12-12-2018

Breaching Together: A Data Science Approach on Firms' Correlated Risk in Information Security

Rahul Dwivedi
Texas A&M University - Central Texas, rahul.dwivedi@tamuct.edu

Sridhar P. Nerur
University of Texas at Arlington, snerur@uta.edu

Jingguo Wang
University of Texas at Arlington, jwang@uta.edu

Breaching Together: A Data Science Approach on Firms' Correlated Risk in Information Security

Research-in-Progress

Rahul Dwivedi
Texas A&M University – Central Texas
rahul.dwivedi@tamuct.edu

Sridhar Nerur
University of Texas – Arlington
snerur@uta.edu

Jingguo Wang
University of Texas – Arlington
jwang@uta.edu

Abstract

This study develops a data science approach to measuring business relatedness of firms, with a view to assessing how this measure (of business proximity) is associated with correlated risk of firms experiencing information security breaches. We analyze textual business descriptions and security risk factors from SEC 10-K filing reports of 33 public firms that were breached at the same time in the last 10 years (2008 – 2017). Specifically, we use text analysis and topic modeling to come up with a measure of breach proximity. The Quadratic Assignment Procedure (QAP), a well-known technique in social network analysis, was used to test for significance of statistical relationships among the various similarity matrices. In preliminary investigations, we found that dyadic relationships between public firms based on their business descriptions and security risk factors from their 10-K filings is significantly correlated with the dyads based on information security breaches for these public firms. We also found geographic proximity and industry type based on two digits SIC code for industry classification to be significantly correlated with the propensity of firms to be breached together.

Keywords

Information Security Breach, Breaching Together, Network Analysis, QAP, Correlated Risks, Topic Modeling

Introduction

Information security breach, also known as data breach, can be defined as unauthorized access or acquisition of data (computerized or not) that compromises "the security, confidentiality or integrity" of proprietary or personal information maintained by a person or an organization (Faulkner, 2007). Theft of disk or portable device with classified data, consumer data obtained by hackers, and theft of proprietary information by insiders are examples of information security breaches. As per one public source, 8064 data breaches have been made public since 2005, resulting in loss of more than 10 billion user records[1]. Although businesses have increased their annual security spending[2,3] many still suffer from heavy financial loss due to security breaches. For example, as per a recent IBM study, the average total cost of data breach across the globe is $3.62 million for the year 2017, and the average cost per lost record is $141[4]. Another recent security breach at a credit reporting agency exposed sensitive personal information of about 143 million US consumers[5].

[1] https://www.privacyrights.org/data-breaches
[2] https://www.gartner.com/newsroom/id/3836563
[3] https://www.sans.org/reading-room/whitepapers/analyst/security-spending-trends-36697
[4] https://www.ibm.com/security/data-breach
[5] https://www.consumer.ftc.gov/blog/2017/09/equifax-data-breach-what-do

Our work focusses on an empirical investigation of correlated risks and failures in information security at the firm level. Cybersecurity risk can be defined as the risk arising from malicious electronic or non-electronic events affecting information technology resources of firms, often resulting in disruption of business and financial loss (Biener, Eling, & Wirfs, 2015; Mukhopadhyay, Chatterjee, Saha, Mahanti, & Sadhukhan, 2013). From a technological standpoint, firms often share correlated risks and vulnerabilities of being breached together due to the usage of common security technologies and connectivity of computer networks (Chen, Kataria, & Krishnan, 2011; Öğüt, Raghunathan, & Menon, 2011). The role of correlated risks have been widely investigated by the cybersecurity insurance community (Baer & Parkinson, 2007; Böhme & Kataria, 2006; Böhme & Schwartz, 2010; Mukhopadhyay et al., 2013). Historically, correlated risks and failures are investigated either within a firm among multiple systems on its own internal networks or across firms on their respective external networks (Chen et al., 2011; Mukhopadhyay et al., 2013).

