Spring 5-14-2015

Communication of Data Breaches Through Financial Statements: A Text Analysis Perspective

Gaurav Bansal
University of Wisconsin – Green Bay, bansalg@uwgb.edu

Amit Deokar
Penn State Erie, avd108@psu.edu

Jie Tao
Dakota State University, jtao16065@pluto.dsu.edu

Communication of Data Breaches Through Financial Statements: A Text Analysis Perspective

Gaurav Bansal
University of Wisconsin – Green Bay
bansalg@uwgb.edu

Amit Deokar
Penn State Erie
avd108@psu.edu

Jie Tao
Dakota State University
jtao16065@pluto.dsu.edu

ABSTRACT

Data breaches of companies in various industry segments have seen a significant increase over the past decade. Consumer data ranging from emails and bank account information to health information has been compromised through such data breaches that have raised grave information security and privacy concerns among the users and the organizations alike. Companies that are experiencing these data breaches have an obligation to communicate information about these incidents to their stakeholders and they do so through their financial reports. In this article, we analyze financial reports from a text analysis standpoint to identify key trends and formulate theoretical propositions. In that regard, we build on legitimacy theory as a foundation, and consider several factors such as the size of the data breach, type of information compromised, and coverage in the media.

Keywords

Data Breaches, Annual Reports, Text Analysis, Sentiment Analysis, Legitimacy Theory

INTRODUCTION

"Information privacy is an important information management issue that continues to challenge organizations" (Culnan and Williams 2009, p. 673). This statement is perhaps no less valued today than it was in 2009 when Culnan and Williams expressed it in reaction to TJ Maxx and ChoicePoint data breaches. Come 2015 and we are witnessing more frequent data breaches that are more costly than ever. The number of breaches and the size of such breaches continues to swell (Quick et al. 2015). According to the Identity Theft Resource Center (ITRC 2015), the number of U.S. data breaches hit a record high of 783 in 2014, which is 27.5% more than the number of breaches reported in 2013 and shows an 18.3% jump over the previous high of 662 breaches in 2010. Equally alarming was the 30% jump in 2013 over the number of breaches in 2012 as reported by ITRC (2014).

Disclosure of data breaches involving consumer information is mandated by an array of state laws in forty-seven states (NCSL 2015). Different states have mandated different terms of when and how notice must be given - which leads to variation in report timing and details of such events (Reuters.com 2014). There has been a recent surge in examining the handling of such breaches, including announcements (as shown in Table 1) - however to date we do not have adequate guidance on how to manage such breaches. There are conflicting suggestions being offered to organizations dealing with such events. On one hand, some studies, such as Symantec Corporation (2011), suggest that the organizations need to do due diligence before responding to such data breaches, and that responding hastily is penalized. On the other and, anecdotal evidence from the Target (Reuters.com 2014) breach suggests that a delayed response proves disastrous for the organization (CNBC 2015).

Building on the research involving the analysis of data breach incidents, we rely on legitimacy and media agenda setting theories, and conduct text analysis of breach related information contained in the firm's financial statements to develop research propositions pertaining to the coverage and sentiments conveyed in such reports. The paper is

organized as follows. The next two sections present the overarching theories and a salient review of the literature. Data analyses are presented next, followed by the results and the research propositions. The last section discusses conclusions, limitations and future research opportunities.

Overarching Theory and Literature Review

We rely on the overarching theoretical lens of Legitimacy Theory to investigate data breach coverage in annual reports and to ground our propositions. Legitimacy Theory argues that a corporation continually attempts to stay legitimate by demonstrating to the society, and its stakeholders in particular, that it is striving to comply with their "expectations" (Chan et al. 2014, p. 59) or "social contracts" (Brown and Deegan 1998, p. 22) which are always evolving and are by no means static or fixed (Brown and Deegan 1998). Corporations rely on various communication mechanisms such as financial statements (Hurst 1970) to communicate their strategies to society in order to stay legitimate. Table 1 provides a salient review of literature covering data / privacy breaches in the information systems area.

