The Dissertation Committee for Yun-Sik Choi
certifies that this is the approved version of the following dissertation:

Towards Better Management of Organizational Cybersecurity

Committee:

Andrew B. Whinston, Supervisor

Gene Moo Lee

Gordon S. Novak Jr.

Risto Miikkulainen

Towards Better Management of Organizational Cybersecurity

by

Yun-Sik Choi

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2019

Dedicated to my loves.

Acknowledgments

First of all, I am indebted to my advisor, Andrew Whinston. I greatly appreciate for giving me continuous opportunities, new ideas, and amazing supports. My best research partner, teacher, and big brother Gene Moo Lee, huge thanks for your guidance and patience towards my doctoral degree. My co-author Jin Hyuk Choi, thanks for your trust and efforts for solving difficult problems together. I'd also like to thank my dissertation committee members Gordon Novak and Risto Miikkulainen for great supports.

I thank you so much for the great memories in the early years in Austin to the CS Avengers "Captain America" Jongwook Kim, "Thor" Eunho Yang, and "Hulk" Sung Ju Hwang. I also thank Sangki Yun and Donghyun Kim for relaxing coffee talks. Fun couple Gunwoong Kim and Hyeeun Kim, thanks for being a great company for various activities. Thanks Jiyoung Lee and Jisang Han for sharing so much with me. Thanks Jeho Oh, Wonjoon Goo and Donghyuk Kim for great comments on my research. WWE buddy Joseph, we had so much memories together. My precious friends from Stonybrook, Youngbum Kim, Jinwoong Hwang, and Youngbum Hur, thanks for being great friends for almost 10 years. And special thanks to my best friend Sunmin, I could not have done this without your continuous sacrifice and support.

My parents, Jung Kwon Choi and Boksil Lee, big thanks for your de-

votion, encouragements, love, and for helping me be myself. My sister Yun Jung, thanks for being the best sister. My first nephew Woojoo Yoo, thanks for the most beautiful smiles. Hyunbok, Hohyun, Hochul, Jiyoung, you guys are of course the most amazing cousins.

Lovely city Austin and amazing state Texas, thanks for being my second hometown. You gave me a lot, and I really enjoyed my 8 years here.

Towards Better Management of Organizational Cybersecurity

Publication No. _____

Yun-Sik Choi, Ph.D.
The University of Texas at Austin, 2019

Supervisor: Andrew B. Whinston

Cybersecurity poses a serious risk to organizations. To manage and improve organizational cybersecurity, one needs to have a technical comprehension of security threats along with an economic understanding of strategies employed by cyber attackers and defenders. In this dissertation, we take both empirical and theoretical approaches to deepen our understanding on the strategies of cybersecurity in three related chapters. First, we conduct an empirical analysis on publicly observed security incidents and developed an organizational security rating system. The rating is composed of botnet, spam, and phishing data from four data sources. By conducting a large-scale field experiment using the rating system, we find a causal relationship between security awareness and protection level. Second, we develop a game-theoretical model that characterizes a real-time dynamic interaction between an unidentified attacker and a defender in Internet Service Provider (ISP) level. Specifically, we propose a Bayesian Nash game in a network security setting. In this

game, a deceptive attacker tries to maximize its profit, and the defender tries to detect the attackers identity. Our equilibrium suggests that the strategic defense of ISP is necessary for the viability of an Internet-based society. Third, we develop a data-driven prediction model for security event detection. We construct a large composite dataset of externally observable organizational security posture and historical cyber incidents. In addition, we use LDA topic modeling on disclosed annual risk reports from organizations (Form 10-K Item 1A) to extract topic features. By leveraging these data, our model effectively predicts future security incidents.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Cybersecurity has become one of the critical issues to individuals, businesses, and governments. Researchers from academia and industry are continuously developing technical solutions, such as threat detection/prevention methods [19, 58], cryptography [20, 79], and access control [65, 81]. Despite these efforts, there are more organizations than ever falling behind on cybersecurity skills due to rapidly evolving new technologies and lack of proper training and investments.[1] While organizations want the highest level of security protection, at the same time, they have limited resources[2] to invest in securing and managing their systems. Even with the state-of-the-art security protection, persistent and sophisticated attackers may come up with new attack methods.[3] Considering the cost and benefits, we argue that organizations will strategically make security investment decisions only when the return on investment is higher than the associated cost. For these reasons, security problems cannot be solved without understanding the economic incentives of the

[1] https://blog.barracuda.com/2017/11/24/more-organizations-than-ever-falling-behind-on-cybersecurity-skills/

[2] https://www.forbes.com/sites/franksorrentino/2016/10/25/dont-let-a-lack-of-resources-compromise-your-cyber-security/

[3] https://www.cso.com.au/article/582065/perfect-security-setup/

organizations and stakeholders involved, both on the side of the attackers and defenders. In this thesis, we introduce three ideas to improve organizational cybersecurity with empirical and theoretical studies.

In the first chapter, we introduce a data-driven security evaluation framework using spam and phishing records, and empirically test how organizations respond to such information. We argue that spam emission can be considered as an indicator of improper internal control on botnet and malware infections as well as distorted incentives causing negative externality issues,[4] while phishing website hosting behavior can be attributed as a pure negative externality issue caused by lack of deterrence policy and responsibility.[5]

The main objective of this study is to find how organizations react in managing two distinct security issues, spam emission and phishing website hosting, when (1) they become aware of such problems and (2) the information is publicized. To achieve the research goal, we conduct a large-scale randomized field experiment on 1,262 organizations in six Pan-Asian countries.[6] We collected data from four reputable spam (CBL, PSBL) and phishing (APWG, Openphish) data sources, then developed a public security advisory website and an email treatment system. We send out advisory emails to the treatment group every 2 months, and ask them to visit our website for more detailed

[4] An externality is a positive or negative consequence (of an economic activity) experienced by unrelated third parties [80].

[5] Based on the Digital Millennium Copyright Act (DMCA) in the United States, ISPs and web hosts may not be found liable for content on a website if they meet certain requirements. https://www.copyright.gov/reports/studies/dmca/dmca_executive.html

[6] List of countries: China, Hong Kong, Singapore, Malaysia, Taiwan, Macau

information. We track the email opening and website visiting activities to measure if the target companies received the treatment. With rigorous econometric analysis, we find heterogeneous treatment effects depending on the characteristics of the security incidents. Organizations in the treatment group have reduced spam volume after opening the email or visiting the website; on the other hand, phishing volume did not change significantly.[7] This result indicates that (1) the security level can be measured with external data sources, (2) organizations react to such information based on the incentives, and (3) we need a stronger security policy on hosting malicious websites.

In the second chapter, we theoretically analyze the strategic interaction between an attacker and defender using a dynamic game model. Game theory has been widely used in a significant number of security studies, where strategic agents are taking actions to achieve an optimal outcome [31, 54, 64]. Game-theoretic analysis for security settings help to allocate limited resources, understand the underlying incentive mechanisms, and balance perceived risks [54].

We describe a continuous time cybersecurity game between a profit-maximizing attacker and an uninformed defender who stops the game based on the noisy observation. The equilibrium of the game characterizes the attacker's strategy of balancing the instantaneous profit and the game duration. In the equilibrium, the defender disconnects the counterpart when the updated

[7] Please see 2.4 for the detailed results.

suspicion level is above a threshold that is endogenously determined. Our analysis implies that strategic defense of Internet Service Providers (ISPs) is necessary for the viability of the Internet-based society.

Using this theoretical foundation, we propose a business model called managed security service with warranty (MSSW) which can be provided by Internet service providers (ISPs). In addition to the traditional Internet connectivity service and managed security service, the MSSW provides cyber insurance (warranty) service in case of security breaches. In other words, the provider will take responsibility for the protection service they provide in case of failure.[8] This model can be especially attractive to small and medium businesses (SMBs) with limited resources for in-house cybersecurity investment. Unlike large companies who can afford well-trained, in-house security professionals, SMBs do not have such resources, which ended up resulting in frequent cyberattacks.

In the last chapter, we develop data-driven prediction models for security event detection. While organizations can use their internal monitoring systems to predict their own cyber risks, it is hard for external entities such as investors to predict such security incidents of the firms of interest. In this context, our model can predict data breaches leveraging publicly available datasets without internal investigation. The dataset is collected from three different perspectives to observe various signs of insecurity. First, we collect

[8]There is no MSSP who provides compensation in case of data breach as of May 2018.

data on malicious activities such as spam, phishing, and botnet. These information can be used as proxies of defense level for organizations [40]. Next, we gather 10-K annual report documents from publicly traded U.S. firms. The documents are analyzed using LDA topic modeling to extract all IT related risks.[9] Lastly, we collected cybersecurity incident records such as Privacy Rights Clearinghouse (PRC)[10], VERIS Community Database (VCDB)[11] and Hackmageddon.[12] With these multidimensional data, we develop security breach prediction models with various classification algorithms including deep neural networks, kNN, and random forests. Our best performing neural network model marks an AUC score of 76.

[9]The term IT risk widely includes system performance related risks and network security related risks.

[10]https://www.privacyrights.org/

[11]http://veriscommunity.net/vcdb.html

[12]https://www.hackmageddon.com/

Chapter 2

Information Disclosure and Security Policy Design: A Large-Scale Randomization Experiment in Pan-Asia

2.1 Introduction

Cyberattacks are imposing serious threats to individuals, organizations, and our society at large. Even with technological advances in secure software and hardware, we are still experiencing an ever-increasing number of cyber-attacks. It is well known that this suboptimal situation in cyberspace is due partly to negative externalities, information asymmetry, and misaligned incentives [4]. This motivates us to explore more effective ways in which to enhance the security awareness of organizations and the public and create proper incentives with which to achieve secure cyber environments. In a recent U.S.-based field experiment, He et al. [40] showed that publicizing the security rankings of organizations against email spam may heighten such organizational awareness towards security issues. Given that cybersecurity is a global issue and each region has its distinct economic and societal environments, there is a need to extend the economics of cybersecurity literature by incorporating various international environments. Specifically, in this chapter, we focus on Pan-Asian countries which show significant economic development as well as rapid

adoption of technologies.

2.1.1 Cybercrime and Related Ordinances in Pan-Asian Countries

According to AIAs Landmark Healthy Living Survey, adults in Hong Kong spend an average of 3.7 hours per day on the Internet. With increasing Internet users, cybercrime is becoming a growing concern. Several pieces of legislation introduced in Hong Kong to fight cybercrime are the Computer Crimes Ordinance enacted in 1993, Telecommunications Ordinance (Cap. 106), Crimes Ordinance (Cap. 200) and Theft Ordinance (Cap. 210), which has been extended to cover computer crimes. However, these pieces of legislation have not been amended or updated for quite some time and are not particularly applicable against the modern and ever-more-complex cybercrime landscape. Another relevant piece of legislation is the Unsolicited Electronic Message Ordinance (UEMO), enacted in 2007 to cover spam. However, there is no legislation that deals with newer types of cybercrimes such as phishing. The Hong Kong Monetary Authority regularly issues statements warning against fraud and phishing cases. The Legislative Council also cites phishing and botnets as the main causes for a 405% increase in IT security incidents over the four years ending in 2015. Considering that phishing has been recognized as a serious threat to businesses and households, it is rather surprising that there is still no direct legislation to deal with this type of crime. Japan has thorough anti-spam legislation in the Act on Regulation of Transmission of Specified Electronic Mail (2009) and has legislation applicable to phishing.

As a result, Japan only saw computer-related crimes to be accounted for only 0.02% of the GDP in 2015. While the degree of legislation on cybercrime varies across Pan-Asia (see Appendix 1 for more information), countries such as South Korea, Singapore, and Malaysia have effective cybercrime legislation in place. They provide legislation covering generic cybercrime while also covering more complex crimes such as fraud, spam, and phishing. This balance between breadth and depth in legislation is something that Hong Kong and other countries in the region can learn from and adapt for own legislative use in the future to keep up with the complex and evolving nature of cybercrime.