In this study, we examine the likelihood of the organizations to have data breach on the same day. We empirically investigate the relationship between relatedness of firms based on their product descriptions and security risk factors as expressed in their 10-K filings and the likelihood of breaching together. The research question we are investigating in this study stems from the reasoning that business proximity between firms, as determined by relatedness of certain organizational attributes, is a likely predictor of their being concurrently breached. In other words, are there any underlying firm characteristics or attributes whose similarity makes firms to have correlated risk of breaching?

Business proximity may be defined as the relatedness of businesses in terms of products, market in which they operate, or the underlying technology used (Shi, Lee, & Whinston, 2015). While there is prior research on shared vulnerabilities ensuing from interconnected computer networks and homogenous software stacks, there is little, or no empirical investigation of how inter-firm relatedness based on certain attributes might affect correlated failures. Our research aims to fill this void.

Most previous studies on security breaches were confined to studying breaches either at the individual level or at the organizational level. For instance, some studies examined effect of security breaches on firm's financial performance (Acquisti, Friedman, & Telang, 2006; Avery & Ranganathan, 2016; Campbell, Gordon, Loeb, & Zhou, 2003; Ko & Dorantes, 2006), effect of compliance policies within organization on security breaches (Ernest Chang & Lin, 2007; Kraemer, Carayon, & Clem, 2009), and economics of information security investments (Gordon & Loeb, 2002; Huang, Hu, & Behara, 2006). Two recent studies investigated the relationship between security investments and breaches for the healthcare industry (Angst, Block, D'arcy, & Kelley, 2017; Kwon & Johnson, 2014). To the best of our knowledge none of the previous studies have attempted to investigate security breaches at the dyadic levels i.e., how the ties between firms based on similarity impact their being breached together.

The remainder of this paper is organized as follows. The next section summarizes some of the previous literature pertinent to our study. The subsequent section provides the rationale for the formulation of our hypothesis. Next, we discuss the procedures used to collect our data, followed by a discussion of our results.

Background Literature and Variables Used

In this exploratory study, we investigate whether firms that are similar in terms of their business descriptions and security risk factors from the 10-K filings are more likely to be breached together. Furthermore, we also explore whether simultaneous breaches are correlated with geographic proximity based on headquartered states for these firms. The next sub sections summarize relevant literature and explains the variables used in this study.

Correlated Risk in Information Security

In a significant work on correlated failures arising as a result of software vulnerabilities shared across organizations, (Chen et al., 2011) propose queuing models for quantifying downtime loss as a function of investment in security technologies, software diversification, and IT resource investments. They further model and analyze the effectiveness of software diversification strategy to deal with correlated failures from different cost benefit perspectives. (Kunreuther & Heal, 2003) develop game-theoretic models addressing the problem of interdependent security where all agents are identical for different real-life scenarios such as airline security, fire protection, vaccinations, and protection against theft and bankruptcy.

Correlated security risks are widely investigated in domain of cyber insurance literature as an effective mechanism for minimizing and managing cyber security incidents. Businesses routinely must manage adverse events. In the context of cyber security, breaches are such adverse events which need to be managed. With rapid growth in Internet, e-commerce and usage of software, along with widespread financial losses due to viruses and breach activities, few insurance companies developed specialized cyber insurance policies in the late 19th and early 20th century (Baer & Parkinson, 2007). Cyberattacks and information security breaches often exploit shared vulnerabilities across interconnected networks resulting in interdependent security risks and hence hindering the growth of cyber insurance market (R. Anderson & Moore, 2009; Ross Anderson & Moore, 2007; Böhme & Kataria, 2006). (Biener et al., 2015) in their empirical analysis comparing cyber risks with operational risks also pointed out the difficulty of insurability of cyber risks because of interconnected nature of computer networks and information systems. (Mukhopadhyay et al., 2013) propose models aimed at evaluating the utility of cyber-insurance products based on the concepts of collective risk modeling theory (Hossack, Pollard, & Zehnwirth, 1999) and argue that there are many benefits of getting cyber-insurance from the perspective of financial trade-offs. (Böhme & Kataria, 2006; Böhme & Schwartz, 2010; Öğüt et al., 2011) provides various frameworks towards modeling correlated risks in the context of cyber-insurance.

