

When Growth Opportunities Meet Cybersecurity Risk: A Topic Modeling Approach

Tracie S. Frost
School of Accounting & Finance, The
Hong Kong Polytechnic University
tracie.frost@polyu.edu.hk

Muktak Tripathi
IE Business School,
IE University
muktak.tripathi@ie.edu

Vincent Y. Zhuang
School of Accounting & Finance, The
Hong Kong Polytechnic University
vincent.zhuang@polyu.edu.hk

Abstract

Using textual analysis to capture firms' business prospects, this paper examines whether firms with high growth prospects are more vulnerable to cybersecurity risk, whether breaches impact growth firms' investment cycles, and how growth firms can protect against data breaches. Our findings suggest that firms with high growth prospects are greater targets for data breaches and provide compelling evidence that the particular firm characteristics that are hallmarks of growth firms may open those firms to greater cybersecurity risk. Further, our findings suggest that high growth firms maintain a robust investment cycle, despite experiencing a data breach, consistent with the growth mindset that these firms adopt. Finally, we provide evidence that growth firms with higher IT awareness can better defend themselves from cyber attacks.

Keywords: business prospect, data breach, high growth opportunity, IT investment, IT awareness

1. Introduction

Data breaches are a growing concern for firms across the globe. In advanced economies, both firm executives and market participants now view cybersecurity risk as one of the foremost global concerns (WEF 2020). This heightened level of concern is justified, considering the surge in major cyberattacks witnessed in recent years. Despite considerable investments made in information security systems, the majority of firms continue to face significant vulnerability to cybersecurity risks¹. As such, there is a need for more research that investigates the determinants of cybersecurity risk so that firms can mitigate the risk of data breaches.

In this study, we examine whether a specific type and characteristic of a firm is more susceptible to data breaches. Employing the opportunity theory of crime Hannon (2002), we ask whether firms with high growth prospects are more likely to sustain a data breach than other firms. Firms with high growth opportunities are characterized by more capacity expansion projects, new product lines, acquisition of other firms, and maintenance and replacement of existing assets – all of which create opportunities for cybercrime (e.g., Chaduvula et al., 2018; Gul, 1999). We also follow our research to the logical conclusion and test whether data breaches impact the ability of

growth firms to execute their investments. Finally, we investigate whether firms with higher IT awareness (as evidenced by the presence of a chief technology officer [CTO]) are better able to fend off breaches.

This topic is of paramount interest to managers, investors, and customers. Particularly, Survey evidence suggests that growth firms focus on investment to the detriment of cyber security (Nelson & Madnick, 2017). Managers of growth companies may take notice of their firms' vulnerability to cyber criminals and take additional measures to protect themselves. Investors may use our study to make informed investment decisions when assessing risks associated with data breaches. Finally, customers of growth firms may take steps to protect themselves, given our evidence that growth firms are more likely the targets of data breaches.

Our sample ranges from 2004 to 2019. We compare levels of data breaches associated with firms' growth opportunities based on a propensity-score matched sample of firms with and without data breaches. To measure growth opportunity, we employ the method of Banker et al. (2022) in using latent Dirichlet allocation (LDA), an unsupervised Bayesian machine-learning approach developed by Blei et al. (2003), to identify growth opportunities from analyst business descriptions. We use a principal component of the LDA measure of growth prospects, textual sentiment, and textual uncertainty to measure growth opportunity. Together, the three components form a three-dimensional measure of growth opportunity that incorporates the firm's opportunities for discretionary expenditures and the risk aversion and uncertainty related to the realization of growth prospects.

Consistent with our expectations, we find that growth opportunities are positively associated with data breaches in our matched sample, providing evidence that fundamental firm characteristics may make firms more vulnerable to cyber attacks. In additional analysis, we find that growth firms continue to invest following a breach and that their investment cycles do not increase, as compared to low growth firms. We also find that growth firms with IT awareness are less likely to suffer a data breach.

We contribute to the literature in the following ways. First, we expand the data breach literature. Prior literature in the accounting space generally focuses on the outcomes of data breaches and ways to mitigate data breaches. Research on firm-specific determinants

URI: <https://hdl.handle.net/10125/107121>

978-0-9981331-7-1

(Communication) research and advisory firm, estimated worldwide spending on information security products of \$124 billion in 2019, representing an increase of 8.8% from 2018.

HICSS

Page 6128

of data breaches is less developed. We expand this literature by identifying a firm characteristic that makes firms more susceptible to data breach. Second, we add to the finance literature on growth opportunity. The corporate finance literature documents that high-growth opportunity firms pose a greater financial risk than low-growth opportunity firms. However, whether growth opportunity is associated with operational risk is less clear. We add to this literature by showing that growth opportunity exposes firms to a higher risk of data breach. Finally, we add to the growing literature that uses textual analysis to identify firm characteristics with our use of LDA and bag-of-words approaches to measure growth opportunity.

2. Literature Review and Hypothesis Development

Data Breaches

Data breach events are increasing in frequency and severity (Ponemon and IBM, 2021), making their study of paramount importance for firms. In the accounting space, most data breach research falls into one of two streams: (1) effects of breaches; (2) actions firms can take to mitigate breaches. In general, the first stream of research indicates that data breaches have negative financial (see e.g., Huang & Wang, 2021; Richardson et al., 2019; Walton et al., 2021) and operational (He et al., 2020) implications for firms.

The second stream of research examines steps that firms take to mitigate breaches. Prior work documents several elements that may reduce cyber security risk, including IT executives (Kwon et al., 2013), non-IT executives with IT expertise (Haislip et al., 2021), board-level technology committees (Higgs et al., 2016), and internal auditors with security expertise (Islam et al., 2018). Moreover, previous research demonstrates that higher cybersecurity awareness and more IT (security) investment are effective in reducing data breaches (Li et al., 2023).