2.1.2 Motivation and Contribution

Motivated by the unique nature and increasing importance of Pan-Asian countries, we extend the work of He et al. [40] by conducting a randomized experiment in this region to test the impact of the publication of security information on security improvement. Specifically, we developed an information security score that reflects an organization's preparedness in terms of cybercrimes. In a manner similar to Moodys and Standard and Poors credit ratings, we build a security evaluation system that can be used as an indicator of the security vulnerabilities of the organizations. The score is constructed from processing large-scale, real-time cyber incident data points from spam emission(CBL, PSBL) and phishing website hosting (APWG, OpenPhish) activities. We argue that organizations would tend to deprioritize security issues when the problems are less likely to directly harm themselves, even though

they create negative externalities to the outside of the companies [4, 68, 78]. Spam and phishing cause significant cost to the email recipients and phishing website visitors, where a significant portion of them are from the outside the organizations. However, there is a notable difference between sending out spam mails and hosting a phishing website. Most spam emails are being sent from Internet connected devices which are compromised by bots [56]. Having bots installed on a company-owned machine may indicate that the organization lacks proper security protection mechanisms . It also means that there is a high possibility of other malware which can be used to harm internal system or steal sensitive data. Thus organizations generating large outbound spam volume can be regarded as ones with insecure information systems. According to our 2017 CBL spam feed, we found that about 55.8% (585,808) among the spam emitting IP addresses (1,048,575) are infected by bots. Depending on the type of bots, the compromised machines can access and steal sensitive internal data and/or participate in DDoS attacks.

Comparing to spam, phishing websites have different underlying mechanisms. While spam can be intermittently emitted by infected computers, phishing websites can only be hosted on dedicated web servers that are operated by the web hosting services. In other words, these phishing websites can be hosted on legitimate hosting services or hijacked websites, depending on the type of attackers. We argue that the organizations hosting phishing websites are more likely to have insufficient security policies and moral hazard against externality [4, 78]. According to our collected data in the focal Pan-

Asian countries, we observe that, among 319 phishing URLs appearing more than three times during 2017, 41.8% of URLs were from legitimate domain (hijacked), and 58.2% was self-registered, or using free hosting companies domains. As phishing attacks involved with the use of legitimate web servers, we expect to observe a different treatment effect on the phishing website hosting, comparing to that of spam emission.

Based on collected spam and phishing data and the associated ranking, we conducted a large-scale randomized field experiment (RFE) to investigate whether informing and publicizing the proposed security score induces an improvement of the organizational security level, which can be measured by the number of reported cybercrime records originated from their networks. To support the experiment, we developed a public treatment website, cybeRatings (https://cyberatings.is.cityu.edu.hk/), to show the scores and rankings of organizations from six Pan-Asian countries and districts (Hong Kong, Mainland China, Singapore, Macau, Malaysia, Taiwan, and Macao). The organizations in the treatment group received three bi-monthly security advisory emails in July, September, and November 2017. The email includes the focal companys security performance report and a personalized URL link for the detailed information in the public website. By visiting our treatment website, the subjects can notice that their security performance is publicized. In addition, with the search function, people can check other companies performance score. Furthermore, we implemented tracking system for both email and website. This enables us to measure treatment effects more precisely by looking at the sub-

jects decision on opening emails and visiting websites. For example, we cannot expect any treatment effects on the companies who never opened our email, or visited website.

Our empirical results show the treatment induced a significant reduction on spam volume, which is consistent with the results from He et al. [40]. In addition, we observed higher treatment effects on companies who actually opened our treatment email, and even higher effects on the organizations who proactively visited our treatment website. Interestingly, we have not observed any significant effect on phishing volume reduction. This may indicate that companies have different incentives in dealing with phishing websites.

This chapter contributes to the literature as following: (1) We publish the first security index website in the Pan-Asian region, using entire population of organizations in 6 target countries who own at least one ASN and valid email address. (2) From rigorous field experiment, we suggest an effective cyber policy design to deal with possible internal threat from botnet, and externality issues on phishing hosts. (3) By using email tracking and web analytics tool, we conduct regression analysis.

2.2 Literature Review

2.2.1 Security Investment Strategies

Researchers from information systems, computer science, and economics are eager to find more efficient solutions to deal with the emergence of endless cybersecurity threats. The root causes of burgeoning cybercrime are discussed

from both technical and economic perspectives. The potential causes include: (i) technical vulnerabilities on the part of organizations, (ii) insufficient economic motivations to counter cybercrimes, and (iii) lack of effective legislation. Without adequate information security measures (e.g., insecure cryptographic protocols, missing anti-virus software), organizations become easy targets for security attacks [5]. To combat technical vulnerabilities, a number of solutions are proposed, for example, spam filtering [18, 26], intrusion detection systems [30, 50, 62], and digital forensics [21, 77]. However, maintaining good information security requires significant investment [34]. Thus, without economic motivation, organizations are reluctant to invest in security infrastructure and countermeasures [4].

As cybersecurity threats are unexpected events and thus hard to predict, it is sometimes difficult to quantify the returns of investment in security adoption [34, 83]. Many organizations do not realize the threats of emerging sophisticated cyberattacks and usually adopt a wait-and-see approach in security investments until a huge security incident significantly affects them [22, 34]. Cyber insecurity is partially due to underinvestment, which is the result of distorted incentives by asymmetric information, network externality, and moral hazard [3, 10]. Legislation can also be a good way to curb cyberattacks to heighten public awareness against cybersecurity threats [29]. Existing works such as Moore and Clayton [55], Quarterman et al. [61], and Tang et al. [76] have documented that security information publication helps improve Internet security condition in the country level. Furthermore, He et al. [40] ex-

tended the literature by proposing an organizational-level security evaluation framework to alleviate the security information asymmetry issue. Specifically, the authors designed a policy for organizations security information disclosures to provide more economic motivations for organizations to improve their Internet security protection. Such disclosure of information helped reduce the information asymmetry issue within organizations. Due to insufficient internal resources and policies, organizations may not have a full understanding of their security problems [29]. In addition, the theory of asymmetric information predicts that organizations will underinvest on cybersecurity when their customers cannot distinguish companies with strong security from those with weak security. Publicizing evaluation reports can force organizations to raise their cybersecurity awareness for the fear of losing customers to their competitors [33, 76]. Furthermore, an industry-level, peer-ranking system may put peer pressure on organizations. In this case, organizations with poor performance could face more pressure from their peers.

2.2.2 Studies on Cyberattacks

To evaluate organizations security levels, this research collected data on two common online scams, namely spamming and phishing. Spam usually consists of unsolicited bulk messages sent out by advertisers to promote their products. Many countries have enacted laws to prevent the spread of spam (e.g., the CAN-SPAM Act in the U.S. and UEMO in Hong Kong). However, adversaries usually use a network of compromised computers (also

known as botnets) to send spam, which can make it difficult to identify the real spammers. Collecting spam data from CBL and PSBL anti-spam block lists, Quarterman et al. [61] developed a public website, SpamRanking.net, for the spam rankings of U.S. companies. Apart from spam, phishing is another one of the latest online crimes that poses a huge threat to financial communities. Bose and Leung [15] conducted research to assess phishing preparedness of Hong Kong banks and compare the performance with the counterparts in Singapore. The study found that companies in both regions perform well in handling bogus phishing websites but need further improvement in handling phishing emails. Also, government advocacy plays an important role to encourage organizations to adopt adequate counter-phishing security measures. Apart from government advocacy, a more in-depth study conducted by Bose and Leung [16] finds that the antecedent factors for firms to adopt counter-phishing measures include credit rating, frequency of phishing attacks, and proliferation of online banking. To maintain the reputation of firms in the area of online banking, organizations tend to adopt more sophisticated anti-phishing measures to safeguard the online security of customers. Adoption of anti-phishing measures may provide a signaling effect to customers that the firms are caring and technologically advanced [17]. Botnet is a neologism combining robot and network. It refers to a collection of computer networks that are contaminated by malware (e.g., virus and Trojan) and controlled by an adversary [73]. After gaining control of a network of computers, the adversary usually use botnets like a group of robots to launch various security

attacks, such as spam, phishing, and denial-of-service attacks. The victims whose computers are contaminated by malware are usually unaware that their computers are being used by the adversary to launch various cyberattacks; such computers are termed zombie computers. Because an adversary uses remote zombie computers to launch cyberattacks, it is very difficult for legal authorities to catch the actual adversary or person. Furthermore, it is difficult for persecutors to collect evidence showing that the adversary launched the cyberattacks. Companies with a weak information security infrastructure have a higher chance of being attacked by malware and becoming a part of a botnet. Therefore, it is important that firms regularly check their corporate information security to ensure that it is up-to-date. While conducting this research, we contacted and received reliable sources of data from international spam and phishing organizations. Based on the volume of spam and phishing from registered domain networks, as measured by ASNs, we developed an information security index that can reflect the security status of a company. As some firms are unaware of their security status, public disclosure of such information may help the firms better evaluate their information security infrastructure. With more information, firms may adopt better security policies and advanced security systems. Hence, it may help firms strengthen their security over time.

2.3 Experiment Design and System Implementation

Hundreds of thousands of personal and business banking details are phished by fake emails and websites. Computers and servers infected with malware or viruses are turned into remotely controlled botnets to send out spam or contribute to DDoS attacks. Email continues to be a popular and effective delivery method for spam, phishing, malware, and, most recently, ransomware. Overall, the proportion of emails that include malware, viruses, or even ransomware is rising dramatically. An organizations Internet security condition is a latent variable that cannot be measured directly. One way to estimate it is by using perceptible data, such as outbound malicious emails and phishing feeds. Symantecs MessageLabs published the 2016 Internet Security Threat Report, which indicates that the global spam volume per day was 24.7 billion messages with an overall email spam rate of 53% in 2015 (Symantec 2016). Among these messages, over 50% of the spam volume was sent by botnets. These infected computers and servers may be used by adversaries as a medium for even more serious cyberattacks, such as phishing, DDoS attacks, identity thefts, hacking, data breaches, and data alterations. Security attacks originating from a corporate network can be a good indicator of weak security infrastructure. In this research, we use: (1) the volume of outbound spam, and (2) real-time phishing intelligence feeds from data sources to construct a comprehensive information security indicator. A voting system Borda count method [1] is used to derive a composite ranking from four constituent rankings from each data source. Organizations with higher Borda counts are ranked

higher, indicating a low security level. All organizations with no volume are ranked equally with the lowest rank.

2.3.1 Large-Scale Randomized Field Experiment

In order to causally test whether publicized security information will induce firms awareness towards their corporate security, and improve their protection level over time, we employ RFE along with econometric analysis as the main evaluation methodology. RFE, also known as randomized controlled trial (RCT), is a well-established evaluation methodology in social science for policy interventions, where the findings can be explained by different factors associated with the interventions or the evaluation [42]. The main advantage of this methodology is the capability of detecting causal relationship in a naturally occurring environment. The subjects in the experiment fall into two equal-sized statistically homogeneous groups, which are divided with stratified and match-pair randomization [57]. The grouping is summarized in figure 2.1. In the control group, there was no treatment. In the public group, three treatment emails were sent to relevant contacts in IT department within each organization to inform their security evaluation results. Each treatment email includes the organizations spam and phishing data, such as total spam mail and phishing website volume, peer rankings in the corresponding industry sectors or certain region, as well as a hyperlink to a designated webpage for the treated organization.