Acquisti et al. (2006)	Stock price analysis	Examines the relationship between firm's market value and data breach announcement.
Bansal and Zahedi (2015)	Structural Equation Modeling	Examines the moderating effects of two types of privacy violation (i.e. hacking and unauthorized sharing) on the trust violation and repair process.
Culnan and Williams (2009)	Theoretical paper	Studies breach incidents at TJ Maxx and Choice Point and provide recommendations for ways organizations can improve their privacy practices.
Garg et al. (2003)	Event study methodology	Examines the relationship between negative abnormal returns and security breach announcements and type of information breached.
Garrison and Ncube (2011)	Longitudinal study	Examines the relationship between five breach types and six types of institutions using longitudinal analysis of breach data spanning over five years.
Ko et al. (2009)	Matched-sampling methodology	Examines the impact of type of breaches (i.e., confidentiality, integrity, and availability) by IT intensiveness of the company breached and the size of the breach.
Wang et al. (2013b)	Text mining of breach reports	Examines the degree of information contained in the breach report pertaining to the incident and the degree of belief formed by the market pertaining to the firm's business value.
Wang et al. (2013a)	Text Mining of financial statements	Uses the text mining approach to relate information security risk factors from annual financial reports to the possibility of future security breaches.

Table 1. Salient Review of Literature on Data Breaches in the Information Systems Area

DATA

As a pilot study, we analyzed four data breaches: TJ Maxx, Citi Bank, Target and eBay by investigating their 10-K reports released immediately after the breach disclosure. We manually selected the mentions of data breaches in the annual reports. We did sentiment analysis using GATE (Cunningham 2002), along with Loughran and McDonald (2014)'s lexicon for positive, negative and uncertain words. Additionally, to understand the media coverage of the data breaches, we analyzed the counts of media exposure using organization names and "data breach" as keywords to search on a media search engine LexisNexis (i.e. "Target" and "Data Breach"). From the text and sentiment analysis on the aforementioned data breaches (via 10-K reports and media exposures), we obtained several insightful findings – which are reported in Table 2 and summarized here: (i) the mentions of data breaches (combining generic and specific) are decreasing in size (use ratio of word counts as proxy) over time; (ii) the more recent 10-K reports contain

more discussions toward data breaches in general, rather than the specific security incidents, and last (iii) the extent of the negative tone is probably positively correlated with the media exposure.

In Table 2: *generic* mentions refer to the portions in the 10-K reports discussing data breaches in general; *specific* mentions are the discussions of the specific incident(s); long term media count is the count of media exposures of the specific data breaches over a calendar year. The numbers in parentheses for generic/specific mentions are their ratios to the total word counts in the respective 10-K reports; the numbers in parentheses for positive, negative, and uncertain word counts are ratios to the word counts of mentions of specific data breaches; while the numbers in parentheses for long term media count are the daily average of media exposures over a calendar year (except for eBay – since the data breach occurred less than a year ago – which is 289 calendar days from breach to March, 6, 2015).

Company Name	Generic Mentions	Specific Mentions	Total Words in 10-K	Accounts Breached (Million Records)	Long term Media Count	Positive Word Count	Negative Word Count	Uncertain Word Count
TJ Maxx (2006)	0 (0.00%)	4,287 (7.80%)	54,943	94	27 (0.074)	19 (0.44%)	151 (3.52%)	50 (1.16%)
Citi Bank (2011)	136 (0.25%)	178 (0.32%)	154,652	0.36	52 (0.142)	1 (0.56%)	6 (3.37%)	1 (0.56%)
Target (2013)	65 (0.12%)	2,400 (4.37%)	37,650	7	990 (2.71)	23 (0.96%)	58 (2.42%)	42 (1.75%)
eBay (2014)	1,430 (2.60%)	348 (0.63%)	77,871	145	234 (0.809)	4 (1.15%)	16 (4.60%)	7 (2.01%)

Table 2. Descriptive Statistics of Pilot Study

THEORETICAL PROPOSITIONS

Legitimacy theory suggests that corporations try to adjust to evolving societal expectations; and it is also known that societal expectations are probably being shaped by omnipresent media more than by anything else. According to Media Agenda Setting Theory there is a causal relationship between the degree of coverage of a particular topic and the degree of salience accorded to the topic by the general public (Ader 1995; Brown and Deegan 1998). Further, Brown and Deegan (1998) note the strong association between the high level of media attention received by adverse incidents and the corresponding high level of attention given by the companies in their annual reports by means of disclosures. The results from our pilot study support this notion. This leads us to posit:

Proposition 1: The larger the media attention, the greater the coverage in annual reports.