Dependent Variable: Breach Proximity

We define breach proximity or breach relatedness as the likelihood of two firms or businesses being breached together. As mentioned previously, although there is some research on correlated security failures, interdependent cyber security, and correlated risks from the technological and cyber insurance perspective, almost all of them followed a game-theoretic modeling approach. Our research adopts a unique data science perspective to empirically test the phenomenon of correlated failures and concurrent breaching using data from the real world. Also, our aim is to explore some of the underlying firm-level antecedents contributing to breach proximity, rather than technology-based variables such as shared vulnerability across software or interconnected computer networks.

Business Similarity

Business similarity or relatedness of businesses have been used as an antecedent in studies on mergers & acquisitions (Shi et al., 2015; L. Wang & Zajac, 2007) and alliance formation (Stuart, 1998). The underlying argument is based on the idea that businesses which are similar in terms of the product, market or technological space, can achieve business synergy easily and hence have higher probability of being successful when merged or become partners compared to dissimilar businesses or firms. There have been a few studies in the past that have looked at similarities of businesses based on their descriptions. For instance, (Shi et al., 2015) came up with a measure of dyadic business proximity for technological firms based on their business descriptions.

Independent Variable: Business Similarity based on Business Descriptions

The United States federal law requires public firms to disclose financial information in the form of various reports on an ongoing basis. Examples include quarterly reports (form 10-Q) and annual reports (form 10-K). As per the SEC website[6], annual report on form 10-K is different from annual report to shareholders and provides a comprehensive overview of the business with its financial condition. Since, the 10-K financial report includes details about the public firms operating in US, they can be used as a measure of business similarity.

(Hoberg & Phillips, 2016) identified related firms based on the business descriptions sections of their 10-K filings. As per the SEC mandates, the business descriptions section of the 10-K filings include significant products offerings by businesses and hence firms offering similar products can be grouped together based on these filings. Analogous to the existing industry classification schemes for public firms such as SIC and NAICS, (Hoberg & Phillips, 2016) proposed classification of industries based on business descriptions from 10-K filings as Text-based network industry classification. (Hoberg & Phillips, 2010) found transactions based on mergers & acquisitions between firms that use similar product descriptions in their 10-K filings to

[6] https://www.sec.gov/fast-answers/answers-form10khtm.html

be more alike than between firms that are dissimilar in terms of their product offerings. Based on these works by (Hoberg & Phillips, 2010, 2016), we argue that 10-K filings from firms can effectively be used as a measure of industry classification with similar firms having similar textual content in these filings. To the best of our knowledge, no empirical study has used this measure of firm similarity - based on textual content from 10-K filings - in the context of information security breaches. Given this backdrop, we contend that business proximity between firms has the potential to shed light on whether they run a higher risk of being breached at the same time.

Independent Variable: Business relatedness based on security risk factors

Public firms often disclose security risk factors associated with their information systems resources in their 10-K public filings. For example, one of the firms included in this research is Automatic Data Processing (ADP), whose security risk factors section from ADP's 10-K filings states that "cybersecurity and privacy breaches may hurt our business, damage our reputation, increase our costs, and cause losses." Similarly, one of Twitter's security risk factor is, "We are unable to combat spam or other hostile or inappropriate usage on our platform." In a significant work on a firm's security risk factor as stated in its 10-K filing and future disclosure of breach announcement by public firms, the decision tree based model proposed by (T. Wang, Kannan, & Ulmer, 2013) associated disclosure of security risk factors with future breach announcements. Our research is similar to that carried out by (T. Wang et al., 2013), in that both studies associate disclosures of security risk factors from 10-K filings with future breach announcements. However, the research proposed in this article is unique because we consider the relationship between security risk factors and a firm's public disclosure of breaches from the social network perspective of firms being breached together. The unit of analysis in this study is not the individual firm but the dyadic relationship between two or more. Thus, we propose that firms that firms with similar security risk factors as disclosed in their public filings are more likely to be the victims of the same breach on the same day (i.e. breaching together) in future in comparison with firms with dissimilar security risk factors. Hence, we extracted security risk factors from 10-K filed in the previous year as when breach is declared publicly.