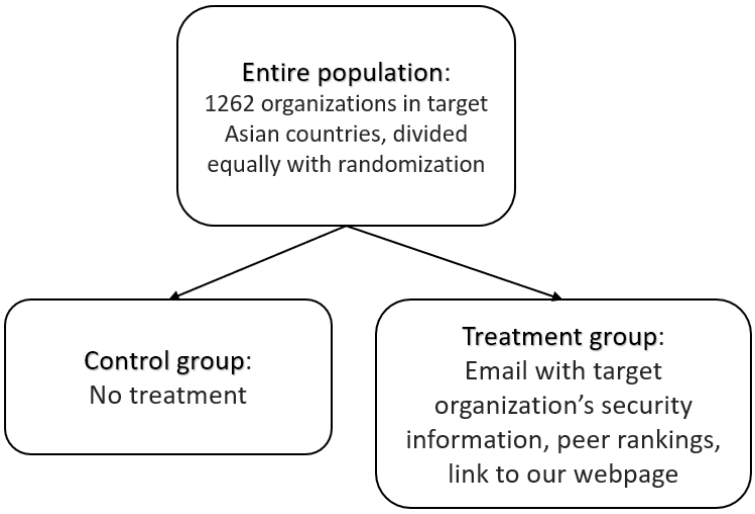


Figure 2.1: Design of the Randomized Field Experiment

2.3.2 Data

Firstly, we collected a full list of 1930 registered ASNs information from the target countries. After mapping the ASNs to registered company names, we created a list of 1293 organizations who own at least one ASN. Lastly, we manually collected and validated corporate email addresses from those organizations, and finalized the list of 1262 organizations. It is important to point out that our field experiment was conducted with a full population of organizations who own at least one registered ASN and a valid email address in six Pan-Asian countries and districts. Table 2.1 shows the number of companies in each country. Figure 2.2 shows the architecture of the entire experiment system. The system is concurrently hosted by the Center for Research on Electronic Commerce (CREC) of the McCombs School of Business at The

Table 2.1: Number of organizations for each country and district

countries and districts	number of organizations	control Group	treatment Group
Hong Kong	309	631	631
Mainland China	309		
Singapore	264		
Malaysia	171		
Taiwan	138		
Macau	4		
Others	67		
Total	1262	1262	

University of Texas at Austin and the Department of Information Systems at the City University of Hong Kong.

The system collects malicious email and website data on a daily basis from various sources: (i) spam/phishing email data from Spamhaus Composite Blocking List (CBL) and Spamikazes Passive Spam Block List (PSBL), (ii) phishing website data feeds from the Anti-Phishing Working Group (APWG) and OpenPhish. CBL and APWGs daily reports are collected to spam and phishing data collector, topaz server, through rsync (a Unix-based file synchronization program), while PSBL and OpenPhish real-time data feeds of the actual spam, phishing contents are stored in topaz server through InterNetNews (inn2). Each spam block list provides daily reports on the total spam volume associated with a complete list of spamming IP addresses. In addition, CBL provides botnet information when available. The data cover

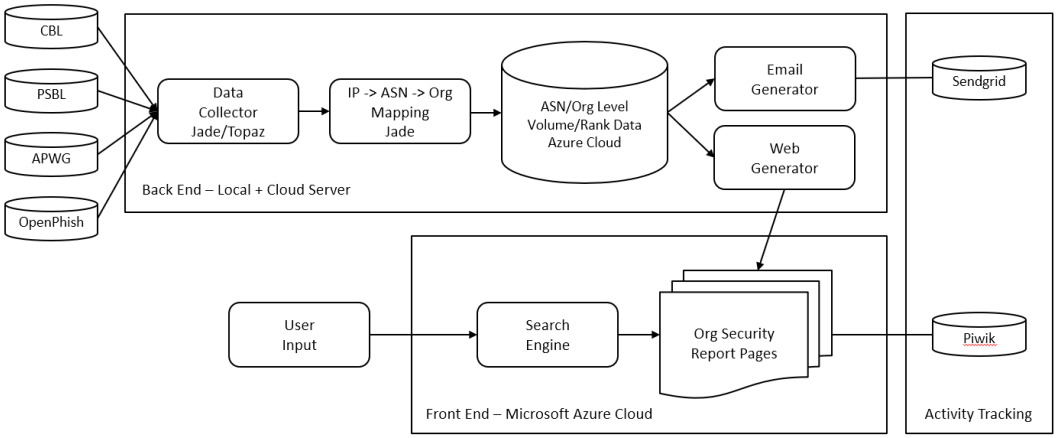


Figure 2.2: System design and implementation

more than eight million IP address, over 190,000 netblocks, and around 21,000 ASNs for 200 countries. PBSL has relatively smaller daily volume compared to CBL, but it provides full email information including raw email header, body, and attachments. APWG provides phishing feeds via eCrime Exchange service and data feeds through phishing data repository (e.g., open and end date, URL, Confidence Level, IP address, etc.). OpenPhish offers daily free phishing intelligence feeds from multiple streams and the analysis is done by applying several prominent phishing detection algorithms. The data repositories of OpenPhish include phishing URL, targeted brand, IP address, country code, ASN info, top-level domain, and discover time. In addition, from raw IP-level data, the organization-level data need to be constructed in order to evaluate an organizations security conditions. Thus, there are three levels of mapping: from IP to netblock; from netblock to ASN; and finally from ASN to organization. With this mapping, it is possible to trace the host organization

of spam mail and phishing websites.

2.3.3 Treatment Channels

Email and website, which are two main treatment channels, play important roles in the experiment design. The email sending system is developed to compose and to send advisory emails bi-monthly to treatment group with customized organizational security reports and URL links to access security ranking web pages. Each security report includes past 3 months of spam volume, number of newly discovered phishing hosts, and peer rankings in the corresponding countries and industry sectors. We provide unsubscription option for organizations who no longer want to receive these emails. By the end of experiment period, we received 2 unsubscription requests, and these organizations were excluded from the analysis.

In addition to the email system, a public website is created to provide organizational security reports to treated organizations and the general public. Visitors can search organizations by names, ASN, industry codes, and country or district names. In the target organizations page, users can select target months from May 2015 to December 2017, and data type (combined overall, CBL, PSBL, APWG, and OpenPhish). It shows daily, monthly volume, and the rankings of firms from three dimensions, namely organization level, industry level, and country level. The website is currently constructed on Microsoft Azure platform to give access to countries who have limited access to the Internet due to the censorship.

Our website outperforms several existing ones (e.g., CBL, Spamhaus, and Cisco) in multiple perspectives:

1. It gives a more complete picture by including smaller spammers. In addition to the top 10 or top 100 spam senders, there are still a lot of organizations sending out a significant amount of spams every day, according to aforementioned data source. Also, organizations who do not have outgoing spam or phishing activities are searchable.

2. It provides organizational-level information on spam and phishing information. Given that many organizations operate multiple Autonomous System Number (ASNs), the metric will combine ASN-level data into organizational one.

3. Instead of snapshot data, it provides continuous and dynamic security information over a long time period from various data sources. With the longitudinal data, people can see how an organizations security situation evolves.

4. It provides unique security ranking data by industry sectors. To correctly define the close competitors, a unified standard industry classification should be enforced such as HSIC (Hong Kong Standard Industrial Classification18) to all Pan-Asian countries in the sense that it is modeled on the United Nations International Standard Industrial Classification of All Economic Activities Revision 2 (ISIC Rev. 2).

2.3.4 Treatment Response Tracking

Besides the website, email tracking and web analytics tools are deployed to check whether employees of an organization have visited the website and become aware of their information security status. Tracking information enables us to perform multiple regression analysis such as Difference in Difference analysis and two-stage least square analysis in Section 2.4. A powerful email management tool Sendgrid is used to track the responses from the treatment group. It provides information for several email related activity including delivery, open, and click. We first check whether the email is successfully delivered to the target mailbox or not. Once it is delivered, the system tells us the information of email opening activity, such as time and IP addresses used to open the email. Also, we track whether the internal links to websites are clicked or not. However, depending on the webmail tool, there are some cases opening action is not traceable. In those cases, we use click information which is always traceable with unique URL link embedded in each email. Web analytics using Piwik is conducted to observe visitor behaviors on the treatment website. We track visitors IP address, location, date and time, opened pages (URLs), duration of visit on each page. By using all available information, we map visitor information to matching organization.

2.4 Empirical Analysis

Our data include 1,262 organizations from 6 Pan-Asian countries and districts: Hong Kong, Mainland China, Singapore, Macau, Malaysia, Taiwan,

Table 2.2: Summary statistics of main variables for empirical analysis.

	description	mean	std.	max	min
cv	CBL volume	151661.8	2269080	1.00e8	0
pv	PSBL volume	147.9001	2698.253	157765	0
av	APWG volume	0.2372	6.1761	456	0
ov	OpenPhish Volume	0.3249	3.1254	105	0
Number of IPs	Total number of IP addresses owned by each company	610223.4	7273093	2.33e8	0
HSIC	Hong Kong Standard Industrial Classification Code			960299	50000
if opened emails	if an organization has opened a treatment email on or before this month	0.2062	0.4048	1	0
if visited website	if an organization has visited our website on or before this month	0.07080	0.2566	1	0

and Macao. Among them, 631 organizations are randomly selected in the treatment group and the rest are in the control group. Since July 2017, we sent out a batch of security information emails to organizations in treatment group every two months. Overall, 565 out of 631 treatment organizations have successfully received at least one treatment email. As a result, we use these organizations and their corresponding control organizations as our empirical analysis data set, total of 1,130. Table 2.2 is the summary statistics of main variables in our analysis.

Table 2.3: Baseline comparison for internal validity

	no control	industry fixed effects	K-S prob (P value)
ln cv_6	0.06324	0.05203	0.934
	(0.2123)	(0.2122)	
ln pv_6	0.05482	0.05312	0.998
	(0.07841)	(0.08130)	
ln ov_6	0.02460	0.02558	1.000
	(0.01978)	(0.02089)	
ln ov_6	0.003235	0.003348	1.000
	(0.007080)	(0.007499)	
ln ip	-0.1309	-0.1580	0.880
	(0.2407)	(0.2483)	
if_social	5.241e-4	9.156e-4	1.000
	(0.02719)	(0.02742)	
HSIC2			1.000

2.4.1 Difference-in-Differences analysis

For the empirical analysis, we use companies spam and phishing volume from July 2017 to December 2017 as companies security measures after our experiment intervention. If an organizations security condition has improved, we would expect its spam and phishing volume to decrease compared with those of control group after our treatment. With the panel data set of organizations spam and phishing information from Jan 2017 to Dec 2017, we apply Difference-in-Differences (DID) model to estimate the treatment effect of our email notification. In particular, email treatment dummy equals to 1 if an organization i is in treatment group and it has successfully received the treatment email in month t. Specifically, the ordinary least squares (OLS)

Table 2.4: DID analysis on monthly security measures

	ln(cv)	ln(pv)	ln(av)	ln(ov)
	(1)	(2)	(3)	(4)
Email_treat	-0.201*	-0.0237	0.0464	0.00577
	(0.115)	(0.0659)	(0.0291)	(0.0353)
Organization fixed effects	yes	yes	yes	yes
Month fixed effects	yes	yes	yes	yes
Constant	-1.256***	-3.940***	-4.439***	-4.350***
	(0.0570)	(0.0355)	(0.0187)	(0.0206)
Observations	13,560	13,560	13,560	13,560
Number of company	1,130	1,130	1,130	1,130

regression function is as follows:

$$y_{it} = \alpha_0 + \alpha_1 email_{it} + \theta_i + \sigma_t + \epsilon_{it} \tag{2.4.1}$$

where y_{it} is one of the security performance measures in our data set. From Table 2.3, we can see the distributions of all main variables are highly skewed, so we use log transformed spam or phishing volume as our dependent variables. Specifically, using CBL spam volume as an example, our dependent variable in analysis is $ln(cv) = log(cv + 0.01)$. In this function, α_1 is our main variable of interest. If α_1 is negative and statistically significant, then compared with organizations in control group, the security performance of those in treatment group has improved after our email intervention. In order to control for organizations time-invariant unobservable characteristics and temporal variation, we also include organization specific (θ_i) and month (σ_t) fixed effects in our regression.

The main results are reported in Table 2.4. We can see that among different security performance measures, our email treatment only significantly influences organizations outbound spam volume measured by CBL. The estimated treatment effect for PSBL spam volume is negative but not statistically significant. On the other hand, for phishing information, there is no evidence showing that our intervention will motivate companies to reduce their phishing volume. The results support our proposition that organizations will have different responses to spam and phishing information. While organizations care about their own potential security issues, they are more reluctant to solve their problems which may bring negative impact to the rest of the world. This can be explained by the negative externality of information security [4].

2.4.2 Heterogeneous Treatment Effects

One possible reason of the insignificant results is that many organizations do not have positive spam or phishing volume during the period of our experiment. As security condition is a relatively hard characteristic to be observed; our existing security measures could not evaluate all organizations cyber security conditions in a very accurate way. Though these organizations security protection levels may have changed, we may lack the ability to precisely measure the difference in our current experiment. Please see detailed numbers in Table 2.5. About 40% of all organizations in our data set show positive spam volume based on CBL. However, only about 22% of them have positive spam volume based on PSBL. For the two phishing volume measures,

Table 2.5: Number of organizations in control and treatment groups with positive spam or phishing volume

	Number of orgs	Number of orgs with positive volume before experiment (treatment)	Number of orgs with positive volume before experiment (control)
cv	1130	228	230
pv	1130	131	120
av	1130	31	27
ov	1130	46	43

there are only about 5% and 8% of organizations with positive volume respectively.

