Bank of America	US Bancorp	2005/2/28	datalost <sup>c</sup>
Bank of America	US Bancorp	2006/3/13	breach <sup>c</sup>
Boeing	Northrop Grumman	1999/6/10	worms <sup>i</sup>
Boeing	Northrop Grumman	2003/1/28	worms <sup>a</sup>
ChoicePoint	ISCO International	2005/2/17	ID theft <sup>c</sup>
ChoicePoint	ISCO International	2005/2/22	ID theft <sup>c</sup>
ChoicePoint	ISCO International	2005/3/5	ID theft <sup>c</sup>
Cisco	Avaya	2004/5/18	codetheft <sup>c</sup>
Cisco	Avaya	2005/5/10	codetheft <sup>c</sup>
Citigroup	JPMorgan Chase	2006/3/8	breach <sup>c</sup>
Citigroup	JPMorgan Chase	2006/3/13	breach <sup>c</sup>
Coca Cola	Pepsi	1997/9/15	attack <sup>a</sup>
Compaq	Gateway	1999/3/30	virus <sup>i</sup>
Compaq	Gateway	2001/2/15	attack <sup>a</sup>
Continental Airlines	AMR	2003/2/6	worms <sup>i</sup>
Countrywide Financial	Fannie Mae	2003/1/28	attack <sup>a</sup>
Cox Communications		2001/8/8	virus <sup>i</sup>
Critical Path	Sun Micro	1999/9/22	breach <sup>c</sup>
CSX	Norfolk Southern	2003/8/21	virus <sup>i</sup>
Dell	IBM	1999/11/19	virus <sup>i</sup>
Dell	IBM	2002/12/11	sitecrashed <sup>a</sup>
Direct TV	EchoStar Communication	2003/1/3	datatheft <sup>c</sup>
DoubleClick	ValueClick	2001/3/30	attack <sup>a</sup>
DoubleClick	ValueClick	2004/7/28	attack <sup>a</sup>
Drug Emporium	Drug Store Com Inc.	2000/1/30	siteshutdown <sup>c</sup>
eBay		2000/2/8	DoS <sup>a</sup>
eBay		2000/2/9	DoS <sup>a</sup>
eBay		2000/2/10	DoS <sup>a</sup>
Estee Lauder	Procter and Gamble	2000/5/5	virus <sup>ia</sup>
FedEx	UPS	2001/8/9	virus <sup>a</sup>
First Data Corp	Fiserv	2000/9/11	break in <sup>c</sup>
Ford Motor	General Motor	2000/5/5	virus <sup>ia</sup>
Ford Motor	General Motor	2005/12/22	datalost <sup>c</sup>
General Electric	Philips Electronics	1999/6/10	worms <sup>i</sup>
Hilton		2005/5/20	breach <sup>c</sup>