Apart from these independent variables, we also used other firm level attributes such as type of industry based on two-digit industry classification SIC codes and similarity of businesses based on their geographic proximity (headquartered in same US state and region) as a measure of firm relatedness. Table 1 shows the list of independent variables along with their usage in literature.

Independent variable	Explanation (data source)	Used in Previous literature
Firm descriptions from 10K filings	Description section in 10K include information about major product offerings (SEC EDGAR system).	(Hoberg & Phillips, 2010, 2016)
Security risk factors from 10K filings	Risk factors in 10K include information about major security risks	(T. Wang et al., 2013)
Headquartered state	As a measure of geographic proximity	(Shi et al., 2015)
Industry type	Type of industry indicated by first two digits of SIC code for industry classification (COMPUSTAT).	(L. Wang & Zajac, 2007; T. Wang et al., 2013)

Table 1. Independent variables (firm level attributes as measures of business relatedness)

Data Collection

The information security breach data for the firms that have been breached together on the same day in the past 10 years (2008 – 2017) is collected from Privacy Rights Clearinghouse dataset. Specifically, the data includes the date the breach was made public, the victim firm(s), location of breach along with type of

breach, and a short description[7]. The breach dataset from Privacy Rights Clearinghouse is one of the most popular publicly available sources of breach information for information security researchers (Avery & Ranganathan, 2016; Kwon & Johnson, 2015; Sen & Borle, 2015). The breach data was collected in July 2017.

The breach dataset also provides us with a short breach description which suggests that some of the firms in our dataset have a client-provider relationship (for example ADP providing payroll services to US airways) while others belonging to similar industry might have collaborative/partnership (such as HP and Symantec with the latter acting as information security vendor to the former) or competitive relationships (such as Bank of America and Citi Group Inc). As argued by (Chen et al., 2011), for partner firms, usage of homogenous software brings many advantages in the form of positive network effects, such as increased compatibility and interoperability.

The SEC 10-K filings about these breached firms are collected from the US Securities and Exchange Commission's (SEC) Electronic Data Gathering, Analysis and Retrieval (EDGAR) system. The resulting network consists of 33 publicly listed firms which have been breached together between the years 2008 and 2017. Other firm level independent variables such as industry type derived from first two-digits of industry classification code of SIC and headquartered state were also used.

Data Analysis and Results

As stated previously, QAP, a popular statistical technique employed by social network analysis researchers, was used to test for correlation and significance of association between the firms which are breached together and various business similarity measures. In statistical parlance, our model comprises dependent matrix of firms that have been breached together in the past 10 years and the independent matrices representing similarity of firms based on different firm characteristics such as industry type, headquartered state, and a matrix for cosine similarity based on business descriptions and security risk factors from SEC 10-K filings.

The dependent matrix is a binary matrix consisting of values 0 (not breached together) or 1 (breached together). Except for cosine similarity matrices of business descriptions and security risk factors derived from SEC 10-K filings, the other two independent matrices were also constructed in the same way, with a 1 indicating firm similarity and 0 otherwise. The values in independent cosine similarity matrix based on 10-K filings are all continuous and less than 1 with diagonal values ignored. To carry out the hypothesis's tests, these dependent and independent matrices are given as inputs to the QAP procedure of a well-known software for social network analysis called UCINET (Borgatti, Everett, & Freeman, 2002). Basic social network analysis is performed using statnet package (Handcock, Hunter, Butts, Goodreau, & Morris, 2008) for social network analysis in R statistical computing platform (Team & others, 2013) using RStudio (RStudio Team, 2015).

Breach network

Our breach network consists of 33 public firms that are affected by the same breach on the same day. These businesses are listed in the Appendix.