There is less research investigating the determinants of data breaches. Despite investing in IT security and human capital, most firms are still exposed to cybersecurity risk (Zhuang et al., 2020). Thus, there is a need for more empirical research to identify where the cybersecurity risk originates. Recently, emerging evidence suggests that breaches may be linked to specific firm attributes. For example, Li and Walton (2023) find an association between data breaches and firm business strategy. Relatedly, La Torre et al. (2018) document a link between firms with significant intellectual capital and data breaches. In this vein, we suggest that a firm's growth opportunity may also be a distinct firm characteristic that increases a firm's cybersecurity risk.

Growth Opportunity and Opportunity Theory of Crime

The opportunity theory of crime suggests that cybercriminals seek out vulnerable victims for their

attacks (e.g., Hannon, 2002; Sen & Borle, 2015). In other words, cybercriminals are more likely to target firms that are easier to breach. Prior work finds that in cybercrime, vulnerabilities in information systems, software, and firmware are associated with data breaches (Sen & Borle, 2015), but the literature also identifies rapid development of new product lines, acquisitiveness, and lack of spending on cybersecurity as hallmarks of cybersecurity risks. Anecdotal evidence also supports these factors as opportunities for cyber criminals to exploit. Take, for example, Amazon, which is a growth firm in our sample. Amazon is notable for aggressive new product development that often involves partnerships with other organizations, as well as high-profile acquisitions. Further, Amazon has faced fines for breaking data protection laws (Leggett, 2021), suggesting weakness in cyber security. Since 2012, Amazon has averaged more than one major data breach per year (Heiligenstien, 2023). Notably, several of these breaches are attributable to companies that Amazon acquired (Zappos, Twitch), partners (third-party sellers), and products (Kindle).

The factors that have played a role in Amazon's data breaches are also characteristics of growth firms in general. Relying on established corporate theories (see e.g., Myers, 1977; Smith & Watts, 1992), we define growth opportunity as the component of firm value resulting from the firm's options to make future investments (i.e., the opportunity to make positive net present value [NPV] investments). Firms often have growth opportunities to collaborate with partners in new product development (Hutchinson and Gul, 2004) and acquisitions (Margsiri et al., 2008). These factors, which are typical of firms with growth opportunities, may increase their cybersecurity risk, as compared to low-growth firms.

Firstly, growth firms undertake more new product lines. Such projects often require sharing of sensitive information such as business intelligence, intellectual property, and customer information with collaborators (Chaduvula et al., 2018). However, these collaborations magnify cybersecurity risk due to differences in security practices between firms. Further, features of new product development such as open design (Anderson, 2002), interdependent subsystems (Kocher et al., 2004), and model integrations (Chaduvula et al., 2018) create vulnerabilities in IT security for growth firms compared to non-growth firms, suggesting that growth firms may be easier targets for cybercriminals.

Secondly, growth firms are more likely than non-growth firms to make acquisitions of other firms. Prior work suggests that business combinations increase cybercriminals' opportunity to breach acquirers and targets because due diligence often overlooks the health of the target's cybersecurity defense systems. An acquisition means buying the target's "past, present and future data security problems" (Okafor, 2021) and potentially creating

vulnerabilities in the combined companies' information systems, software, and firmware (Sen and Borle 2015) that cybercriminals can exploit.

Third, growth firms may choose to spend less money on cyber security than their non-growth counterparts in pursuit of extreme growth (e.g., Gordon & Smith, 2007; Nelson & Madnick, 2017). Gordon and Smith (2007) note that cybersecurity activities compete with other organizational activities such as new product development, R&D, and M&A, while Nelson and Madnick (2017) find that the majority of Chief Information Officers in their survey of highly-innovative firms believe that their firms take on too much cybersecurity risk in order to achieve their growth objectives. Such findings suggest that growth firms' focus on positive NPV projects may outweigh their concern for cyber security, making growth opportunity a unique identifier for breach likelihood. Lack of attention to and spending on cybersecurity may create weaknesses in systems and software that make it easier for cybercriminals to breach growth firms as compared to non-growth firms.

In summary, growth opportunity firms' propensity for new product development, business combinations, and prioritization of growth over security may make them easier breach targets for cybercriminals, as suggested by prior findings that associate these three growth firm characteristics with higher-than-average cybersecurity risk. Therefore, based on the opportunity theory of crime which suggests that cybercriminals are more likely to target vulnerable breach targets, and our argument that the unique characteristics of growth firms create vulnerabilities in information systems, software, and firmware, we frame our hypothesis as follows:

Hypothesis: Firms with high growth opportunity are more likely to sustain a data breach than other firms.

3. Research Design

Data and Sample Selection

We follow prior research and use Privacy Rights Clearinghouse (see privacyrights.org) to identify 829 data breaches reported by public companies between 2004 and 2019 (e.g., He et al., 2020; Li & Walton, 2023). Privacy Rights Clearinghouse maintains a list of breach incidents collected from various sources, such as US state governments and firm disclosures. Our sample begins in 2004 because it is the first year privacyrights.org collected breach data. We end our sample period in 2019 because this is the last year for which we have calculated growth opportunity measures. We include all types of breach incidents and include rolling counts of multiple breaches at the same firm. We further require firms to have available data

from Compustat to calculate the growth opportunity measure and control variables in our models. We retain 336 public company data breaches.