For the reason mentioned above, we repeated the analysis above using subset of organizations with positive spam and phishing volume in the beginning of our experiment. If our treatment emails are effective, we should observe that spam or phishing volume from these organizations have a larger reduction after the experiment. The results are reported in Table 2.6. Compared with data in Table 2.4, we find that the magnitude of the treatment effect for CBL spam volume is larger. More importantly, the treatment effect for PBL spam volume is significantly negative and the magnitude is very close to that of CBL. This further indicates that our email treatment motivates organizations to improve their security protection, leading to less outbound spam volume. However, for the phishing performance, we still could not find evidence of reducing phishing volume.

Table 2.6: DID analysis on monthly security measures with non-zero metrics organizations

	cv (log)	pv (log)	av (log)	ov (log)
	(1)	(2)	(3)	(4)
email_treat	-0.4591*	-0.4381**	0.4931	-0.2563
	(0.2428)	(0.2119)	(0.3875)	(0.3944)
organization fixed effects	yes	yes	yes	yes
month fixed effects	yes	yes	yes	yes
constant	3.388***	-1.896***	-1.998***	-1.662***
	(0.1276)	(0.1367)	(0.2844)	(0.2320)
observations	5,472	3,144	744	1104
number of company	456	262	62	92

2.4.3 Two-stage Least Squares Analysis

One potential reason of the relatively weak treatment effect is that employees of these treated organizations may not actually think over our emails. For example, the successfully delivered emails may not be opened at all. On the other hand, some organizations may pay more attention to our treatment by visiting our website through the link in the email.

Table 2.7 shows the organizations responses on our treatment by tracking data from email tracking tool (Sendgrid) and web analytics tool (Piwik). In the table, treatment group is divided into two subgroups, based on spam and phishing records in 2017. Among 565 organizations who successfully received our treatment email, 257 (45.5%) had emitted at least one spam email or hosted at least one phishing website. Email opening rate shows that 6% more organizations who have zero volume endogenously decided to open the

Table 2.7: Email open/website visit counts among 565 companies who received our treatment email.

	Number of organizations in the treatment group			
Volume from all data sources	Total	Opened Email (/Total)	Visited website (/Open)	Multiple Visits (/Visit)
Orgs with no spam & phishing	308	150 (48.7%)	44 (29.3%)	33 (75.0%)
Orgs with 1+ spam/phishing	257	110 (42.8%)	32 (29.0%)	25 (78.1%)
Total	565	260 (46.0%)	76 (29.2%)	58 (76.3%)

email titled Security Advisory Report for (organization name), sent from email address advisory@cityu.edu.hk . It tells us that organizations who have better protection care more about security related news (Z-score = 1.4011, p-value = 0.08076, one-tailed). However, once the email was opened, website visit rates and multiple visit rates are nearly identical within the minimal error rate (± 1) between the two groups.

One econometric challenge of estimating these treatment effects is that these actions including opening emails and visiting website are endogenously determined by treated organizations. Directly estimating the regressions with the corresponding dummies may lead to biased estimators. Taking advantage of our randomization, we use the dummy variable indicating whether the security measure is from a treatment organization after July 2017 as an instrumental variable (IV) for organizations decisions to open an email or to visit our website. Since only organizations in the treatment groups can receive our

Table 2.8: 2SLS for treatment effects on opening treatment emails

	ln(cv)	ln(pv)	ln(av)	ln(ov)
	(1)	(2)	(3)	(4)
open a treatment email	-0.591*	-0.165	0.120	0.00636
	(0.346)	(0.190)	(0.0874)	(0.104)
organization fixed effects	yes	yes	yes	yes
month fixed effects	yes	yes	yes	yes
constant	-1.256***	-3.940***	-4.439***	-4.350***
	(0.0570)	(0.0355)	(0.0187)	(0.0206)
observations	13,560	13,560	13,560	13,560
number of company	1,130	1,130	1,130	1,130

treatment emails, so monotonicity condition is satisfied in our case. Then we apply two-stage least square regression (2SLS) to estimate the local average treatment effects of opening our email and visiting our website (Imbens and Angrist, 1994). The specific regression functions are as follows:

$$D_{it}^{*} = \gamma_0 + \gamma_1 email_{it} + \epsilon_{it} \tag{2.4.2}$$

with the observed email opening or website visiting indicator, D_{it} related to the unobserved latent index, D_{it}^{*} , by

$$D_{it}^{*} = \begin{cases} 1, & D_{it}^{*} > 0 \\ 0, & D_{it}^{*} <= 0. \end{cases} \tag{2.4.3}$$

And the dependent variable y_{it} is related to the treatment by the equation

$$y_{it} = \beta_0 + \beta_1 D_{it} + \mu_{it}. \tag{2.4.4}$$

The results for the local average treatment effect of opening an email and visiting our website are reported in Table 2.8 and Table 2.9. All the standard

Table 2.9: 2SLS for treatment effects on visiting our website

	ln(cv)	ln(pv)	ln(av)	ln(ov)
	(1)	(2)	(3)	(4)
visit treatment website	-1.931*	-0.539	0.392	0.0208
	(1.141)	(0.624)	(0.290)	(0.340)
organization fixed effects	yes	yes	yes	yes
month fixed effects	yes	yes	yes	yes
constant	-1.256***	-3.940***	-4.439***	-4.350***
	(0.0572)	(0.0355)	(0.0187)	(0.0206)
number of observations	13,560	13,560	13,560	13,560
number of organizations	1,130	1,130	1,130	1,130

deviations are robust and clustered at company level. Similar to the results in Table 4, only the coefficient of companies spam volume based on CBL is negative and significant. However, the magnitude of the coefficient is much larger (-0.591 and -1.931), indicating that organizations who indeed opened the emails and visited our website tend to perform better. More specifically, outbound spam volume from organizations which opened our emails has decreased by 44.1% and that from organizations which visited our website is reduced by about 85.4%. There can be two potential mechanisms to explain our results: 1. Only organizations which opened our treatment emails have received our treatment, leading to enhanced security performance; 2. Organizations who chose to open our emails or even visited our websites are those who are more vigilant about potential security threats. Hence, they are more likely to improve their security safety measures after receiving our treatment emails.

Table 2.10: 2SLS for treatment effects on opening treatment emails with non-zero metrics organizations

	ln(cv)	ln(pv)	ln(av)	ln(ov)
	(1)	(2)	(3)	(4)
Open a treatment email	-1.440*	-2.016**	1.182	-1.451
	(0.811)	(0.832)	(1.209)	(1.964)
Organization fixed effects	yes	yes	yes	yes
Month fixed effects	yes	yes	yes	yes
Constant	3.388***	-1.896***	-1.998***	-1.662***
	(0.127)	(0.139)	(0.284)	(0.232)
Number of observations	5,472	3,144	744	1,104
Number of organizations	456	262	62	92

Table 2.11: 2SLS for treatment effects on visiting our website with non-zero metrics organizations

	ln(cv)	ln(pv)	ln(av)	ln(ov)
	(1)	(2)	(3)	(4)
Visit treatment website	-4.915*	-7.989**	5.403	-3.824
	(2.836)	(3.940)	(6.840)	(5.446)
Organization fixed effects	yes	yes	yes	yes
Month fixed effects	yes	yes	yes	yes
Constant	3.388***	-1.896***	-1.998***	-1.662***
	(0.128)	(0.143)	(0.290)	(0.232)
Number of observations	5,472	3,144	744	1,104
Number of organizations	456	262	62	92

2.5 Discussion and Concluding Remarks

Using a large-scale randomized field experiment, we empirically study how security evaluation publication affects organizational security levels in Pan-Asian countries and districts. To measure the pre- and post-experimental information security risk level of the organizations, we use two distinct perceptible cyberattack data: outbound spam volume and phishing websites. To increase security awareness in the general public and increase economic motivations on the part of organizations, security performance rankings were published on our project website. In doing so, an organization with a weak information security level may have faced a threat of reputation loss among customers. From a series of regression analysis on two different types of security attacks, we found evidence that the security report publication has a statistically significant effect in reducing spam volume. The treatment effects gradually increased from receiving emails (results from DID) to opening emails and visiting the website (results from 2SLS).

On the other hand, we do not find a statistically significant effect on phishing website hosting behavior. There are two possible explanations for this: First, web hosting companies do not have economic incentives to eliminate phishing websites as they are legitimate customers of the hosting services. This can be considered to be a negative externality issue. Second, there is a lack of strict phishing-related policies in Pan-Asia compared to those geared toward spamming activities, and those in force impose less liability risk for the website hosting services. Following this line, some ISPs and hosting services

have policies which pass responsibilities on to their customers. Although we did not have statistically significant results in phishing reduction, we observed anecdotal cases in which our treatment induced positive changes: among 46 treated companies who hosted phishing websites according to OpenPhish data, six of them actually eliminated all phishing websites within one or two months after their first response (opened an email and/or visited the website) to our treatment. Based on the other phishing data from APWG, among 31 organizations hosting phishing websites, four addressed the issues fully. This may suggest that the provided information was appreciated and induced a certain level of improvement in the subjects security protection level. To summarize, our results from the empirical analysis suggest that security monitoring websites, such as cybeRatings, can be effective in terms of reducing botnet activities represented by outgoing spam volume. At the same time, we observed that organizations have different incentives in terms of managing phishing attacks. This has a policy implication in that stronger regulations may be needed to internalize the negative externalities resulting from organizations hosting phishing websites.

As a functional direction, we are currently preparing multiple extensions of our experiment in terms of communication channels and scope. First, we will use massive social media platforms (e.g., Twitter, Facebook, Weibo, and WeChat) to share the security reports with the treated organizations. One unique advantage of using a social media treatment compared to an email treatment is that social media are closely followed by customers and strategic

partners. As such, information disclosure on social media may lead to more pronounced reactions from the treatment organizations. In addition, by using direct messages in social media channels, deliverability could be improved from the relatively low email opening rate (46%). In order to avoid some spillover effects to the control group, the treatment effect on social media will only be applied to the public treatment group. The effect of social media treatment can be measured by the difference between organizations who only received treatment emails and those whose security reports are also disseminated via social media.

The second extension is to expand the scope of the experiment. Organizations preparedness in terms of cybercrimes has become a global debate in this globally hyper-connected economy. It is also possible that the designs or regulations that are effective in the U.S. or Pan-Asia may not work on other continents [47]. As cybersecurity issues are not isolated to specific countries, it is necessary and beneficial for all countries to collaborate on this issue. As our spam and phishing data include information from more than 200 countries worldwide, we plan to generate and publicize security reports for other countries organizations. Doing so will increase the population size of the experiment, which was a limitation for our study in Asia. With a larger sample size, we plan to add more treatment groups with different email contents. For example, one group will receive emails with general spam/phishing activity information that are similar to those used in the current study, and the other group will receive more comprehensive information, including information on

the actual botnets installed in the subjects system, possible threats from the botnets, a detailed list of IP addresses involved in the cybercrime, and possible measures to mitigate the issue.

Chapter 3

To Disconnect or Not: A Cybersecurity Game

3.1 Introduction

Our daily lives, business operations, and government services heavily rely on the Internet infrastructure [25]. As our dependency on the cyberspace has ever increased, however, we also witness an ever-growing number of cyber threats and associated financial damages.[1] Data driven society makes personal information more valuable, and gives strong motivations to cyber attackers who mainly seek out financial gain. Well-organized hackers operate as for-profit businesses seeking to maximize profit, where the profit can be based on the cumulative attack intensity.[2] The defending side, such as cybersecurity service providers, and governments, is also taking active measures to cope with cyberattacks. Governments are actively implementing new cyber policies,[3] while using intelligence to track down high-profile cyber attackers.[4] Cybersecurity service providers vigorously analyze new threats every day and

[1] Annual Internet security threat report (2016). Symantec Corporation.

[2] For example, the attack intensity can be volume of spam emails, or intensity of distributed denial-of-service (DDoS) attacks.