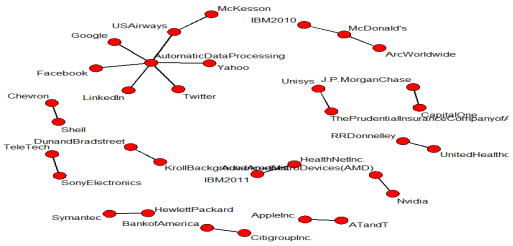

Figure 1. Breach Network of 33 Public Firms

[7] https://www.privacyrights.org/data-breaches

Note that there are two separate nodes for IBM, as there are two separate breaches in the years 2010 and 2011 and we have used the corresponding SEC 10-K filings from IBM for both the years for the respective breaches in our analysis.

Method of Analysis

We have used the MALLET (i.e., MAchine Learning for LanguagE Toolkit) implementation of the Latent Dirichlet Analysis (LDA) technique to perform topic modeling over our data set of firm descriptions. The number of topics is chosen based on elbow method as used in cluster analysis (Kodinariya & Makwana, 2013). One of the outputs from the mallet program gives us the probability loadings of each of the individual firm descriptions on each topic. This output can then be utilized to create a cosine similarity matrix for the firm descriptions based on the argument that firms with similar probability distributions across topics would be more like each other than those with dissimilar probability loadings for those topics. The cosine similarity matrix we got from this step can be used as an input to the next step of statistical analysis using Quadratic Assignment Procedure (QAP).

Quadratic Analysis Procedure (QAP) was performed using UCINET, a social network analysis software. Within UCINET, we have used the multiple regression variant of QAP known as MR-QAP, where the dependent matrix is the breach network for the 33 public businesses and independent matrices are cosine similarity matrices from 10-K filings, geographic proximity matrix (0 for firms headquartered in different states and 1 for those headquartered in the same state) and matrix of business similarity based on SIC codes from COMPUSTAT (again 0 for firms with different two digits of SIC codes and 1 for exactly same SIC codes). We ran 5000 permutations for the MR-QAP analysis for each model using UCINET to obtain the given results. The analysis steps are summarized in Figure 2 and the results from MR-QAP analyses are in Table 3.

.

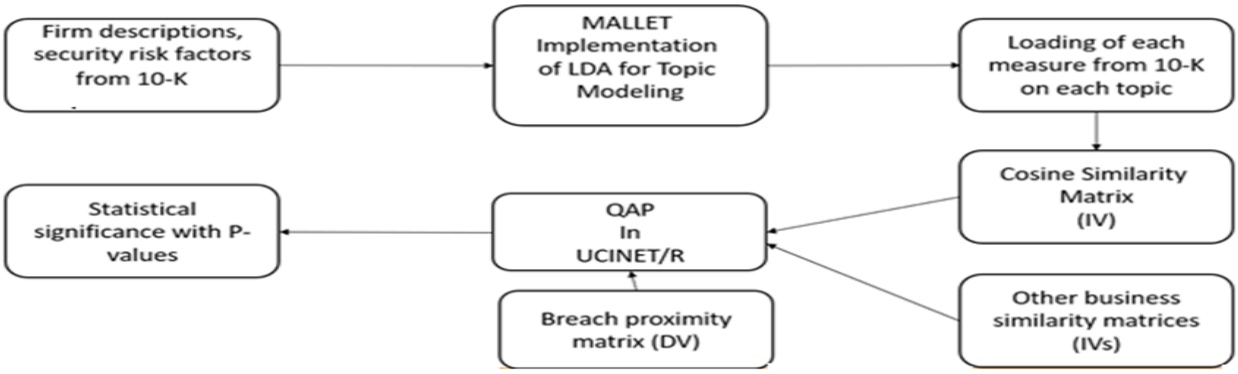

Figure 2. Steps carried out during analysis

QAP Results

Table 3. Results from MR-QAP analysis (p<0.1 *, p <0.05 **, p<0.01 *)** **Dependent variable: breach_proximity_matrix**		
	Model (1)	Model (2)
cosine_similarity_matrix_10K_description	0.02320**	
cosine_similarity_matrix_preprocessed_10K_description	0.00260***	0.01080**
Cosine_similarity_matrix_security_risk_factors	0.04979**	0.04479**
geographic_proximity_matrix_HQ_state	0.20076	0.08118*
industry_type_matrix_2_digit_SIC	0.00480***	0.03719**