<sup>c</sup>Confidentiality, <sup>i</sup>Integrity, <sup>a</sup>Availability

Company	Control Company	Event Date	Type of Incident
Hewlett Packard	IBM	2001/2/15	attack <sup>a</sup>
Intel	AMD	1999/3/30	virus <sup>i</sup>
Intel	AMD	1999/6/10	worms <sup>i</sup>
Knight Ridder	Pulitzer	2003/9/10	attack <sup>a</sup>
Lockheed Martin	Northrop Grumman	1999/3/30	virus <sup>i</sup>
Marriott International		2005/12/28	datalost <sup>a</sup>
Mastercard		2003/2/19	hackinfo <sup>c</sup>
Mastercard	American Express	2005/6/19	attack <sup>c</sup>
McGraw-Hill	Moodys	2000/2/22	theft of data <sup>c</sup>
MCI WorldCom	Nextel	1998/12/21	virus <sup>i</sup>
MCI WorldCom	Nextel	1999/6/18	virus <sup>i</sup>
MCI WorldCom	Nextel	2001/12/6	securitybreach <sup>c</sup>
Merrill Lynch	Goldman Sachs	1999/3/30	virus <sup>i</sup>
Microsoft	IBM	1997/6/23	hacker <sup>a</sup>
Microsoft	IBM	1999/3/30	virus <sup>i</sup>
Microsoft	IBM	1999/6/10	worms <sup>i</sup>
Microsoft	IBM	1999/8/31	attack <sup>a</sup>
Microsoft	IBM	2000/10/27	attack <sup>c</sup>
Microsoft	IBM	2000/11/8	attack <sup>c</sup>
Microsoft	IBM	2001/1/25	DoS <sup>a</sup>
Microsoft	IBM	2001/1/26	DoS <sup>a</sup>
Microsoft	IBM	2001/8/10	worms <sup>i</sup>
Microsoft	IBM	2001/8/30	breach <sup>c</sup>
Microsoft	IBM	2001/11/5	breach <sup>c</sup>
Microsoft	IBM	2002/8/23	breach <sup>c</sup>
Microsoft	IBM	2003/8/15	worms <sup>ia</sup>
Microsoft	IBM	2004/2/13	codelost <sup>c</sup>
Microsoft	IBM	2004/4/14	breach <sup>i</sup>
Microsoft	IBM	2006/10/13	breach <sup>i</sup>
National Discount Brokers		2000/2/25	siteattack <sup>a</sup>
Network solutions		1999/7/3	siteattack <sup>a</sup>
New York Times	Dow Jones	1998/9/14	attack <sup>a</sup>
New York Times	Dow Jones	2002/7/12	deface <sup>a</sup>
Nike		2000/6/22	siteattack <sup>a</sup>
Sabre		2000/6/24	breach <sup>c</sup>
SBC		1999/6/10	worms <sup>i</sup>
SCO	IBM	2003/12/15	attack <sup>a</sup>
SCO	IBM	2004/2/2	virus <sup>i</sup>
SCO	IBM	2004/11/29	deface <sup>a</sup>
Siebel	PeopleSoft	2003/1/24	worm <sup>a</sup>
Southern Company	Unisource Energy	1999/6/10	worm <sup>i</sup>
Symantec	McAfee	1999/6/10	worm <sup>i</sup>
TD Ameritrade	Charles Schwab	2006/10/24	hack in account <sup>c</sup>
TJX	Macy's	2007/1/19	credit card info <sup>c</sup>
TJX	Macy's	2007/2/22	credit card info <sup>c</sup>
TJX	Macy's	2007/3/30	credit card info <sup>c</sup>
TJX	Macy's	2007/6/12	credit card info <sup>c</sup>
TJX	Macy's	2007/10/25	credit card info <sup>c</sup>
T-mobile (Deutsche Telekom AG)	Sprint	2005/1/13	hack in account <sup>c</sup>
ToysRus		1999/11/8	sitecrashed <sup>a</sup>
TransWorldAirlines	SkyWest	2000/3/21	Security breach <sup>c</sup>
USA Today (Gannett)	Tribune	2002/7/12	deface <sup>a</sup>

<sup>c</sup>Confidentiality, <sup>i</sup>Integrity, <sup>a</sup>Availability

Company	Control Company	Event Date	Type of Incident
Verisign	Entrust	2002/3/21	siteattack <sup>a</sup>
Walt Disney	CBS	2000/9/27	DoS <sup>a</sup>
Washington Mutual	Wachovia	2003/2/6	worm <sup>ia</sup>
Wells Fargo	US Bancorp	2006/3/13	breach <sup>c</sup>
Yahoo	Infospace	2000/1/11	disruption <sup>a</sup>
Yahoo	Infospace	2000/2/8	DoS <sup>a</sup>
Yahoo	Infospace	2000/2/9	DoS <sup>a</sup>
Yahoo	Infospace	2000/2/10	DoS <sup>a</sup>
Yahoo	Infospace	2004/7/27	virus <sup>i</sup>
Yahoo	Infospace	2005/3/24	phisher <sup>c</sup>

<sup>c</sup>Confidentiality, <sup>i</sup>Integrity, <sup>a</sup>Availability

#### Appendix D. Stock Price Reactions from Information Security Incidents

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

$$R_{it} = \beta_0 + \beta_1 R_{mt} + \varepsilon_{it} \quad (\text{D-1})$$

where  $R_{it}$  denotes company  $i$ 's return at period  $t$  which equals to  $(p_t - p_{t-1}) / p_{t-1}$ . 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.  $\beta_0$  and  $\beta_1$  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:

$$AR_{it} = R_{it} - \hat{\beta}_0 - \hat{\beta}_1 R_{mt} \quad (\text{D-2})$$

As shown by equation (A-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.,  $(\sum_{t=1}^N \sum_{t_0}^{t_1} AR_{it}) / N$ , 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.