[Insert Table 1 Here]

It is possible that some firms are more prone to be targeted for data breaches than others. To address this possibility, we use a propensity score matching (PSM) approach to identify a set of control firms that are most similar to the treatment firms in their likelihood of being breached (Shipman et al., 2017). Following Higgs et al. (2016) and Xu et al. (2019)², we match on firm size (*Size*), measures of performance (*ROA*) and valuation (*Book-to-Market*), and complex multinational operations (*Foreign Operations*) consistent with larger, more profitable firms with more complex structures being more attractive breach targets. We also include measures of financial flexibility (*Cash Holdings*, *Liquidity*), obligations (*Leverage*), and commitments (*Dividend*) because more financially sound firms are better equipped to take steps that reduce breaches. We match on intangible assets (*Intangibles*) and tangible assets (*PP&E*) as firms with these assets are more (less) likely to possess significant capital that would make them attractive to hackers. Finally, we include employees (*Employees*) to account for larger firms' increased susceptibility to malware and phishing scams (Eaton et al., 2019).³ We match all breached firms, including those affected by data hacks or denial-of-service (DoS) attacks, with the non-breached firms using the nearest neighbor matching, without replacement, based on the firm-level propensity⁴. The propensity score is estimated using a binary choice Probit model specified below in Eq. (1):

$$\begin{aligned} Breach_j = & Growth\ Opportunity_j + ROA_j \\ & + Dividends_j + Foreign\ Operations_j \\ & + Cash\ Holdings_j + Liquidity_j + \\ & + Employees_j + PP\&E_j + Intangibles_j \\ & + Leverage_j + Book-to-Market_j + Size_j + \varepsilon_j \end{aligned} \quad (1)$$

where *Breach* is an indicator variable equal to 1 if the firm has suffered a data breach, and 0 otherwise. All variables are defined in the Appendix. Our final sample is 336 firms with data breaches and 336 control firms. The total number of firm-year observations is 9,150; however, we lose 828 firm-year observations in the creation of lagged variables, so our final sample is 8,322 firm-year observations. Panel A of Table 2 summarizes the sample selection.

Panel B of Table 2 contains descriptive statistics for our pooled sample. Approximately 23.4 percent of firm years are affected by a data breach. The distribution of *Growth Opportunity* is skewed to the right, indicating that most firms in our sample have relatively low *Growth Opportunity*, along with the

² See Table 1 for detailed variable definitions.

³ In untabulated tests, we include controls for the time-variant effects of CEO, CFO, and Directors, and our results are unchanged.

⁴ We also used a 1:3 match and the results are robust.

presence of a smaller number of firms with very high *Growth Opportunity*. Firms mean (median) investment cycle is 11.6% (10.2%), which is, relative to total assets, the rate of operating cash flows spent on the capital expenditures and acquisitions in lag ($t-1$), current (t), and leading ($t+1$) years.⁵ The average investment cycle of 11.6% (0.116) corresponds to 42.3 days of 365 days in a year and a turnover rate of 8.7 times (365/42.3). Panel C of Table 2 presents the means (medians) of control variables for the PSM sample. No significant differences exist between samples. [Insert Tables 2a, 2b, 2c Here]

Growth Opportunity Measure

Prior research in cyber security provides evidence that textual information gleaned from public disclosures is able to predict data breach likelihood (Florackis et al., 2022). Following this literature we employ the methodology of Banker et al. (2022) in identifying three vectors of growth opportunity based on forward-looking textual measures of firms' business prospects, sentiment, and financial uncertainty which we suggest are associated with the likelihood of data breach. The first vector, called "business prospects" is obtained through LDA topic analysis of firms' business descriptions. These descriptions are prepared by analysts as an introduction to equity-based research reports and investment recommendations, following the standards set by professional bodies like the Chartered Financial Analyst institute and regulatory mandates. Thus, these business descriptions serve as reliable representations of the firms' actual business operations. By utilizing this approach, the assessment of business prospects captures the firm's growth potential within the framework of its financial, regulatory, and technological limitations (Smith & Watts, 1992).

To measure growth prospects, we use a topic modeling approach in which we employ LDA on S&P business descriptions sourced from the Compustat North American database's business descriptions File. LDA is a statistical technique utilized for analyzing the textual content of original documents to uncover underlying themes present within a large body of text. LDA identifies topics based on the probability of words co-occurring within documents. This approach mitigates researcher bias involved in manually coding the textual content of S&P descriptions into business prospects. It also eliminates the need for pre-determined word dictionaries or topic categories. LDA enables the classification of extensive text collections and finds widespread application in business research for identifying topics within financial corpora (e.g., Brown et al., 2020; Dong et al., 2018; Dyer et al., 2017; Huang et al., 2018).

We undertake the following steps to carry out the topic analysis and identify firms' business prospects.

First, we generate a database of text using S&P business descriptions from the Compustat database. After cleaning the text, we identify the most useful number of topics based on a 20% random training sample. We obtain maximum likelihood estimates for the number of topics, and we use that number to perform LDA across the entire textual corpus. Using the top 10 words for each topic, we identify topic-word combinations that relate to several dimensions of business prospects. In particular, we find that variations in business prospects are associated with industry-specific topics, such as information technology and energy, that suggest systematic differences in growth opportunities between industries. We also find variations in more general topics such as geographical diversification and mergers and acquisitions, which are not specific to any single industry but indicate investment-related options available to firms.

Based on prior literature that indicates that textual tone has incremental explanatory power above content-based textual analysis and financial variables alone, we also incorporate sentiment and financial uncertainty in the measure of growth opportunity (e.g., Li, 2010; Loughran & McDonald, 2011). We quantify sentiment, the second vector, using the Harvard General Inquirer and Diction word lists to capture the psychological sentiment of the firm from S&P business descriptions. The ratio of positive words to total words less the ratio of negative words to total words describes firms' expected payoffs from their forward-looking business prospects. The third vector, financial uncertainty, is constructed using Loughran and McDonald (2011) financial uncertainty list. The ratio of financial uncertainty words to total words describes the firm's emphasis on financial risk and uncertainty with regard to realizing the firms' growth opportunities. The combination of these three components creates a comprehensive 3-dimensional measure of growth opportunity. This measure takes into account not only the firm's potential for discretionary spending but also factors in its level of risk aversion and the uncertainty surrounding the realization of business prospects. By incorporating these elements, the measure provides a more holistic understanding of the firm's growth potential.

To create the Growth Opportunity measure from the three text-based components, we employ principal component analysis (PCA). PCA serves the purpose of reducing dimensionality and deriving weights that account for endogeneity resulting from organizational decisions related to business integration and operations. By employing PCA, we are able to combine the three individual components of *Growth Opportunity* into a composite index, representing a singular variable that captures the growth opportunity set. Confirmatory factor analysis further validates that

⁵ E.g., the relationship between operating cash flows and investments (see Dechow, 1994; Dechow et al., 1998; Khan & Watts, 2009).

each of the three dimensions of *Growth Opportunity* carries unique information content, contributing distinctively to the overall measure. Banker et al. (2022) provides a validation of the individual components of Growth Opportunity.