Table 3. MR-QAP Results with corresponding p-values

Based on two variants of our MR-QAP models, we found that business relatedness based on cosine similarity matrix derived from business descriptions, similarity of businesses based on security risk factors derived from 10-K filings for public firms and type of industry based on the first two digits of SIC code are significantly correlated with breach proximity for the 33 firms under analyses. In future extension of this work, we propose to include business size (in terms of number of employees) and firm revenue as independent variables as well.

Discussion

This study employs a unique data science approach to exploring information security breaches at the organizational level. The dyadic and the social network viewpoint is also novel, as, to the best of our knowledge, no prior studies have used such techniques to analyze security breaches. We believe our study is an important first step towards understanding why firms are concurrently breached. We explore the question of whether firms that are breached together share some common attributes. This is an important research question that needs to be urgently addressed by cyber-security researchers. While breaches have been studied before, we believe our study is unique in its approach.

Contribution to Research

Our study demonstrates how specific text mining techniques of topic modeling and cosine similarity of firms based on their business descriptions can be applied to understand concurrent breaches. The increasing availability of textual data in both structured and unstructured form provides a unique opportunity to information systems researchers to use certain text analysis techniques to answer research questions in different contexts. In addition to contributing to methodology, our study offers new perspectives on how researchers may investigate the factors that impact cyber security breaches. Our study provides an example of how such data analytic methods can be applied to the context of correlated risks in information security. Our usage of certain firm level characteristics as antecedents contributing towards quantification of correlated risks paves way towards future research in this direction.

Furthermore, it is our hope that our application of a well-known social network analysis technique - Quadratic Assignment Procedure - for statistical analysis of network-based data will encourage other information systems researchers to use this under-utilized technique.

Implications for Practice

Cyber-security involves securing organizations against security breaches. It is evident from numerous surveys and studies that although businesses have been increasingly spending millions of dollars with each passing year on securing their information infrastructure, many of them invariably experience security breaches. From a practical viewpoint, we believe that firms can always learn from the failure of other firms. For instance, a business can take proactive measures in securing their resources, if a similar firm has been breached recently. Our study demonstrates that business proximity of a firm to another increases the likelihood of its being breached should the other one be infiltrated. Employees and managers responsible for safeguarding organizational resources as well as Chief Security Officers (CSO) should not only be upgrading their software for any potential vulnerabilities but must also remain informed about breaches affecting other businesses.

From the perspective cyber-insurance, our approach may be of practical importance to cyber risk insurer, responsible for insuring information infrastructure of multiple firms. It's possible that many of these firms may have similar portfolios or characteristics. Our approach may assist cyber risk insurer to not only effectively design insurance products for businesses based on their similarity but may also help in computing premiums based on breaches which may have recently affected similar organizations.

The social networking perspective that is used in our study can be very useful in understanding interdependent security vulnerabilities. Understanding the network characteristics will enable CSOs/security managers to anticipate the impact that any given node (i.e., firm/business) will have on the entire network. Furthermore, such an understanding will help them decide where to deploy scarce resources to mitigate the risks of cyber-attacks, or to stem the spread of, say, a virus through the network.