## Appendix E. Cluster Analysis and Concept Links

The cluster analysis is performed as follows using SAS<sup>®</sup> 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  $1 + \sum_j (p_{ij} \log_2(p_{ij}) / \log_2(n))$ , where  $p_{ij} = f_{ij} / g_i$ ,  $g_i$  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  $\sum_{h=1}^k \omega_h f_h(x|\mu_h, \Sigma_h)$ , where  $\mu_h, \Sigma_h$  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  $(\omega_h^j f_h(x|\mu_h^j, \Sigma_h^j)) / (\sum_i \omega_i^j f_i(x|\mu_i^j, \Sigma_i^j))$  (SAS Institute Inc 2004). The iteration terminates if the likelihood value of two iterations is less than  $\varepsilon > 0$  or a maximum of five iterations are reached (SAS Institute Inc 2004). The text mining results are discussed in section 4.3.2.

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  $\text{Max}(4, A, B)$ .  $A$  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 F. Variable Definitions

<b>Variable</b>	<b>Definition</b>
$m$	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.
$\alpha$	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.
$\rho$	Proportion of privacy sensitive customers which is defined between zero and one.
$\delta$	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 $t$ .
$\lambda$	Mean-time-to-failure
$b$	Change of failure rate across time
$F_m(t)$	The probability of system failure (CDF) of two non-repairable component across time $t$ .
$\psi$	False acceptance rate (FAR) of a biometric system which is determined by the selected threshold.
$\varphi$	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
$\varepsilon$	The percentage change of customers, which depends on different systems. Therefore, we use ten different percentages for our analysis. $\varepsilon_1$ ( $\varepsilon_4, \varepsilon_7, \varepsilon_{10}$ ) 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 component and one biometric system). $\varepsilon_2$ ( $\varepsilon_5, \varepsilon_8$ ) 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 component and one biometric

	system). $\varepsilon_3$ ( $\varepsilon_6, \varepsilon_9$ ) 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 component and one biometric system).
--	--

## Appendix G. Conditions that Make the New Authentication System More Preferable

## Panel A. Shift to biometric system

<p><i>implementation costs:</i></p> $c_{bio} < c_n - V_{net\_bio} - F_{bio}(t; \bar{s})(V_{bio} + L_{bio}) + F_n(t)(V_n + L_n)$ <p>if <math>c_n - V_{net\_bio} - F_{bio}(t; \bar{s})(V_{bio} + L_{bio}) + F_n(t)(V_n + L_n) &gt; 0</math></p>
<p><i>percentage of privacy sensitive customers:</i></p> $\frac{-Y - \sqrt{Y^2 - 4XZ}}{2X} < \rho < \frac{-Y + \sqrt{Y^2 - 4XZ}}{2X}$ $X = F_{bio}(t; \bar{s})(1 - m)\varepsilon_3\varepsilon_4$ $Y = F_{bio}(t; \bar{s})(m\varepsilon_4 - m\delta\varepsilon_2\varepsilon_4) - (1 - m)\varepsilon_3 - F_n(t)m\varepsilon_1$ $Z = c_{bion} - c_n + m\delta\varepsilon_2 + F_{bio}(t; \bar{s})L_{bio} - F_n(t)L_n$ <p>if <math>-Y + \sqrt{Y^2 - 4XZ} &gt; 0</math> and <math>-Y - \sqrt{Y^2 - 4XZ} &gt; 0</math></p>
<p><i>percentage of convenience sensitive customers:</i></p> $\delta < \frac{c_n - c_{bio} + (1 - m)\rho\varepsilon_3(1 - F_{bio}(t; \bar{s})\rho\varepsilon_4) - F_{bio}(t; \bar{s})(m\rho\varepsilon_4 + L_{bio}) + F_n(t)(V_n + L_n)}{m\varepsilon_2(1 - F_{bio}(t; \bar{s})\rho\varepsilon_4)}$ <p>if <math>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) &gt; 0</math></p>
<p><i>market share:</i></p> $m > \frac{c_{bio} - c_n - \rho\varepsilon_3 + F_{bio}(t; \bar{s})\rho^2\varepsilon_3\varepsilon_4 + F_{bio}(t; \bar{s})L_{bio} - F_n(t)L_n}{\rho\varepsilon_3 + \rho\varepsilon_1F_n(t) - \delta\varepsilon_2(1 - F_{bio}(t; \bar{s})\rho\varepsilon_4) - \rho\varepsilon_4F_{bio}(t; \bar{s})(1 - \rho\varepsilon_3)}$ <p>if both the denominator and the nominator are positive or negative</p>
<p><i>expected losses:</i></p> $F_{bio}(t; \bar{s})L_{bio} < c_n - c_{bio} - m\delta\varepsilon_2 + (1 - m)\rho\varepsilon_3 - F_{bio}(t; \bar{s})(V_{bio}) + F_n(t)(V_n + L_n)$ <p>if <math>c_n - m\delta\varepsilon_2 + (1 - m)\rho\varepsilon_3 - F_{bio}(t; \bar{s})(V_{bio}) + F_n(t)(V_n + L_n) &gt; 0</math></p>