[3] https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/executive-order-improving-critical-infrastructure-cybersecurity.

[4] Recently conducted randomized field experiment provided a causal evidence that proper policy implementation can significantly mitigate cyberattacks [He et al. 2016].

develop solutions for their clients in both hardware and software levels. Recently, advanced machine learning methods such as deep learning are being used for zero-day attack protection [35, 38].

Even with the technological advances in the security systems and the policy implementations, cyber risks will continue to exist due to the strategic actions by the financially motivated attackers. In that sense, cyber risk is a factor that has to be "managed" rather than something that can be eliminated, and cyberinsurance market can be a solution to manage the cyber risks. According to Wall Street Journal, cyberinsurance is one of the fastest growing products in the United States in 2016.[5] An insurance broker Marsh stated that the cyberinsurance market has grown to more than $2 billion in gross written premiums in 2014, and has potential to grow to $20 billion by 2025.[6]

Cyberinsurance market has unique characteristics compared with conventional insurance markets (e.g., health and automobile). In the traditional insurance scheme, risk assessment is done with the actuarial tables, which calculate probable outcome based on various risk factors. This actuarial data have been established from statistics on a long history of risk data. However, such table for cyber risks is yet to be established. For example, cyberinsurance rate surged 32% in the first half of 2015. From 2015 to 2016, insured customers experienced 10% to 150% of premium rate increases.[7] It clearly shows that the

[5]https://www.theguardian.com/technology/2016/oct/26/ddos-attack-dyn-mirai-botnet.
[6]Mmc cyber handbook, (2016). Marsh & McLennan Companies.
[7]Marketplace realities (2016). MarketScout.

cyber insurance market is premature due to the unpredictable nature of the cyber security domain. Moreover, unlike the auto accidents that occur due to mistakes of drivers or local weather conditions, cyber incidences are deliberate outcomes by the strategic attackers.[8]

Internet Service Providers (ISPs) can play a pivotal role in the cyberinsurance. In the Internet infrastructure, ISPs act as the central hubs that inter-connect with other ISPs and the Internet backbones. Thus, ISPs have visibility to a broad range of Internet traffic, which can provide ISPs informational advantages. Besides the role of hubs, ISPs also provide access service to the end users. It is rather obvious that, if ISPs proactively monitor and block malicious activities, cyber-attacks do not reach to the end users' systems. Such service falls into a broad concept called "managed security services" in the industry.[9] Managed security services can be especially attractive to small and medium businesses (SMBs) with limited resources. Unlike large companies who can manage the security with well-trained in-house professionals, SMBs suffer from frequent online attacks.[10] It is well known that cyber attackers strategically select the easy targets.[11] Reports say that 60% of SMBs suffered from cyberattacks went out of businesses in six months.[12] Managed security

[8]In the case of such unique risk characteristics, reinsurance can be a solution, where domain experts assess and buy risks then resell them to other insurers to create syndicates. Besides the commercial success of reinsurance market such as in Lloydś of London, mathematical foundations for reinsurance have been established in the academia [13, 45, 53].

[9]Interestingly, only a handful of ISPs such as Verizon and AT&T are providing such services.

[10]State of cybersecurity in small & medium-sized businesses, (2016). Ponemon Institute.

[11]Flipping the economics of attacks. Ponemon Institute.

[12]http://www.denverpost.com/2016/10/23/small-companies-cyber-attack-out-of-

services with cyberinsurance could be a possible solution to this issue with affordable cost to the SMBs.

Although ISPs have aforementioned informational and structural advantages, only a handful of ISPs are providing managed security services, but none of such services provide warranty coverage in the event of loss of their clients. Similarly, while security software publishers develop more secure software, they do not provide any liability.

In this article, we propose a new cybersecurity-driven business model for ISPs. In addition to the traditional Internet access services, ISPs can provide managed security services with cyberinsurance for cyber risks. In other words, ISPs provide liability insurance to their clients with the exchange of insurance premiums. From the clients' perspective, they do not need to staff in-house security experts, as the security services are outsourced to the ISP. In addition, clients only have limited liability in case of security incidences, as the insuring ISP compensates the loss. To manage cyber risks, ISPs will make strategic investment (e.g., multiple layers of protections from the network level to the end point) to build secure environments to their clients.[13] While large ISPs like AT&T and Verizon may not have strong incentive to embrace this new business model, we believe that emerging ISPs can boost their market shares by adopting the cybersecurity-driven business model.

---

business/.

[13]To prepare for the contingency of huge losses, ISPs should work with reinsurance markets to form syndicates as in [53].

To build a theoretical foundation of such cyberinsurance business model, we develop a dynamic game model and analyze the strategic interaction between an attacker and a defender in the cyber space. Specifically, the defender is the ISP providing managed security services with insurance and the attacker is either a host or a set of hosts that create Internet connections with the clients in the defending ISP. The game starts when the user initiates her connection with the defender. She can be either a compromised user (e.g., a PC with malware, controlled by a bot master) or an innocent user. The role of the players and the concept of the Bayesian Nash equilibrium in our model is described below.

- **The defender** cannot directly observe the actual identity of the user, so the model has asymmetric information structure. But he can dynamically update the suspicion level (the probability that the user is indeed an attacker) by observing the stream of the user's total actions.[14] The defender minimizes his expected costs by stopping the game (disconnect the user) based on the observation stream. The total cost for the defender is sum of (i) the cumulative damage by the attacker (if exists) and (ii) false positive blocking cost in case the blocked user was an innocent one. The strategic decision is to choose the optimal blocking time to minimize the expected total cost.

[14]The user's total actions consist of malicious activities (controlled by the bot master) and some noise (regular activities by the owner of the device).

- **The attacker** dynamically chooses the attack intensity to maximize the expected profit that can be obtained until she is blocked by the defender. The optimal attacking strategy should be determined with the consideration that higher attack intensity incurs higher immediate payoff but earlier termination (block) of the game. Therefore, the attacker's strategy at time t depends on the suspicion level at that time, which is computed by the defender based on the observation of the signal process up to time t.

- **The Bayesian Nash equilibrium** in this game consists of (i) the attacker's optimal strategy and (ii) the defender's optimal stopping time and the suspicion level adjustment formula.

We prove that there exists a unique Bayesian Nash equilibrium in this game model, and find explicit expressions for the optimal attacking intensity for the attacker and the blocking threshold for the defender.

To the best of our knowledge, our model is the first to explore dynamic game between strategic informed player and uninformed player who not only updates his belief but also optimally terminate the game based on the belief. Our game framework is suitable for network security, since the most obvious remedy for possible threat is blocking suspicious users. The distinctive feature of our model is to endogenize defender's blocking policy of when to terminate the game, based on the available information, in the continuous time framework.

We fully analyze the equilibrium in our game model and obtain several interesting implications.

- Our model gives answer to the following question: Will the cyber-attack explode as the attacker's maximum attack capacity increase? This is a natural question, since the Internet capacity is keep expanding these days, and more Internet capacity implies more attack capacity. We compare our equilibrium result and three benchmark cases, and conclude that for the viability of the Internet-based society, the defender's roles of updating suspicion level and blocking suspicious users are essential, and also the defender should take into account the strategic nature of the attacker. Otherwise, the expected costs of the defender explodes as the maximum attack capacity getting higher. See Section 3.3 for details.

- We describe and interpret the equilibrium behaviors of the players. We list just a few of them here: The equilibrium blocking threshold decreases as the maximum attack intensity increases, and as the defender's ability of filtering decreases; The attacker chooses maximum attack-intensity for sufficiently low suspicion level, then gradually decreases the attack-intensity as the suspicion level increases, i.e., the attacker tries to lower the possibility of being blocked when she is highly suspected; Both attacker and defender are less active when the suspicion level is close to the blocking threshold, etc. See Section 3.4 for details.

- The model support our claim that ISPs should be the insurance provider,

in two different aspects. First, we illustrate how much insurance premium can be saved if ISPs take the insurance liability and plays the role of the defender in our model, compared to the traditional insurance provider. It turns out that the amount of saving is significant if the maximum attack capacity is high. Secondly, we provide a simple formula for the probability of the event that the user is blocked eventually in equilibrium. The formula can be used for the calibration of the initial suspicion level (an important model parameter), and the defender should play the game multiple times with different users to make this calibration more accurate. ISPs are in the optimal position for such tasks. See Section 3.5 for details.

Attacker in our model strategically control her action to hide her identity from defender, and this ingredient of model is connected to deception literature in game theory. Hendricks and McAfee [43] consider a one-shot game with sender and receiver, and describes how the signaling technology affects equilibrium strategy of the attacker. Crawford [28] shows that in the interaction of rational and boundedly rational types of players, the deception can be used by rational type. Aumann and Maschler [6] study dynamic game of incomplete information in discrete time framework.

Our model is related to insider trading literature in finance. Kyle [48] describes interaction of market makers and an insider who has long-lived private information about an asset value, and studies equilibrium pricing of the market makers and dynamic trading of the insider. A version of Kyle model

studied in Back and Baruch [7] is closer to our model: their insider has private information of binary asset value and our attacker has private information about her identity which is binary (hacker or not); their market makers update price of assets and our defender updates suspicion level.

The closest model to ours is the model in Anderson and Smith [2], where the private information and profit structure of attacker is similar to our model. Accordingly, the equilibrium attacking intensities are analogous. The major difference between these two models is the role of defender. Anderson and Smith [2] consider continuum of myopic defenders who instantaneously choose mixed strategy of binary actions. On contrast, we consider a single long-lived defender who plays the critical role of blocking the user (i.e., terminates the game) when the suspicion level reaches certain threshold. Furthermore, our defender is not myopic because he chooses the optimal threshold to minimize the expectation of overall costs until the end of the game.

Our model can be also considered as a game version of the sequential testing literature in mathematics statistics. Based on the continuous observation of user's action, our defender's task is to test sequentially the hypothesis whether the user is an attacker or not, and to find the optimal stopping policy to minimize expected costs. In other words, the defender side narrative in our model is to solve a sequential testing problem for hypothesis about the drift part of an observed Wiener process. The key difference between our game model and the sequential testing problem is that our defender is dealing with the strategic attacker who takes into account the defender's strategy: the

sequential test results are provided by an adversarial agent.

3.2 The Model

We consider a continuous time game between a user (who can be either an attacker or an innocent user) and a defender. The identity of the user is represented by a random variable θ which can take two values 0 or 1: $\theta = 1$ means that the user is an attacker and $\theta = 0$ means that the user is an innocent user. The defender is uninformed about the value of θ, and has prior $q_0 = \mathbb{E}[\theta] \in (0, 1)$ which is the defender's initial estimation of probability that the user is an attacker. In case the user is an innocent one, the user performs no malicious action. Otherwise, the attacker chooses her attack intensity Δ_t dynamically over time $t \geq 0$. We set a constant $M > 0$ as the upper bound for attack intensity, i.e., $0 \leq \Delta_t \leq M$ for all $t \geq 0$.

We consider the defender who can disconnect the user and terminate the game, based on the observation of the user's action. We assume that the defender's observation of the user's action flow $\Delta_t 1_{\{\theta=1\}} dt$ is obscured by a noise term $\sigma\, dW_t$, where $\sigma > 0$ is a constant and $(W_t)_{t\in[0,\infty)}$ is a standard Brownian motion independent of θ.[15] In other words, the signal process $(Y_t)_{t\in[0,\infty)}$ the defender can observe is expressed as:

$$dY_t = \Delta_t 1_{\{\theta=1\}} dt + \sigma\, dW_t. \tag{3.2.1}$$

[15]One way to interpret this is to consider $\Delta_t dt$ as the malicious action by generated by the bot master and $\sigma\, dW_t$ as the normal traffic by the original owner of the computer.