Limitations and Future Research

As pointed out previously, our study is a preliminary investigation of security breaches from the perspective of business relatedness. We believe that the question of business relatedness as an antecedent to breach proximity should be explored using different variables. Similarity of firms based on number of employees i.e. firm size and annual revenue, are important variables which need to be included in future extensions of these analyses. The use of similar security technologies (and hence having vulnerability proximity with each other to some extent) is an important antecedent to breach proximity of firms. Lack of availability of such data prevented us from using such variables in our analysis. Our study comprises only 33 public firms, and although some of the previous MR-QAP studies use similar or smaller network sizes (Coelho et al., 2015; Tsai & Ghoshal, 1998), we would like to replicate this study using larger sample size. Furthermore, we would like to extend our research to private firms as well.

Conclusion

In conclusion, this study provides an essential first step towards analyzing security breaches at the dyadic level for organizations that are similar on certain characteristics. Though exploratory in nature, our study shows how certain data science-based techniques, such as topic modeling and cosine similarity on textual contents, and statistical techniques such as QAP for networked data can be successfully applied in the context of security breaches. We found that for a sample of networked data comprising 33 public firms, business similarity of firms based on the business descriptions and security risk factors as declared in their 10-K filings is significantly correlated with their propensity to be breached together. Also, geographic proximity in terms of headquartered state and industry type to which business belong based on two digits of SIC code are significantly correlated with likelihood of being breached together. This study, with future extensions we have been working on, can help researchers and practitioners better understand information security breaches from this unique perspective of business relatedness or similarity.

Appendix

List of 33 public firms in our breach network

AT&T, Advanced Micro Devices, Apple Inc, Arc Worldwide, Automatic Data Processing, Bank of America, Capital One, Chevron, Citi Group Inc, Dun and Bradstreet, Facebook, Google, Health Net Inc, Hewlett Packard, IBM in 2010, IBM in 2011, J P Morgan Chase, Kroll Background America, LinkedIn, McDonald's, McKesson, Nvidia, RR Donnelley, Shell, Sony Electronics, Symantec, TeleTech, The Prudential Insurance Company of America, Twitter, US Airways, Unisys, United Healthcare and Yahoo.

References

Acquisti, A., Friedman, A., & Telang, R. (2006). Is there a cost to privacy breaches? An event study. *ICIS 2006 Proceedings*, 94.

Anderson, R., & Moore, T. (2009). Information security: where computer science, economics and psychology meet. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *367*(1898), 2717–2727. https://doi.org/10.1098/rsta.2009.0027

Anderson, Ross, & Moore, T. (2007). Information security economics–and beyond. In *Annual International Cryptology Conference* (pp. 68–91). Springer.

Angst, C. M., Block, E. S., D'arcy, J., & Kelley, K. (2017). When do IT security investments matter? Accounting for the influence of institutional factors in the context of healthcare data breaches. *MIS Quarterly*, *41*(3), 893–916.

Avery, A., & Ranganathan, C. (2016). Financial Performance Impacts of Information Security Breaches.

Baer, W. S., & Parkinson, A. (2007). Cyberinsurance in it security management. *IEEE Security & Privacy*, *5*(3).

Biener, C., Eling, M., & Wirfs, J. H. (2015). Insurability of cyber risk: An empirical analysis. *The Geneva Papers on Risk and Insurance-Issues and Practice*, *40*(1), 131–158.

Böhme, R., & Kataria, G. (2006). Models and Measures for Correlation in Cyber-Insurance. In *WEIS*.

Böhme, R., & Schwartz, G. (2010). Modeling Cyber-Insurance: Towards a Unifying Framework. In *WEIS*.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). Ucinet for Windows: Software for social network analysis.

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

Chen, Kataria, & Krishnan. (2011). Correlated Failures, Diversification, and Information Security Risk Management. *MIS Quarterly*, *35*(2), 397. https://doi.org/10.2307/23044049

Coelho, C. G., Falótico, T., Izar, P., Mannu, M., Resende, B. D., Siqueira, J. O., & Ottoni, E. B. (2015). Social learning strategies for nut-cracking by tufted capuchin monkeys (Sapajus spp.). *Animal Cognition*, *18*(4), 911–919. https://doi.org/10.1007/s10071-015-0861-5

Ernest Chang, S., & Lin, C. (2007). Exploring organizational culture for information security management. *Industrial Management & Data Systems*, *107*(3), 438–458. https://doi.org/10.1108/02635570710734316

Faulkner, B. (2007). Hacking into data breach notification laws. *Fla. L. Rev.*, *59*, 1097.