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

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

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

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

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)

<p><i>implementation costs:</i></p> $c_{nn} < c_{nbio} - V_{net\_nn} + V_{net\_nbio} - F_{nn}(t)(V_{nn} + L_{nn}) + F_{nbio}(t; \bar{s})(V_{nbio} + L_{nbio})$ <p>if <math>c_{nbio} - V_{net\_nn} + V_{net\_nbio} - F_{nn}(t)(V_{nn} + L_{nn}) + F_{nbio}(t; \bar{s})(V_{nbio} + L_{nbio}) &gt; 0</math></p>
<p><i>percentage of privacy sensitive customers:</i></p> $\rho < \frac{-Y - \sqrt{Y^2 - 4XZ}}{2X} \text{ or } \rho > \frac{-Y + \sqrt{Y^2 - 4XZ}}{2X}$ $X = F_{nbio}(t; \bar{s})(1 - m)\varepsilon_9\varepsilon_{10} - F_{nn}(t)(1 - m)\varepsilon_6\varepsilon_7$ $Y = F_{nbio}(t; \bar{s})(m\varepsilon_{10} - m\delta\varepsilon_8\varepsilon_{10}) - (1 - m)(\varepsilon_6 - \varepsilon_9) - F_{nn}(t)(m\varepsilon_7 - m\delta\varepsilon_5\varepsilon_7)$ $Z = c_{nbio} - c_{nn} - m\delta(\varepsilon_5 - \varepsilon_8) + F_{nbio}(t; \bar{s})L_{nbio} - F_{nn}(t)L_{nn}$ <p>if <math>-Y + \sqrt{Y^2 - 4XZ} &gt; 0</math> and <math>-Y - \sqrt{Y^2 - 4XZ} &gt; 0</math></p>
<p><i>percentage of convenience sensitive customers:</i></p> $\delta < \frac{c_{nn} - c_{nbio} + (1 - m)\rho(\varepsilon_9 - \varepsilon_6) - F_{nbio}(t; \bar{s})[(m + (1 - m)\rho\varepsilon_9)\rho\varepsilon_{10} + L_{nbio}] + F_{nn}(t)[(m + (1 - m)\rho\varepsilon_6)\rho\varepsilon_7 + L_{nn}]}{m(\varepsilon_8 - \varepsilon_5) + F_{nn}(t)m\rho\varepsilon_5\varepsilon_7 - F_{nbio}(t; \bar{s})m\rho\varepsilon_8\varepsilon_{10}}$ <p>if <math>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}] &gt; 0</math></p>
<p><i>market share:</i></p> $m > \frac{c_{nn} - c_{nbio} - \rho(\varepsilon_6 - \varepsilon_9) + F_{nn}(t)\rho^2\varepsilon_6\varepsilon_7 - F_{nbio}(t; \bar{s})\rho^2\varepsilon_9\varepsilon_{10} + F_{nn}(t)L_{nn} - F_{nbio}(t; \bar{s})L_{nbio}}{\rho(\varepsilon_9 - \varepsilon_6) + \delta(\varepsilon_8 - \varepsilon_5) + F_{nn}(t)\rho\varepsilon_7[\rho\varepsilon_6 - \delta\varepsilon_5 - 1] + F_{nbio}(t; \bar{s})\rho\varepsilon_{10}[1 - \delta\varepsilon_8 - \rho\varepsilon_9]}$ <p>if both the denominator and the nominator are positive or negative</p>
<p><i>expected losses:</i></p> $F_{nn}(t)L_{nn} < c_{nbio} - c_n + V_{net\_nbio} - V_{net\_nn} + F_{nbio}(t; \bar{s})(V_{nbio} + L_{nbio}) - F_{nn}(t)(V_{nn})$ <p>if <math>c_{nbio} - c_n + V_{net\_nbio} - V_{net\_nn} + F_{nbio}(t; \bar{s})(V_{nbio} + L_{nbio}) - F_{nn}(t)(V_{nn}) &gt; 0</math></p>