Empirical Specifications

We employ ordinary least squares (OLS) regression to test the relationship between growth opportunity and data breach as follows:

$$\begin{aligned} \ln \text{Breach}_{j,t} = & \text{Growth Opportunity}_{j,t-1} + \text{ROA}_{j,t} \\ & + \text{Dividends}_{j,t} + \text{Foreign Operations}_{j,t} \\ & + \text{Cash Holdings}_{j,t} + \text{Liquidity}_{j,t} \\ & + \text{Employees}_{j,t} + \text{PP\&E}_{j,t} \\ & + \text{Intangibles}_{j,t} + \text{Leverage}_{j,t} \\ & + \text{Book-to-Market}_{j,t} + \text{Size}_{j,t} + \text{FirmFE} \\ & + \text{YearFE} + \varepsilon_{j,t} \end{aligned} \quad (2)$$

where the variable of interest is the principal component of the textual measure of growth opportunity from S&P business descriptions, sentiment, and uncertainty. *lnBreach* is the natural logarithm ($\ln(x+1)$) of one of three variables of data breaches: hack (*lnData-Hack*), DoS attack (*lnDoS-Attack*), or data breach (*lnData-Breach*). Control variables are defined in the Appendix. We include firm and year fixed effects and cluster robust standard errors by year.⁶ The hypothesis predicts that the coefficient of *Growth Opportunity* is positive.

4. Results

We document our findings for H1 in a multivariate setting using Eq. (2) in Table 3. In column (1), the dependent variable is *lnData-Breach*, and *Growth Opportunity* loads positively and significantly (0.066, $p\text{-value} < 0.01$), suggesting that firms with high *Growth Opportunity* are greater targets for data breaches. When we break data breaches into hacks (*lnData-Hack*) and DoS attacks (*lnDoS-Attack*), we find that *Growth Opportunity* is positively associated with both (*lnData-Hack* 0.064, $p\text{-value} < 0.01$; *lnDoS-Attack* 0.006, $p\text{-value} < 0.01$). These findings support hypothesis 1 and provide evidence that the particular firm characteristics that are hallmarks of growth firms may open those firms to greater cybersecurity risk. Our results are economically significant. *Growth Opportunity* increases the likelihood of data breach by 6.8 percent on average, and a one standard deviation increase of *Growth Opportunity* increases the likelihood of data breach by 29 percent. **[Insert Table 3 Here]**

Next, we investigate an outcome of data breaches on firm performance that is particularly pertinent to growth firms. Specifically, we examine the impact of

data breaches on the length of the investment cycle. We expect that data breaches will lengthen the investment cycle on average. However, following prior research which finds that firms with a strong focus on innovation do not cut back on R&D after a breach, while firms with no focus on innovation intensity do cut back on their R&D (He et al., 2020), we anticipate that high growth firms continue the pace of investment following a breach, compared to low-growth firms. Table 4 presents our results. While our results do suggest that data breaches are associated with longer investment cycles on average (0.003, $p\text{-value} < 0.01$ and 0.004, $p\text{-value} < 0.01$ in columns [1] and [4], respectively), we also find that firms with high *Growth Opportunity* are less likely to experience an increase in investment cycle than their low-growth counterparts. In column (3), when the dependent variable is *Invest-Cycle (t+1)*, the coefficient on *lnBreach* is -0.005 for high-growth firms, whereas the coefficient on *lnBreach* in column (2) is 0.008 ($p\text{-value} < 0.01$) for low growth firms⁷.

On average, in column (1), cybersecurity breach increases investment cycles by 2.6% more than the firms' average rate ($\exp(0.003/0.0116)$). This equates to 1.1 days per investment cycle or 9.5 days per fiscal year, corresponding to a reduction in turnover rate of 8.4 times from 8.7 times. These findings emphasize the focus on maintaining investment in positive NPV projects for high growth firms. **[Insert Table 4 Here]**

We have argued that growth firms' unique characteristics make them more attractive targets for cybercriminals. We now investigate whether growth firms can reduce their cyber vulnerability. This test is important in understanding how companies whose business models naturally expose them to greater cybersecurity risk may mitigate that risk. Prior work suggests that firms which signal higher IT awareness by employing a CIO and/or a CTO have fewer data breaches (e.g., Kwon et al. 2013; Li and Walton 2023). Having a CIO or CTO indicates that the firm has information security maturity and a strong security culture (Kwon et al. 2013) and may discourage cybercriminals from targeting the firm because a firm with a strong culture of IT security and awareness is likely to manage information security risks more effectively (Li and Walton 2023). To investigate whether a signal of IT awareness discourages cybercriminals from targeting growth firms, we interact *Growth Opportunity* with our measure of *IT awareness*, whether the firm has a CIO and/or a CTO. Our findings are presented in Table 5, Panel A. Consistent with our reasoning, we observe a negative and significant coefficient on the interaction (-0.082, $p\text{-value} < 0.01$; -0.010, $p\text{-value} < 0.05$; and -

⁶ Our results remain unchanged when we cluster standard errors by firm or firm and year instead of year. We also perform a lead-lag analysis in which we include controls at year $t-1$ in addition to year t . Our results are qualitatively unchanged.

⁷ In the interest of space, we omit our findings for *lnData-Hack* and *lnDoS-Attack*; however, the results are consistent with *lnBreach*.