Legitimacy Theory argues that corporations strive to maintain their legitimacy by trying to achieve "moving" expectations. Amidst revelations by Edward Snowden about government snooping of users' data, the ensuing debate between security and privacy, and the frequent occurrences of much publicized data breaches, it is believed that privacy is increasingly losing ground (Sterner 2014). The Center for the Digital Future (2014), in its recent report, suggests that the percentage of extremely privacy concerned individuals has declined for the fourth consecutive year. Thus, we argue that the societal concerns regarding the gravity of the data breach incident are declining and so are the "expectations." Moreover, companies are also increasingly averse in disclosing their data breaches promptly as evident from the delays observed in recent high profile breaches such as Target, Kaiser Foundation Health Plan, Uber, and USPS. As such, we argue that the extent of attention given by the companies in their annual reports following such incidents is generally declining. The results from our pilot study support this notion. Hence, we propose that:

Proposition 2: The annual report coverage of data breach incidents is declining over time.

From text analysis of the annual reports, we observe that in recent data breach incidents, companies acknowledge these issues in a broader sense, placing a general disclaimer and putting them in the context of the increasing number of information security threats and data breach incidents. The attention given to the specific data breach incident is much less than the broader contextual discussion around this issue.

We find support for this observation in the literature as well. Lindblom (1994) describes various strategies that organizations adopt as they rationalize adverse events and legitimize them through communications such as annual reports. One such strategy is that of altering the perceptions of stakeholders by means of diversion techniques that bring focus on other related issues. Another strategy is that of trying to change outsiders' performance expectations about the company. Both of these strategies align with our observations of diluting the attention from the data breach incidents by discussing generic security breach information and possibilities in the annual reports. As such, we posit:

Proposition 3: The more recent the data breach, the greater the "general" coverage (as opposed to the breach "specific" coverage) in annual reports.

Legitimacy Theory (Brown and Deegan 1998) suggests that an organization will be rebuked and penalized if its operations and reporting are not aligned with public expectations. This viewpoint has been validated in the context of corporate reporting and disclosures. Hogner (1982) conducted a longitudinal study of annual reports of US Steel Corporation over 80 years and found that disclosures in reports were essentially well-articulated responses to social expectations of the corporation as held by the external stakeholders and the general public. We also find support for this rationale through our text analysis in which greater media attention raises social expectations from the stakeholders about acknowledging the adverse nature of the incident, which in turn leads to more negative sentiments expressed in the disclosures through annual reports. Accordingly, we posit:

Proposition 4: The greater the media attention, the more negative the sentiments exhibited in the annual reports.

It seems that the retail industry provides greater coverage in their annual reports than any other business type. It could be due to the fact these organizations stand to lose the most from a loss of consumer trust and subsequent attrition. Echoing similar sentiments, the Acquisti et al. (2006) study found that in the wake of a data breach the retail industry segment is likely to be affected more because of the low switching costs for consumers as compared to other segments, such as banks, where the switching costs are higher. Moreover, the results from our pilot study support this notion. Based on this rationale, we propose:

Proposition 5: Annual report coverage is dependent on the industry type.

Concluding Remarks and Future Work

In this study we have attempted to tease out the patterns observed in how data breach incidents are conveyed to the company stakeholders in a formal manner through financial reports. Our observations lead us to posit that the discussion in financial reports is influenced by a number of factors including how much media attention has been received for data breach incidents, the impact of information loss occurred, and the general breach announcement burnout experienced by stakeholders due to increased number of incidents. Further, the coverage can be viewed as a response to media attention in terms of the negative sentiment expressed in the reports to align with public expectations, and the use of deflection techniques such as greater coverage of general security issues and disclosures, as compared to providing greater detail about the incident itself.

We plan to extend this study further by investigating social media coverage in addition to news media coverage of data breach incidents. Also, the issue of corporate governance in this context as a factor influencing annual report coverage may be studied as well (Chan et al. 2014). We believe this research stream has strong potential to provide insights into the communication approaches used by public companies in responding to adverse events in general.