Based on the available information up to time t obtained by the observation of the signal process Y, the defender can updated the suspicion level q_t (the defender's estimation of probability that the user is an attacker), i.e.,

$$q_t = \mathbb{P}(\theta = 1 | \mathcal{F}_t^Y), \tag{3.2.2}$$

where $(\mathcal{F}_t^Y)_{t\in[0,\infty)}$ is the filtration generated by the process Y. We assume that the signal process Y is public information and known to the attacker. This implies the admissibility condition $\Delta_t \in \mathcal{F}_t^Y$ for the attacker strategy. Using [Lipster and Shiriyaev Theorem 8.1], the filtering equation (3.2.2) produces the following Stochastic Differential Equation (SDE) that q_t should satisfiy:[16]

$$\begin{aligned} dq_t &= \frac{1}{\sigma^2}\Big(\mathbb{E}[\theta\,\Delta_t 1_{\{\theta=1\}}|\mathcal{F}_t^Y] - \mathbb{E}[\theta|\mathcal{F}_t^Y]\cdot\mathbb{E}[\Delta_t 1_{\{\theta=1\}}|\mathcal{F}_t^Y]\Big) \\ &\qquad \cdot\Big(dY_t - \mathbb{E}[\Delta_t 1_{\{\theta=1\}}|\mathcal{F}_t^Y]dt\Big) \\ &= \frac{q_t(1-q_t)\Delta_t}{\sigma^2}\Big(dY_t - q_t\Delta_t dt\Big) \end{aligned} \tag{3.2.3}$$

We consider a game between the attacker and defender, and the Bayesian Nash equilibrium consists of (i) the attacker's optimal strategy and (ii) the defender's optimal stopping strategy.

(i) Attacker's profit maximization problem

The attacker obtains instantaneous profit of $\Delta_t dt$ through her malicious actions. The game is terminated when the attacker's identity is determined by

[16] In (3.2.3), the term $dY_t - q_t\Delta_t dt$ is innovation (or surprise) the defender perceives. (3.2.3) says that the adjustment of q is proportional to the surprise. The term $q_t(1-q_t)\Delta_t/\sigma^2$ describes how sensitively the belief changes with respect to the surprise.

outside factor[17], (at time T) or the defender disconnects the user (at time τ_p). We assume that T is independent of θ and $(W_t)_{t\in[0,\infty)}$, and has exponential distribution, $\mathbb{P}(T > t) = e^{-rt}$ with a constant $r > 0$. The stopping time τ_p is defined as[18]

$$\tau_p = \inf\{t \geq 0 : q_t \geq p\}, \tag{3.2.4}$$

i.e., the defender disconnect the user when the suspicion level q is above certain threshold $p \in [0, 1]$.

The attacker seeks the optimal strategy Δ to maximize her expected cumulative profit until the game is over:

$$\max_{0\leq(\Delta_t)_{t\in[0,\infty)}\leq M} \mathbb{E}\Big[\int_0^{T\wedge\tau_p} \Delta_t dt \,\Big|\, \theta = 1\Big]. \tag{3.2.5}$$

The attacker recognizes that her actions affect the stopping time τ_p in such a way that more aggressive actions (larger Δ) will increase the suspicion level q faster through the defender's Bayesian update, and eventually terminate the game sooner (smaller τ_p).

(ii) Defender's cost minimization problem

We assume that the defender has two types of costs - cumulative cost from malicious actions of attacker in case $\theta = 1$, and one-time cost of *false alarm* if the defender disconnects an innocent user. To describe the class of admissible strategies of the defender, let $\mathfrak{T}$ be the set of all stopping times with

[17] It can be caused by a security patch, bug fix, or blacklisting from the other defenders

[18] τ_p is a stopping time respect to the filtration $(\mathcal{F}_t^Y)_{t\in[0,\infty)}$.

respect to the filtration $(\mathcal{F}_t^Y)_{t\in[0,\infty)}$. The defender's goal is to find the optimal disconnecting strategy to minimize the expected total costs:

$$\min_{\tau\in\mathcal{T}} \mathbb{E}\Big[\Big(\int_0^{T\wedge\tau} \Delta_t dt\Big)\cdot 1_{\{\theta=1\}} + l_f \cdot 1_{\{\theta=0,\,\tau<T\}}\Big], \tag{3.2.6}$$

where the constant $l_f > 0$ is the one-time-cost of blocking an innocent user. We can check that two extreme cases produce trivial optimal stopping strategy: In case $l_f = 0$, then the defender should block the user immediately, i.e., set $\tau \equiv 0$; If $l_f = \infty$, then the defender never block the user, i.e., set $\tau \equiv \infty$. For $0 < l_f < \infty$, we will see that the optimal p satisfies $0 < p < 1$.

As a real-world example, we can apply our model to the botnet and botherder situation. Bot-herders[19] never attack targets with their own machines, and use their botnet as a front line army. Each botnet consists of many individual bots which are compromised Internet connected devices such as PCs, tablets, or IoT devices. In most cases, the owners of compromised devices will use their machines daily without being aware of the existence of bots. This regular activity is modeled as noise dW_t in (3.2.1), and the malicious activity of a bot is denoted by $\Delta_t dt$. The defender updates the suspicion level q_t based on the signal Y_t, which is sum of the regular activity of the owner of the device and the malicious activity generated by the bot. In addition, most ordinary users do not try to hide themselves or jump around different IP addresses for their privacy. In our model, the goal of the defender is not

[19]Who are running many bot-infected devices(botnet) to attack targets

eliminating botherders which is very complicate and difficult task, but blocking compromised hosts who are directly affecting the customers. Depends on the type of the botherders, some bots remain silently and attack strategically under noise[20], and some others attack vigorously and get detected earlier. We consider several different situations regarding the type of the attacker in Section 3.3.

The optimal stopping strategy τ is endogenously determined by the defender's optimization problem. This feature of termination of game based on the suspicion level is the distinctive ingredient of our model compared to existing Bayesian game models: Kyle (1985) [48] considers fixed terminal time; Back and Baruch (2004) [7] and Anderson and Smith (2013) [2] set the terminal time as an independent random time (as T in this paper) which is not part of equilibria.

In this framework, our goal is to study Bayesian Nash equilibrium that consists of the attacker's optimal strategy and the defender's optimal stopping strategy. The definition of the equilibrium is following.

Definition 3.2.1. *Consider a constant $p \in (0,1)$ and a Lipschitz continuous function $\alpha : [0,1] \to [0,M]$. Let $(Y_t)_{t\geq 0}$ be as in (3.2.1) and τ_p as in (3.2.4). We say that the pair (p,α) is a Bayesian Nash equilibrium if following (1) and (2) hold.*

[20] Jaku botnet example: https://www.helpnetsecurity.com/2016/05/05/jaku-botnet-targeted-attacks/

(1) (Attacker's optimal intensity) Let $(q_t)_{t\geq 0}$ *obeys the SDE*[21]

$$dq_t = \frac{q_t(1-q_t)\alpha(q_t)}{\sigma^2}\Big(dY_t - q_t\alpha(q_t)dt\Big). \tag{3.2.7}$$

Then, $\big(\alpha(q_t)\big)_{t\geq 0}$ *is the solution of the attacker's profit maximization problem, i.e.,*

$$\big(\alpha(q_t)\big)_{t\geq 0} \in \underset{0\leq(\Delta_t)_{t\in[0,\infty)}\leq M}{\arg\max}\ \mathbb{E}\Big[\int_0^{T\wedge\tau_p}\Delta_t dt\ \Big|\ \theta=1\Big]. \tag{3.2.8}$$

(2) (Defender's optimal stopping) Let $(q_t)_{t\geq 0}$ *obeys the SDE* (3.2.7) *with* $\Delta_t = \alpha(q_t)$*. Then, the stopping time* τ_p *solves the defender's cost minimization problem:*

$$\tau_p \in \underset{\tau\in\mathcal{T}}{\arg\min}\ \mathbb{E}\Big[\Big(\int_0^{T\wedge\tau}\alpha(q_t)dt\Big)\cdot 1_{\{\theta=1\}} + l_f\cdot 1_{\{\theta=0,\,\tau<T\}}\Big], \tag{3.2.9}$$

where $\mathcal{T}$ *be the set of all stopping times with respect to the filtration* $(\mathcal{F}_t^Y)_{t\geq 0}$*.*

To clarify, the exogenously given model parameters are q_0, r, M, σ and l_f, and the endogenously determined quantities through our Markovian equilibrium are α (attacker's optimal strategy) and p (defender's optimal stopping policy).

Theorem 3.2.2. *There exists a Bayesian Nash equilibrium* (p, α)*.*

The explicit formula of the equilibrium are given in the appendix.

[21]This SDE is originated from the update of belief, (3.2.2), and its SDE form (3.2.3).

3.3 Viability of the Internet-based Society

Expansion of the Internet capacity due to the increase of network-enabled devices and faster Internet speed introduced more cybersecurity related issues. For example, a DDoS attack from tens of millions of IoT devices caused several hours of blackout on DNS servers operated by Dyn.[22] It was the biggest DDoS attack with an estimated throughput of 1.2 terrabits per second. Also, the Internet Protocol version 6 (IPv6) enables a lot more devices to be connected to Internet space.[23] For cyber attackers, more Internet capacity implies more attack capacity, that is, higher M in our model.

In this section, we extract our model's implication regarding this issue of increasing attack capacity. We analyze the relationship between M (the maximum attack capacity) and the equilibrium costs of the defender. In addition to our original game model, we consider three auxiliary cases below, as stepping stones to the path to our equilibrium (the 4th case). We explain motives of the attacker or defender to be more 'strategic': Each case should naturally evolve to the next case, and finally we reach the case of our equilibrium.

3.3.1 Case of No-defence

To begin with, we consider the case the defender does nothing (no-defence). Then there is no reason for attacker to hide her identity, and the

[22] https://www.theguardian.com/technology/2016/oct/26/ddos-attack-dyn-mirai-botnet
[23] http://iot6.eu/ipv6_advantagess_for_iot

attacker will choose attack intensity as M (the maximum attack capacity) all the time. The corresponding expected profit $V_1(q_0)$ of the attacker and expected cost $C_1(q_0)$ of the defender are

$$\begin{aligned} V_1(q_0) &:= \mathbb{E}\Big[\int_0^T M\,dt \;\Big|\; \theta = 1\Big] = \frac{M}{r} \\ C_1(q_0) &:= \mathbb{E}\Big[\Big(\int_0^T M\,dt\Big)\cdot 1_{\{\theta=1\}}\Big] = \frac{q_0 M}{r} \end{aligned} \tag{3.3.1}$$

We observe that $C_1(q_0) \to \infty$ as $M \to \infty$. This means that if there is no defense mechanism at all, then the Internet-based society may not be viable when the maximum attack capacity is very high.

3.3.2 Case of Non-strategic Attacker vs. Defender

The previous case is obviously not a desired situation for the defender side. Therefore, it is natural to expect that the defender does some defense activities: He blocks the user to minimize his expected costs. In this subsection, we consider the case the attacker is not strategic, i.e., the attacker is not aware the role of the defender and just chooses the attack-intensity as M all the time.

Proposition 3.3.1. *In the case of non-strategic attacker vs. defender, the optimal blocking threshold $\tilde{p}$, the minimal expected costs of the defender $C_2(q_0)$, and the corresponding expected profit of the attacker $V_2(q_0)$ have following asymptotic behavior:*

$$\lim_{M\to\infty} \tilde{p} = 1, \quad \lim_{M\to\infty} C_2(q_0) = 0, \quad \lim_{M\to\infty} V_2(q_0) = 0. \tag{3.3.2}$$

The explicit formula for $\tilde{p}, C_2, and V_2$ are given in the appendix.

Proposition 3.3.1 says that as $M \to \infty$, the expected costs of the defender becomes negligible. Here is an intuitive explanation. When $\Delta_t = M$, the adjustment equation (3.2.3) becomes

$$dq_t = \frac{q_t(1-q_t)M}{\sigma^2}\Big(dY_t - q_t M dt\Big). \tag{3.3.3}$$

We observe that for bigger M, the the belief process q_t reacts to the surprise $(dY_t - q_t M dt)$ more sensitively, i.e., q_t moves toward the true state of θ (identity of the user) more quickly. This means that it is easier for the defender to detect the existence of the attacker, so the lifetime of the attacker decreases and the expected costs diminishes accordingly. This also implies the increase of the blocking threshold $\tilde{p}$. In summary, if the attacker is non-strategic and the defender uses the blocking strategy accordingly, then the increase of the maximum attack capacity M eventually harms the attacker's profit and makes attacker to be more *detectible* to the defender.