0.084, p-value<0.01 in columns [1], [2], and [3], respectively). This result supports the findings of Nelson and Madnick (2017) that firms which rely on innovation and product development have fewer cyber security breaches when they have IT leadership. **[Insert Table 5 Panel A Here]**

Lastly, we perform a cross-sectional test based on firm life cycle following (Dickinson, 2011). We divide our sample into two groups representing firms that are in a cash-growing cycle (*Introduction*, *Growth*, or *Mature* firms) or a cash-shrinking cycle (*Shake-out* or *Declining* firms). We suggest that firms in the *Introduction*, *Growth*, or *Maturity* phases of the firm life cycle are likely to be able to take advantage of growth opportunities to a greater extent than those firms in the *Shake-out* or *Declining* phases, given their access to cash. Building on our primary findings, as firms in cash-growing cycles take advantage of their cash stores to invest in positive NPV projects, we expect that those projects may increase their risk of data breach through the channels of acquisitions, new product development, and hyper-focus on growth. Our findings are presented in Table 5, Panels B and C.

In Panel B, our results suggest that firms with higher growth opportunities in cash-growing cycles are more likely to experience a data breach than firms with few growth options. Conversely, in Panel C we observe that firms whose cash flow is shrinking (e.g., those firms that do not have the means to invest as freely in positive NPV projects) are less likely to experience data breaches. However, even in this subsample, we find that those with greater growth opportunities are subject to more data breaches than their lower-growth counterparts. These results provide additional evidence that growth firms' characteristics may make them more vulnerable to data breaches than low growth firms. **[Insert Table 5 Panel B & C Here]**

5. Conclusion

We provide evidence that the type and characteristics of firms may make them more susceptible to data breaches. In addition, we show that firms with a growth focus continue to pursue that growth focuses even following a data breach by maintaining the speed of their investment cycle. Further, following prior research, we show that growth firms may defend themselves against cybersecurity risk by implementing IT security steps in the firm.

Our research complements practice literature that suggests that cybersecurity risk is associated with certain types of firms, particularly those which operate in environments that expose them to excess risk and that those firms may take certain steps to reduce their risk (Aiyer et al., 2022). The novel innovation of our study is that we find that those firms can be identified through textual analysis of common analyst reports. As such, our work contributes to the data breach, accounting and finance, and textual analysis literature.

References

- Aiyer, B., Caso, J., Russell, P., & Sorel, M. (2022). *New survey reveals \$2 trillion market opportunity for cybersecurity technology and service providers*. . <https://shorturl.at/mtFY6>
- Anderson, R. (2002). *Security in Open versus Closed Systems – The Dance of Boltzmann, Coase and Moore* Open Source Software : Economics, Law and Policy, Toulouse, France.
- Banker, R. D., Frost, T. S., & Tripathi, M. K. (2022). The Determinants of InformationWeek 500 Selection and Its Implications: A Textual Analysis Approach. *Journal of Information Systems*, 36(1), 81-109.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Brown, N. C., Crowley, R. M., & Willott, W. B. (2020). What Are You Saying? Using topic to Detect Financial Misreporting. *Journal of Accounting Research*, 58(1), 237-291.
- Chaduvula, S. C., Dachowicz, A., Atallah, M. J., & Panchal, J. H. (2018). Security in Cyber-Enabled Design and Manufacturing: A Survey. *Journal of Computing and Information Science in Engineering*, 18(4).
- Dechow, P. M. (1994). Accounting earnings and cash flows as measures of firm performance: The role of accounting accruals. *Journal of Accounting and Economics*, 18(1), 3-42.
- Dechow, P. M., Kothari, S. P., & L. Watts, R. (1998). The relation between earnings and cash flows. *Journal of Accounting and Economics*, 25(2), 133-168.
- Dickinson, V. (2011). Cash Flow Patterns as a Proxy for Firm Life Cycle. *The Accounting Review*, 86(6), 1969-1994.
- Dong, W., Liao, S., & Zhang, Z. (2018). Leveraging Financial Social Media Data for Corporate Fraud Detection. *Journal of Management Information Systems*, 35(2), 461-487.
- Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics*, 64(2), 221-245.
- Eaton, T. V., Grenier, J. H., & Layman, D. (2019). Accounting and Cybersecurity Risk Management. *Current Issues in Auditing*, 13(2), C1-C9.
- Florackis, C., Louca, C., Michael, R., & Weber, M. (2022). Cybersecurity Risk. *The Review of Financial Studies*, 36(1), 351-407.
- Gordon, L. A., & Smith, R. (2007). Incentives for improving cybersecurity in the private sector: A cost-benefit perspective. *Congressional Testimony*.
- Gul, F. A. (1999). Growth opportunities, capital structure and dividend policies in Japan. *Journal of Corporate Finance*, 5(2), 141-168.