References

Acquisti, A., Friedman, A., and Telang, R. 2006. "Is there a cost to privacy breaches? An event study," in the proceedings of the *International Conference on Information Systems*.

Ader, C. R. 1995. "A longitudinal study of agenda setting for the issues of environmental pollution," *Journalism and Mass Communication Quarterly* (72:2), pp. 300-311.

Bansal, G., and Zahedi, F. 2015. "Trust Violation and Repair: The Information Privacy Perspective," *Decision Support Systems* (71), pp. 62-77.

Brown, N., and Deegan, C. 1998. "The public disclosure of environmental performance information-a dual test of media agenda setting theory and legitimacy theory," *Accounting and Business Research* (29:1), pp. 21-41.

Center for the Digital Future. 2014. "The 2014 Digital Future Project,"

Chan, M. C., Watson, J., and Woodliff, D. 2014. "Corporate governance quality and CSR disclosures," *Journal of Business Ethics* (125), pp. 59–73.

CNBC. 2015. "Why did Target take so long to report the breach?," (http://www.cnbc.com/id/101287567)

Culnan, M. J., and Williams, C. C. 2009. "How ethics can enhance organizational privacy: Lessons from the Choicepoint and TJX data breaches," *MIS Quarterly* (33:4), pp. 673-687.

Cunningham, H. 2002. "GATE, a General Architecture for Text Engineering ," *Computers and the Humanities* (36), pp. 223-254.

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

Garrison, C. P., and Ncube, M. 2011. "A longitudinal analysis of data breaches," *Information Management & Computer Security* (19:4), pp. 216-230.

Hogner, R. H. 1982. "Corporate social reporting: Eight decades of development at US steel," *Research in Corporate Performance and Policy* (4), pp. 243-250.

Hurst, J. W. 1970. *The legitimacy of the business corporation in the law of the United States 1780-1970*, The University Press of Virginia: Charlottesville: .

ITRC. 2014. "2013 Data Breach Category Summary," (http://www.idtheftcenter.org/ITRC-Surveys-Studies/2013-data-breaches.html)

ITRC. 2015. "Identity Theft Resource Center Breach Report Hits Record High in 2014," (http://www.idtheftcenter.org/ITRC-Surveys-Studies/2014databreaches.html)

Ko, M., Osei-Bryson, K.-M., and Dorantes, C. 2009. "Investigating the impact of publicly announced information security breaches on three performance indicators of the breached firms," *Information Resources Management Journal* (22:2), pp. 1-21.

Lindblom, C. K. 1994. "The implications of organisation legitimacy for corporate social performance and disclosure," in the proceedings of the *Paper presented at the Critical Perspectives on Accounting Conference, New York. NY.*

Loughran, T., and McDonald, B. 2014. "Measuring readability in financial disclosures," *Journal of Finance* (69), pp. 1643-1671.

NCSL. 2015. "Security Breach Notification Laws " (http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx)

Quick, M., Hollowood, E., Miles, C., and Hampson, D. 2015. "http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/," (http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/)

Reuters.com. 2014. "U.S. companies allowed to delay disclosure of data breaches," (http://www.reuters.com/article/2014/01/16/us-target-data-notification-idUSBREA0F1LO20140116)

Sterner, E. 2014. "The security vs. privacy debate is already over, and privacy lost," (http://www.washingtonexaminer.com/the-security-vs.-privacy-debate-is-already-over-and-privacy-lost/article/2545407)

Symantec Corporation. 2011. "2010 Annual Study: U.S. Cost of a Data Breach," (http://www.symantec.com/content/en/us/about/media/pdfs/symantec_ponemon_data_breach_costs_report.pdf)

Wang, T., Kannan, K. N., and Ulmer, J. R. 2013a. "The Association Between the Disclosure and the Realization of Information Security Risk Factors," *Information Systems Research* (24:2), pp. 201-218,494-495.

Wang, T., Ulmer, J. R., and Kannan, K. 2013b. "The textual contents of media reports of information security breaches and profitable short-term investment opportunities," *Journal of Organizational Computing and Electronic Commerce* (23:3), p. 200.