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

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, *24*(1), 1548.

Hoberg, G., & Phillips, G. (2010). Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *Review of Financial Studies*, *23*(10), 3773–3811. https://doi.org/10.1093/rfs/hhq053

Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, *124*(5), 1423–1465.

Hossack, I., Pollard, J. H., & Zehnwirth, B. (1999). *Introductory statistics with applications in general insurance*. Cambridge University Press.

Huang, C. D., Hu, Q., & Behara, R. S. (2006). Economics of Information Security Investment in the Case of Simultaneous Attacks. In *WEIS*. Citeseer.

Ko, M., & Dorantes, C. (2006). The impact of information security breaches on financial performance of the breached firms: an empirical investigation. *Journal of Information Technology Management*, *17*(2), 13–22.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal*, *1*(6), 90–95.

Kraemer, S., Carayon, P., & Clem, J. (2009). Human and organizational factors in computer and information security: Pathways to vulnerabilities. *Computers & Security*, *28*(7), 509–520. https://doi.org/10.1016/j.cose.2009.04.006

Kunreuther, H., & Heal, G. (2003). Interdependent security. *Journal of Risk and Uncertainty*, *26*(2–3), 231–249.

Kwon, J., & Johnson, M. E. (2014). PROACTIVE VS. REACTIVE SECURITY INVESTMENTS IN THE HEALTHCARE SECTOR. *Manage Inf Syst Q*, *38*(2), 451–471.

Kwon, J., & Johnson, M. E. (2015). The Market Effect of Healthcare Security: Do Patients Care about Data Breaches? In *WEIS*.

Mukhopadhyay, A., Chatterjee, S., Saha, D., Mahanti, A., & Sadhukhan, S. K. (2013). Cyber-risk decision models: To insure IT or not? *Decision Support Systems*, *56*, 11–26. https://doi.org/10.1016/j.dss.2013.04.004

Öğüt, H., Raghunathan, S., & Menon, N. (2011). Cyber Security Risk Management: Public Policy Implications of Correlated Risk, Imperfect Ability to Prove Loss, and Observability of Self-Protection: Cyber Security Risk Management. *Risk Analysis*, *31*(3), 497–512. https://doi.org/10.1111/j.1539-6924.2010.01478.x

RStudio Team. (2015). *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/

Sen, R., & Borle, S. (2015). Estimating the Contextual Risk of Data Breach: An Empirical Approach. *Journal of Management Information Systems*, *32*(2), 314–341. https://doi.org/10.1080/07421222.2015.1063315

Shi, Z., Lee, G. M., & Whinston, A. B. (2015). Toward a better measure of business proximity: Topic modeling for industry intelligence.

Stuart, T. E. (1998). Network Positions and Propensities to Collaborate: An Investigation of Strategic Alliance Formation in a High-Technology Industry. *Administrative Science Quarterly*, *43*(3), 668. https://doi.org/10.2307/2393679

Team, R. C., & others. (2013). R: A language and environment for statistical computing.

Tsai, W., & Ghoshal, S. (1998). SOCIAL CAPITAL AND VALUE CREATION: THE ROLE OF INTRAFIRM NETWORKS. *Academy of Management Journal*, *41*(4), 464–476. https://doi.org/10.2307/257085

Wang, L., & Zajac, E. J. (2007). Alliance or acquisition? a dyadic perspective on interfirm resource combinations. *Strategic Management Journal*, *28*(13), 1291–1317. https://doi.org/10.1002/smj.638

Wang, T., Kannan, K. N., & Ulmer, J. R. (2013). The Association Between the Disclosure and the Realization of Information Security Risk Factors. *Information Systems Research*, *24*(2), 201–218. https://doi.org/10.1287/isre.1120.0437