- Haislip, J., Lim, J.-H., & Pinsker, R. (2021). The Impact of Executives' IT Expertise on Reported Data Security Breaches. *Information Systems Research*, 32(2), 318-334.
- Hannon, L. (2002). Criminal Opportunity Theory and the Relationship Between Poverty and Property Crime. *Sociological Spectrum*, 22(3), 363-381.
- He, C. Z., Frost, T., & Pinsker, R. E. (2020). The Impact of Reported Cybersecurity Breaches on Firm Innovation. *Journal of Information Systems*, 34(2), 187-209.
- Heiligenstien, M. (2023). *Amazon Data Breaches: Full Timeline*. <https://firewalltimes.com/amazon-data-breach-timeline/>
- Higgs, J., Pinsker, R., Smith, T., & Young, G. (2016). The Relationship between Board-Level Technology Committees and Reported Security Breaches. *Journal of Information Systems*, 30(3), 79-98.
- Huang, A. H., Lehavvy, R., Zang, A. Y., & Zheng, R. (2018). Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach. *Management Science*, 64(6), 2833-2855.
- Huang, H. H., & Wang, C. (2021). Do Banks Price Firms' Data Breaches? *The Accounting Review*, 96(3), 261-286.
- Islam, M. S., Farah, N., & Stafford, T. F. (2018). Factors associated with security/cybersecurity audit by internal audit function: An international study. *Managerial Auditing Journal*.
- Khan, M., & Watts, R. L. (2009). Estimation and empirical properties of a firm-year measure of accounting conservatism. *Journal of Accounting and Economics*, 48(2), 132-150.
- Kocher, P., Lee, R., McGraw, G., Raghunathan, A., & Ravi, S. (2004, 7-11 July 2004). Security as a new dimension in embedded system design. Proceedings. 41st Design Automation Conference, 2004.,
- Kwon, J., Ulmer, J. R., & Wang, T. (2013). The Association between Top Management Involvement and Compensation and Information Security Breaches. *Journal of Information Systems*, 27(1), 219-236.
- La Torre, M., Botes, V. L., Dumay, J., Rea, M. A., & Odendaal, E. (2018). The fall and rise of intellectual capital accounting: new prospects from the big data revolution. *Meditari accountancy research*.
- Leggett, T. (2021). *Amazon hit with \$886m fine for alleged data law breach*. BBC. <https://www.bbc.com/news/business-58024116>
- Li, F. (2010). The Information Content of Forward-Looking Statements in Corporate Filings—A Naïve Bayesian Machine Learning Approach. *Journal of Accounting Research*, 48(5), 1049-1102.
- Li, T., & Walton, S. (2023). Business Strategy and Cybersecurity Breaches. *Journal of Information Systems*, 1-26.
- Li, W. W., Leung, A. C. M., & Yue, W. T. (2023). Where is IT in Information Security? The Interrelationship among IT Investment, Security Awareness, and Data Breaches [Article]. *Mis Quarterly*, 47(1), 317-342.
- Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35-65.
- Margsiri, W., Mello, A. S., & Ruckes, M. E. (2008). A Dynamic Analysis of Growth via Acquisition. *Review of Finance*, 12(4), 635-671.
- Myers, S. (1977). Determinants of corporate borrowing. *Journal of Financial Economics*, 5(2), 147-175.
- Nelson, N., & Madnick, S. E. (2017). Studying the Tension Between Digital Innovation and Cybersecurity. Americas Conference on Information Systems,
- Okafor, R. C. (2021). *Cybersecurity Due Diligence in Mergers & Acquisitions Transactions*. SSRN. <https://ssrn.com/abstract=3915861>
- Ponemon and IBM. (2021). *Cost of a data breach report 2021*. <https://www.ibm.com/downloads/cas/ojdvqgry>
- Richardson, V., Smith, R., & Watson, M. (2019). Much Ado about Nothing: The (Lack of) Economic Impact of Data Privacy Breaches. *Journal of Information Systems*, 33(3), 227-265.
- Sen, R., & Borle, S. (2015). Estimating the Contextual Risk of Data Breach: An Empirical Approach. *Journal of Management Information Systems*, 32(2), 314-341.
- Shipman, J. E., Swanquist, Q. T., & Whited, R. L. (2017). Propensity Score Matching in Accounting Research. *The Accounting Review*, 92(1), 213-244.
- Smith, C. W., & Watts, R. L. (1992). The investment opportunity set and corporate financing, dividend, and compensation policies. *Journal of Financial Economics*, 32(3), 263-292.
- Walton, S., Wheeler, P. R., Zhang, Y., & Zhao, X. (2021). An Integrative Review and Analysis of Cybersecurity Research: Current State and Future Directions. *Journal of Information Systems*, 35(1), 155-186.
- Xu, H., Guo, S., Haislip, J. Z., & Pinsker, R. E. (2019). Earnings Management in Firms with Data Security Breaches. *Journal of Information Systems*, 33(3), 267-284.
- Zhuang, Y., Choi, Y., He, S., Leung, A. C. M., Lee, G. M., & Whinston, A. (2020). Understanding Security Vulnerability Awareness, Firm Incentives, and ICT Development in Pan-Asia. *Journal of Management Information Systems*, 37(3), 668-693.

Table 1: Variable Definitions

Variable	Description
Breach	The natural log of the number of hacks (<i>Data-Hack</i>), denial-of-service attacks (<i>DoS-Attack</i>), or data breaches (<i>Data-Breach</i>)
Growth Opportunity	Growth opportunity measured as an industry-year peer-firm mean-adjusted principal component of (1) the output from the LDA topic model derived from S&P business descriptions, (2) the sentiment of business descriptions as measured by the Harvard General Inquirer and Diction word lists, and (3) the uncertainty in business descriptions as measured by Loughran and McDonald (2011) financial uncertainty list.
Investment Cycle	The firm-year investment cycle is estimated as the fitted value of the following regression: $\text{OperatingCF}_t = c1 * \text{CapEx}_t + c2 * \text{CapEx}_{t-1} + c3 * \text{CapEx}_{t+1} + c4 * \text{AcqEx}_t + c5 * \text{AcqEx}_{t-1} + c6 * \text{AcqEx}_{t+1} + \text{intercept} + \epsilon,$ where, the OperatingCF is cash flows from operating activities (OANCF) scaled by total assets (AT) in prior year ($t-1$), CapEx is capital expenditures (CAPX) scaled by total assets (AT) in prior year ($t-1$), and AcqEx is the sum of all types of acquisition expenditures (ACQINVT+ ACQPPE+ ACQLNTAL+ ACQINTAN+ ACQGDWL+ ACQAO) scaled by total assets (AT) in prior year ($t-1$).
ITAwareness	An indicator variable equals to one if the yearly executive job title is identified as “chief information officer,” “CIO,” “chief technology officer,” or “CTO”, else the value is zero. The information is obtained from ExecuComp for the yearly firm-executive level estimations.
ROA	Earnings before extraordinary items (IB) divided by total assets (AT).
Dividends	Total common and preferred dividends (DVT) scaled by total assets (AT).
Foreign Operations	An Indicator variable which equals to 1 for firms operating in foreign or multinational countries with a non-zero foreign currency adjustments (FCA), else the value is 0.
Cash Holdings	Cash and cash equivalents (CHE) scaled by total assets (AT).
Liquidity	Total current assets (ACT) less total current liabilities (LCT) scaled by total assets (AT).
Employees	Number of Employees (EMP) scaled by total assets (AT).
PP&E	Net plant property and equipment (PPNET) scaled by total assets (AT).
Intangibles	Total intangible assets (INTAN) scaled by total assets (AT).
Leverage	The ratio of total liabilities (LT) divided by total assets (AT).
BTM	The ratio of book value of equity (CEQ) divided by market value of equity (PRCC_F*CSHO).
Size	The natural logarithm of total assets (AT).