3.3.3 Case of Strategic Attacker vs. Naïve Defender

In the previous case, (originally) non-strategic attacker will realize that her expected profit vanishes as M increases. Then she will naturally consider the defender's blocking strategy and choose attacking intensity strategically. Therefore, we now assume that the attacker is strategic, but the defender does not realize that the attacker is strategic. Still, the defender updates the suspicion level, but the adjustment equation (3.2.3) is driven by the assumption

that the attack intensity is always M. To be specific, the $(q_t)_{t\in[0,\infty)}$ is the solution of the following SDE,

$$dq_t = \frac{q_t(1-q_t)M}{\sigma^2}\Big(dY_t - q_t M dt\Big) \quad \text{with} \quad dY_t = \Delta_t 1_{\{\theta=1\}}dt + \sigma\, dW_t, \tag{3.3.4}$$

where $(\Delta_t)_{t\in[0,\infty)}$ is the attackers possible strategy. Then, (3.3.4) will not produce the filtering equation (3.2.2) if the attack intensity Δ_t is different from M (the defender's guess).

Proposition 3.3.2. *In the case of strategic attacker vs. naïve defender*[24]*, the maximum expected profit of attacker $V_3(q_0)$, and the expected cost of defender $C_3(q_0)$ have following asymptotic behavior:*

$$\lim_{M\to\infty} V_3(q_0) = \infty, \quad \lim_{M\to\infty} C_3(q_0) = \infty. \tag{3.3.5}$$

The explicit formula for C_3 and V_3 are given in the appendix.

Proposition 3.3.2 says that as $M \to \infty$, the expected costs of the defender also goes to ∞. Here is an intuitive explanation. Similarly as in the previous case, for bigger M, q_t moves toward the true state of θ more quickly. Now the 'strategic' attacker makes the situation quite different from the previous case. Recall that $\tilde{p}$ is the optimal blocking threshold in Proposition 3.3.1 and a is a constant in (5.0.1). For large enough M, we have following explicit formula for optimal attack intensity $\tilde{\alpha}$,

$$\tilde{\alpha}(q_t) = \begin{cases} M, & \text{if } \ q_t \in [0, \tilde{q}^*] \\ 0, & \text{if } \ q_t \in (\tilde{q}^*, \tilde{p}] \end{cases} \quad \text{where} \quad \tilde{q}^* = \frac{1}{1+(\frac{1+a}{2a^2})^{\frac{1}{1+2a}}(\frac{1-\tilde{p}}{\tilde{p}})}. \tag{3.3.6}$$

[24]This defender is 'naive' in the sense that she believes that the attacker is non-strategic and always chooses $\Delta_t = M$.

In words, the attacker chooses not to attack at all when the suspicion level q_t is relatively high ($q_t > \tilde{q}^*$), then the suspicion level will quickly drop with high probability. The attacker resumes the malicious activity when q_t is small enough ($q_t \leq \tilde{q}^*$). By exploiting defender's naïveness, the strategic attacker can make expected profit to ∞ as $M \to \infty$.

3.3.4 Case of Strategic Attacker vs. Defender

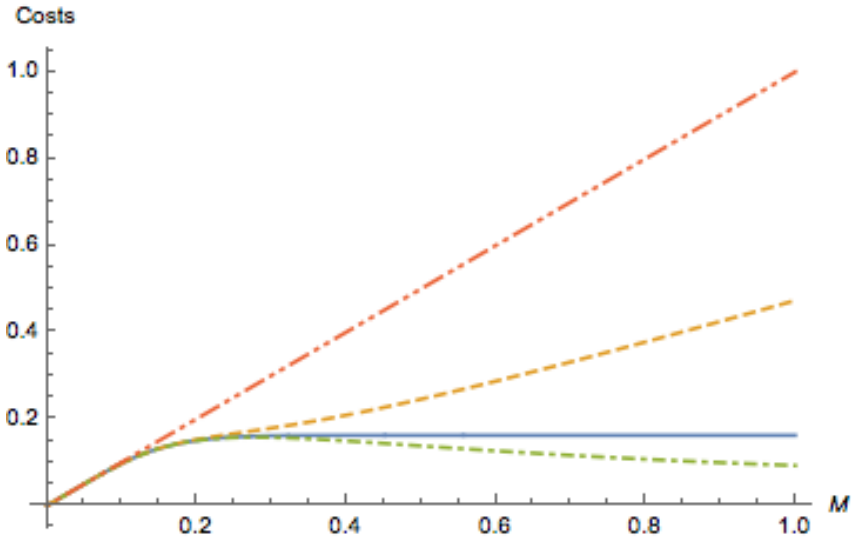
In the previous case, the naïve defender's expected cost blows up as $M \to \infty$. Therefore, (originally) naïve defender will naturally perceive the strategic behavior of the attacker and incorporates it to the adjustment of the belief process $(q_t)_{t\in[0,\infty)}$. This is the equilibrium concept in Definition 3.2.1.

Proposition 3.3.3. *The equilibrium in Definition 3.2.1 produces following asymptotic result for the expected cost of the defender $C_e(q_0)$ and the optimal blocking threshold p:*

$$\begin{aligned} \lim_{M\to\infty} C_e(q_0) &= \begin{cases} l_f(1-q_0)e^{-\varphi(1-\frac{\sigma q_0}{l_f\sqrt{\pi r}(1-q_0)})}, & q_0 \in [0,p) \\ l_f(1-q_0), & q_0 \in [p,1] \end{cases} \\ \lim_{M\to\infty} p &= \frac{l_f\sqrt{\pi r}}{l_f\sqrt{\pi r}+\sigma} \end{aligned} \tag{3.3.7}$$

Comparing the optimal attack strategies in Proposition 3.3.2 and Proposition 3.3.3, we observe that the attacker becomes more careful in Proposition 3.3.3 and *smooths out* her extreme behavior,[25] because she knows that

[25]The value of $\tilde{\alpha}$ in (3.3.6) is M or 0 only, and it is discontinuous in q_t. On the other hand, the optimal attack strategy α in Proposition 3.3.3 is continuous in q_t. See appendix for the expression of α.



$C_1\,(-\cdot\cdot-)$, $C_2\,(-\cdot-\cdot-)$, $C_3\,(---)$, $C_e\,(---)$
$q_0 = 0.01$, $r = 0.01$, $\sigma = 1$, $l = 1$

Figure 3.1: Expected costs for different cases, as functions of M

the defender is now considering the strategic behavior of the attacker.

In our equilibrium model (Definition 3.2.1) with strategic attacker and strategic defender, Proposition 3.3.3 implies that the expected costs and the optimal threshold stabilize as $M \to \infty$. Figure 3.1 illustrates that the relationships between M and the defender's expected costs for different cases, C_1, C_2, C_3 and C_e. From the aforementioned four cases, we derive the model implication for the requirements regarding the viability of the Internet-based society when the maximum attack capacity M is very high: (i) The defender's roles of updating suspicion level and blocking suspicious users are essential. (ii) The defender's updating procedure should rely on the right perception of

attacker (non-strategic or strategic).[26]

3.4 Equilibrium Analysis

In this section we examine the equilibrium behaviors of the attacker and defender in our game. As in Anderson and Smith (2013), the quantity $\frac{r\sigma^2}{M^2}$ plays an important role for the description of the equilibrium.

3.4.1 Blocking Threshold

In the equilibrium of our continuous time Bayesian game model, the most distinctive feature (compared to existing literature on insider trading or deception) is that the defender updates the suspicion level and terminates the game if the suspicion level is above certain threshold p, and the threshold is endogenously determined by the defender's cost-minimization problem.

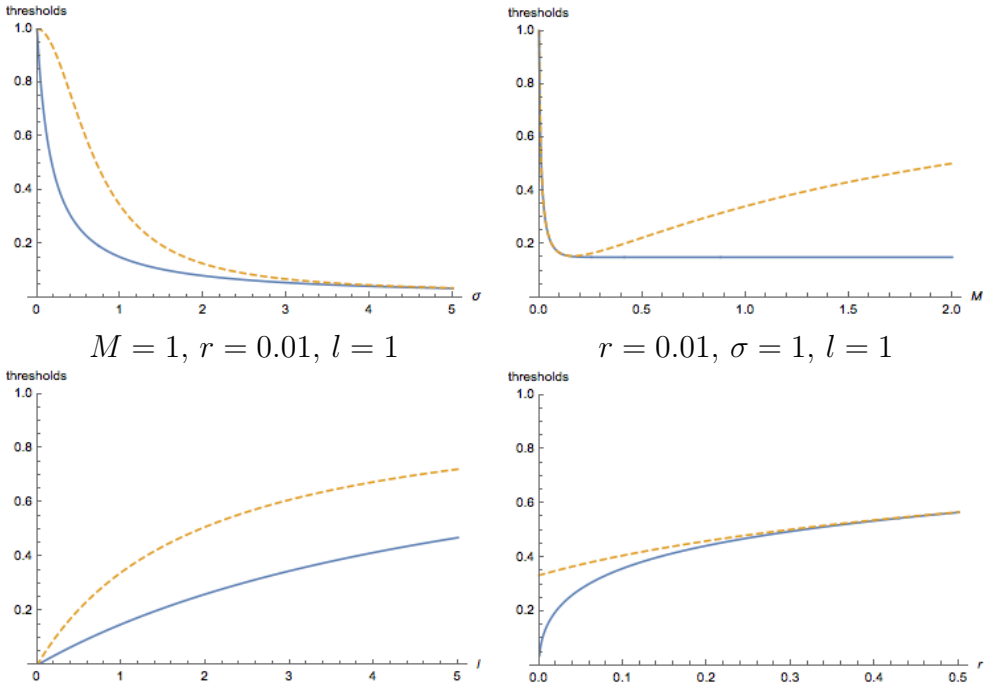
$$p = \begin{cases} \frac{(1+a)rl_f}{(1+a)rl_f+aM}, & \text{if } \frac{r\sigma^2}{M^2} > 1 \\ \frac{c\,l_f\sqrt{\pi r}}{c\,l_f\sqrt{\pi r}+\sigma}, & \text{if } \frac{r\sigma^2}{M^2} \leq 1 \end{cases} \tag{3.4.1}$$

where a, b, c are constants (depending on r, σ, M) defined in (5.0.1). For comparison, we also consider $\tilde{p} = \frac{(1+a)r\,l_f}{(1+a)r\,l_f+aM}$ in Proposition 3.3.1, the optimal blocking threshold in case the defender deals with a non-strategic attacker.

Proposition 3.4.1. *(1) $p \leq \tilde{p}$.*

(2) p decreases in M and σ, and increases in l_f and r.

[26] We may think $C_e - C_3$ as the defender's cost for the *underestimation* of the attacker, which goes to ∞ as $M \to \infty$.



$M = 1,\ r = 0.01,\ l = 1$ $\qquad$ $r = 0.01,\ \sigma = 1,\ l = 1$

$M = 1,\ r = 0.01,\ \sigma = 1$ $\qquad$ $M = 1,\ \sigma = 1,\ l = 1$

Figure 3.2: p (—) and $\tilde{p}$ (- - -) for varying l, r, σ and M

Figure 3.2 illustrates Proposition 3.4.1. The intuition for Proposition 3.4.1 (1) is obvious: The defender will be more careful and lower the blocking threshold when he encounters the strategic attacker, rather than non-strategic one. The intuition for Proposition 3.4.1 (2) is following: (i) Larger σ makes the observation $(Y_t)_{t\geq 0}$ more noisy and less informative for the defender. Accordingly, the attacker will be more aggressive since her identity is harder to be detected, and the expected cost of the defender will increase. Therefore, for larger σ, the defender will be more cautious and lower the blocking threshold. (ii) Larger l_f (*false alarm* cost) makes the defender more reluctant to block the user, therefore, induces higher equilibrium blocking threshold. (iii) Recall that T represents the random termination time of the game and has the exponential distribution $\mathbb{P}(T > t) = e^{-rt}$. If we increase r, then the defender has a better chance of blocking the user without concern of the false alarm cost. Therefore, larger r makes the defender to rely more on the random termination of game, and the equilibrium block threshold to increase. (iv) The M dependence is more subtle than the others. We consider non-strategic attacker case first. If M increases, than the non-strategic attacker's instantaneous profit increases (downward effect for $\tilde{p}$) but the defender updates the suspicion level more sensitively on the signal (see (3.3.3)), i.e., the attacker's identity is more revealing (upward effect for $\tilde{p}$). These upward and downward effects on $\tilde{p}$ can be seen in Figure 3.2 for varying M, first decreasing then increasing. When M is large enough, the existence of the non-strategic attacker is very 'revealing'. In contrast to $\tilde{p}$, p is monotonically decreasing on M. From

this observation, we deduce that for large enough M, the strategic attacker refrains herself from aggressive actions and mitigates the revealing effect. This observation is consistent with the attacker's behavior in equilibrium (see (2) in Proposition 3.4.2).