Table 2 – Panel A: Sample Selection

Sample Procedure	Observation (firm-year)
Total firm-year observations for firms’ financials obtained from Compustat North America for 19 years for calendar years from 2001-2019 fiscal-end and 17,907 unique firms.	151,513
Merge: Sample of data hacks for calendar years from 2004-2020 and 544 unique publicly listed firms, those with a non-missing CIK value.	
Merge: Sample for denial-of-service attacks for calendar years from 2011-2019 and 60 unique publicly listed firms, those with a non-missing CIK value.	
Less: firm-year observations with missing values for the growth opportunity variables.	66,559
Less: firms without necessary data to calculate control variables.	15,003
Pre-Matched Sample: firm-year observations for 19 years from 2001-2019 fiscal-end and 7,184 unique firms, where 336 of such unique publicly listed firms are subject to either data hacks (330 firms), denial-of-service attacks (13 firms), or both (6 firms).	69,951
Matched Sample: firm-year observations for 19 years from 2001-2019 fiscal-end and 672 unique firms, including 336 breached and control firms, respectively.	9,150
Less: firm-year with missing values for the prior period ($t-1$)	828
Analysis sample: Total firm-year observations for the regression estimation, including 18 years from 2002-2019 fiscal-end with 664 unique publicly listed firms of which 336 breached and 328 control firms remain, having at least (min) a period of two years before a data breach.	8,322

Table 2 – Panel B: Descriptive Statistics

Variable	Mean	Std.Dev	Min	Q1 (p25)	Median (p50)	Q3 (p75)	Max
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Data-Hack</i>	0.220	0.584	0.000	0.000	0.000	0.000	7.000
<i>DoS-Attack</i>	0.014	0.194	0.000	0.000	0.000	0.000	5.000
<i>Datar-Breach</i>	0.234	0.660	0.000	0.000	0.000	0.000	11.000
<i>Growth-Opportunity</i>	0.000	0.433	-1.370	-0.192	-0.076	0.066	4.427
<i>ROA</i>	0.034	0.201	-6.461	0.016	0.051	0.089	0.354
<i>Dividends</i>	0.018	0.033	0.000	0.000	0.006	0.024	0.382
<i>Foreign Operations</i>	0.361	0.480	0.000	0.000	0.000	1.000	1.000
<i>Cash Holdings</i>	0.160	0.165	0.000	0.041	0.100	0.220	0.852
<i>Liquidity</i>	0.157	0.232	-5.148	0.026	0.128	0.273	0.847
<i>Employees</i>	0.006	0.009	0.000	0.001	0.003	0.006	0.086
<i>PP&E</i>	0.236	0.204	0.0002	0.081	0.171	0.329	0.918
<i>Intangibles</i>	0.256	0.222	0.000	0.059	0.199	0.405	0.812
<i>Leverage</i>	0.600	0.343	0.045	0.421	0.581	0.722	8.745
<i>Book-to-Market</i>	0.378	1.016	-23.342	0.194	0.351	0.595	9.739
<i>Size</i>	8.248	1.984	0.249	6.908	8.383	9.757	12.041
<i>Investment Cycle</i>	0.116	0.055	-0.495	0.090	0.102	0.126	2.316

Table 2 – Panel C: Propensity Score Matched Covariate Tabulations

Variables	Breached: No	Breached: Yes	Standardized Mean Difference	p-value
	(1)	(2)	(3)	(4)
Mean (Std. Dev.)				
Firms N=	336	336		
<i>ROA</i>	0.011 (0.163)	0.026 (0.217)	0.081	0.293
<i>Dividends</i>	0.017 (0.025)	0.016 (0.023)	0.020	0.798
<i>Foreign Operations</i>	0.330 (0.418)	0.352 (0.392)	0.053	0.490
<i>Cash Holdings</i>	0.167 (0.172)	0.174 (0.153)	0.038	0.621
<i>Liquidity</i>	0.142 (0.281)	0.159 (0.212)	0.069	0.374
<i>Employees</i>	0.006 (0.010)	0.006 (0.008)	0.006	0.941
<i>PP&E</i>	0.223 (0.190)	0.237 (0.202)	0.071	0.358
<i>Intangibles</i>	0.261 (0.207)	0.258 (0.205)	0.015	0.845
<i>Leverage</i>	0.622 (0.377)	0.601 (0.338)	0.060	0.434
<i>Book-to-Market</i>	0.375 (0.928)	0.382 (0.350)	0.010	0.895
<i>Size</i>	8.022 (1.982)	8.066 (1.916)	0.022	0.772

Table 3: Growth Opportunity and Likelihood of Data Breach

Dependent variable:	<i>lnData-Breach</i> (t)	<i>lnData-Hack</i> (t)	<i>lnDoS-Attack</i> (t)
Coef. (t-stat.)	(1)	(2)	(3)
<i>Growth-Opportunity</i> _{j,t-1}	0.066 *** (4.444)	0.064 *** (4.179)	0.006 *** (4.535)
<i>ROA</i> _{j,t}	0.083*** (3.553)	0.076*** (3.491)	0.013*** (3.703)
<i>Dividends</i> _{j,t}	0.128(0.898)	0.182(1.296)	-0.107*(-1.822)
<i>Foreign</i> _{j,t}	0.002(0.210)	0.003(0.423)	0.003(0.628)
<i>Cash</i> _{j,t}	-0.097***(-3.077)	-0.110***(-3.750)	0.025(1.446)
<i>Liquidity</i> _{j,t}	0.036(1.661)	0.041*(1.861)	-0.000(-0.033)
<i>Employees</i> _{j,t}	-1.725(-1.572)	-2.085*(-1.855)	0.687*** (5.327)
<i>PP&E</i> _{j,t}	-0.078(-1.519)	-0.089*(-1.783)	0.033** (2.410)
<i>Intangibles</i> _{j,t}	-0.276***(-7.218)	-0.270***(-7.210)	-0.005(-0.488)
<i>Leverage</i> _{j,t}	0.093*** (3.808)	0.085*** (3.552)	0.015*** (4.871)
<i>Book-to-Market</i> _{j,t}	0.007*** (2.917)	0.007*** (2.901)	0.000(1.258)
<i>Size</i> _{j,t}	0.037*** (3.991)	0.036*** (4.010)	0.003(1.442)
Fixed Effects	Firm and Year		
Standard Errors	Robust-cluster by Year		
Observations (N)	8,322	8,322	8,322
Adjusted R ²	0.584	0.584	0.394

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4: Data Breach and Investment Cycle in Leading Years

Dependent var:	<i>Invest-Cycle (t+1)</i>			<i>Invest-Cycle (t+2)</i>		
	<i>Average: Baseline</i>	<i>Growth-Opportunity_{j,t-1}:</i>		<i>Average: Baseline</i>	<i>Growth-Opportunity_{j,t-1}:</i>	
		<i>Low</i>	<i>High</i>		<i>Low</i>	<i>High</i>
Coef. (t-stat.)	(1)	(2)	(3)	(4)	(5)	(6)
lnData-Breach_{j,t}	0.003* (1.899)	0.008** (2.524)	-0.005** (-2.311)	0.004** (2.193)	0.011*** (3.632)	-0.006** (-2.269)
<i>ROA_{j,t}</i>	0.019*(2.064)	0.001(0.068)	0.035***(2.751)	0.003(0.238)	-0.019(-0.458)	0.013(1.398)
<i>Dividends_{j,t}</i>	0.010(0.356)	0.015(0.342)	-0.004(-0.072)	0.070***(2.289)	0.059(1.325)	0.073(1.734)
<i>Foreign_{j,t}</i>	-0.000(-0.101)	0.001(0.254)	-0.002(-0.839)	0.004***(2.913)	0.003(1.334)	0.005***(2.186)
<i>Cash_{j,t}</i>	0.039*(2.091)	0.082(1.636)	0.022(1.192)	0.022(1.167)	0.049(1.407)	0.001(0.043)
<i>Liquidity_{j,t}</i>	-0.035(-0.898)	-0.094(-0.908)	-0.001(-0.128)	-0.032(-0.949)	-0.074(-0.905)	-0.004(-0.592)
<i>Employees_{j,t}</i>	2.319***(2.562)	3.126***(2.641)	1.691***(-5.727)	1.789***(2.639)	2.659(1.274)	1.442***(-3.944)
<i>PP&E_{j,t}</i>	0.018(0.418)	-0.005(-0.075)	0.039(1.562)	-0.035(-0.849)	-0.068(-0.712)	-0.017(-0.704)
<i>Intangibles_{j,t}</i>	0.009(0.667)	0.001(0.051)	0.014(0.692)	0.018(1.171)	0.015(0.515)	0.016(0.993)
<i>Leverage_{j,t}</i>	-0.002(-0.257)	0.002(0.154)	-0.001(-0.247)	-0.003(-0.583)	-0.004(-0.235)	0.000(0.010)
<i>Book-to-Market_{j,t}</i>	-0.003(-1.529)	-0.002(-1.049)	-0.003**(-2.623)	-0.002(-1.364)	-0.001(-1.099)	-0.002(-1.140)
<i>Size_{j,t}</i>	-0.015***(-3.139)	-0.019*(-1.932)	-0.013***(-3.663)	-0.018***(-3.370)	-0.021**(-2.257)	-0.015***(-4.241)
Fixed Effects	Firm and Year					
Standard Errors	Robust-cluster by Year					
Observations (<i>N</i>)	7,014	3,477	3,537	6,379	3,147	3,232
Adjusted R ²	0.350	0.267	0.468	0.341	0.268	0.427

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5, Panel A: Mitigatory Effect of IT Awareness and Likelihood of Data Breaches

Dependent variable:	<i>lnData-Hack (t)</i>	<i>lnDoS-Attack (t)</i>	<i>lnData-Breach (t)</i>
Coef. (t-stat.)	(1)	(2)	(3)
<i>Growth-Opportunity_{j,t-1}</i>	0.092*** (5.635)	0.013*** (5.713)	0.097*** (5.998)
<i>ITAwareness_{j,t}</i>	0.019** (2.737)	0.005* (1.800)	0.019** (2.673)
<i>Growth-Opportunity_{j,t-1}*ITAwareness_{j,t}</i>	-0.082*** (-3.203)	-0.010** (-2.324)	-0.084*** (-3.130)
Controls	Included		
Fixed Effects	Firm and Year		
Standard Errors	Robust-cluster by Year		
Observations (<i>N</i>)	32,966	32,966	32,966
Adjusted R ²	0.629	0.443	0.629

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5, Panel B & C: Growth Opportunity and Data Breach for Firms in Shakeout and Decline Lifecycle Stages

	Panel B			Panel C		
Dependent variable:	<i>lnData-Hack_{j,t}</i>	<i>lnDenial-Attack_{j,t}</i>	<i>lnData-Breach_{j,t}</i>	<i>lnData-Hack_{j,t}</i>	<i>lnDenial-Attack_{j,t}</i>	<i>lnData-Breach_{j,t}</i>
Coef. (t-stat.)	(1)	(2)	(3)	(4)	(5)	(6)
<i>Growth-Opportunity_{j,t-1}</i>	0.073*** (4.408)	0.008** (2.036)	0.075*** (4.456)	0.037* (1.689)	-0.000 (-0.080)	0.037* (1.679)
Controls	Included					
Fixed Effects	Firm and Year					
Standard Errors	Robust-cluster by					
	Year	Year	Year	Year	Year	Year
Observations (<i>N</i>)	5,573	5,573	5,573	2,749	2,749	2,749
Adjusted R ²	0.569	0.365	0.566	0.621	0.414	0.619

Note: *p<0.1; **p<0.05; ***p<0.01.