Observe that the gap between p and $\tilde{p}$ increases in M. This implies that when the maximum attack capacity is high, it is important for the defender to notice that the attacker is strategic. Otherwise, if the defender naïvely assumes that the attacker is non-strategic, then he will choose blocking threshold much higher than p (the truly optimal one) and will suffer very high expected costs (see Figure 3.1).

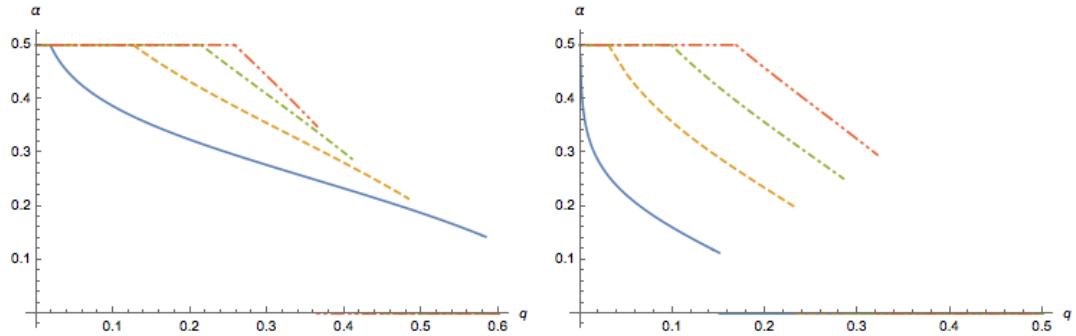
3.4.2 Attacker's Strategy

The attacker dynamically optimizes the attack intensity to maximize the expected profit, under the consideration that the defender updates the suspicion level by the signal process. The explicit expression of the equilibrium implies the following property of the optimal strategy of the attacker.

Proposition 3.4.2. *(1) If* $\frac{r\sigma^2}{M^2} \geq 1$*, the attacker chooses maximum attack intensity, i.e,* $\alpha = M$*, all the time regardless of the suspicion level.*

(2) If $\frac{r\sigma^2}{M^2} < 1$*, the attacker chooses maximum attack intensity* M *when* $q_t \leq q^*$*. After* q_t *exceed* q^**, the attack intensity gradually decreases as* q_t *increases. The expression of* q^* *is in* (5.0.1)*.*

Figure 3.3 describes Proposition 3.4.2. This behavior of the attacker is



$\sigma = 0.4$ (—), 0.6(− − −),
0.8 (-·-·-), 1.0 (-···-··-)
$M = 0.5$, $r = 0.1$, $l = 1$

$r = 0.01$ (—), 0.03 (—),
0.05 (-·-·-), 0.07 (-···-··-)
$M = 0.5, \sigma = 1$, $l = 1$

Figure 3.3: Graphs of $\alpha(q)$

similar to that of Anderson and Smith (2013), in the sense that the attacker does *deception*: When the suspicion level q_t is high, the attacker reduces attack intensity to mitigate the increase of q_t. The key difference between our model and one in Anderson and Smith (2013) is that our defender terminates the game when q_t reaches the blocking threshold p, therefore, our attacker's behavior can be interpreted as sacrificing the current profit to extend the lifetime of the game.

Figure 3.3 also shows that the attack intensity increases as σ or r increases: (i) If there is more noise (larger σ), then it is easier for the attacker to hide her identity, so the attack intensity will be higher. (ii) If we increase r, then there is more chance for the random termination of game, which makes the attacker's *deception* less valuable. Therefore, the attacker will focus more on the current profit and be aggressive as r increases.

3.4.3 Defender's Adjustment of Suspicion Level

The defender updates q_t by (3.2.7) in equilibrium, and q_t becomes the belief of the defender (i.e, satisfies (3.2.2)). In the adjustment equation (3.2.7) for q_t, the sensitivity of the movement of q_t with respect to the signal process Y is

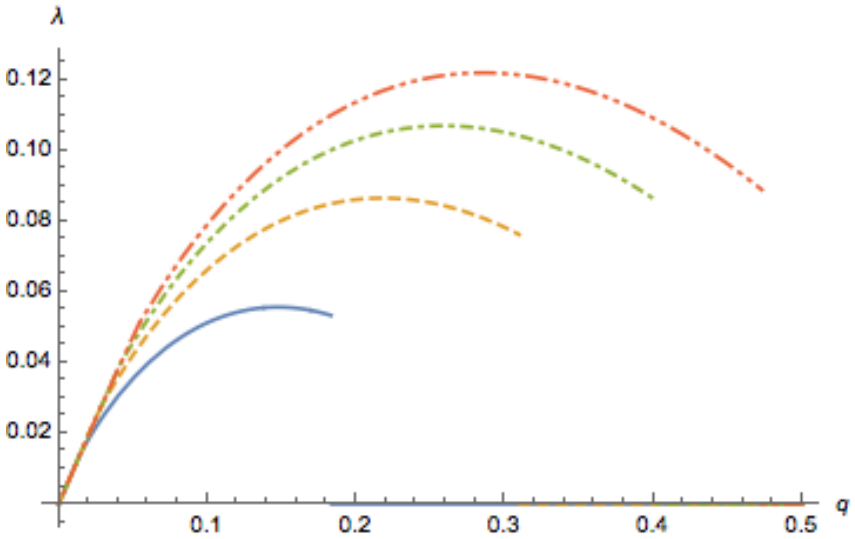
$$\lambda(q_t) := \frac{q_t(1-q_t)\alpha(q_t)}{\sigma^2}. \tag{3.4.2}$$

The following proposition implies that the update of the suspicion level q_t becomes less sensitive to the signal dY when q_t approaches the blocking threshold, or the false alarm cost l_f decreases.

Proposition 3.4.3. *(1) If* $\frac{r\sigma^2}{M^2} < 1$*, then* $\lambda'(p) < 0$.
(2) λ *is increasing in* l_f.

Figure 3.4 illustrates Proposition 3.4.3. Proposition 3.4.3 (1) implies that λ decreases on q near p. Here is an economic intuition. If $\frac{r\sigma^2}{M^2} < 1$, then Proposition 3.4.2 implies that the attacker will be less aggressive when q_t is close to p. In other words, the attacker's portion α becomes relatively small in the signal dY, and the signal becomes less informative for the defender. Therefore, when the suspicion level is close to the blocking threshold, both attacker and defender become *less active*.

We also give an intuitive explanation for (2) in Proposition 3.4.3. According to Proposition 3.4.1, p is increasing in l_f, hence it is enough to explain why $\lambda(q_t)$ increases in p. For higher blocking threshold p, the attacker will be



$l = 0.4$ (—), 0.8 (- - -), 1.2 (-·-·-), 1.6 (-···-···-)
$M = 1$, $r = 0.1$, $\sigma = 1$

Figure 3.4: Graphs of $\lambda(q)$

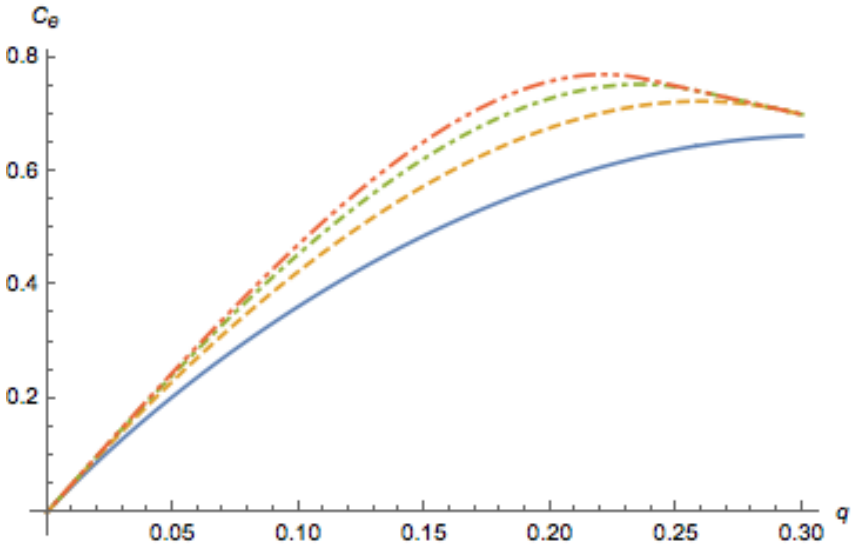
more aggressive since it is harder to be blocked. Then the signal Y will be more informative for the defender, so λ will be bigger.

3.5 Business Model for the ISP - Managed Security Service with Warranty

We claim that ISPs should provide MSSW for the clients who cannot afford a state of the art, in-house security system and cybersecurity experts. In this way, the liability of ISPs becomes a financial motive for strategic defense, as in our game model. The ISP will engage in more efficient defense to reduce the costs, and the expected societal costs related to cyberattacks will decrease accordingly. This produces a win-win situation for the clients, ISPs, and society.

The game model in this paper supports the claim that ISPs are in the suitable position to take the role of defender. The defender reduces the expected costs by strategically blocking the user based on the observation of the signal process. The "observe & update" role of the defender can be more effectively performed by ISPs than individual hosts, because ISPs have collective knowledge of the state of the Internet. For instance, the defender need to assign the initial suspicion level q_0. The following proposition can be used to estimation q_0.

Proposition 3.5.1. *For $0 < q_0 < p$, the probability that the defender eventu-*



$\sigma = 1.0\,(—),\ 1.5(- - -), 2.0\,(\text{-}\cdot\text{-}\cdot\text{-}),\ 2.5\,(\text{-}\cdots\text{-}\cdots\text{-}),$
$M = 0.5,\ r = 0.1,\ l = 1$

Figure 3.5: Graphs of $C_e(q)$

ally blocks a user is given by

$$\mathbb{P}(\tau_p < \infty) = \frac{q_0}{p}. \tag{3.5.1}$$

According to (3.5.1), $q_0 \approx p \cdot$ (ratio of blocked users). To make this calibration more accurate, the defender is supposed to be in the position to play multiple games with different users. Naturally, ISPs are in the optimal position for such tasks since dealing with multiple entities is their original job.

In Figure 3.5, we observe that the expected costs of the defender increase over the noise term σ. This means that ISP with better filtering ability (small σ) can reduce the MSSW service fee and attract more customers.

Even though some ISPs including AT&T and Verizon are providing MSS, the market for the service is very limited and underdeveloped. To explain this situation, we extend our game model to include the *monitoring cost*. To be specific, we modify the defender's cost minimization problem (3.2.9) in Definition 3.2.1 the following:

$$\tau_p \in \arg\min_{\tau \in \mathcal{T}} \mathbb{E}\Big[\Big(\int_0^{T\wedge\tau} \alpha(q_t)dt\Big) \cdot 1_{\{\theta=1\}} + \int_0^{T\wedge\tau} l_s\, dt + l_f \cdot 1_{\{\theta=0,\, \tau<T\}}\Big], \tag{3.5.2}$$

where the constant $l_s \geq 0$ represents the monitoring cost.[27] If the monitoring cost l_s is too high, it is better not provide such service.

By the same way as in Theorem 3.2.2, we prove that there exists an equilibrium if $l_s < r\ l_f$. The following result implies explanation for the premature state of the MSSW market.

Proposition 3.5.2. *Assume that $l_s < r\ l_f$. Then there exists an equilibrium in Definition 3.2.1 with the defender's cost minimization problem* (3.5.2)*. The defender's equilibrium expected cost is less than $\frac{q_0 M}{r}$ (cost without defense) if (i) M is large enough, or (ii) l_s is small enough.*

Proposition 3.5.2 indicates that the MSSW is profitable if the attacker have large attack capacity or less chance of random termination, or the monitoring cost is low. We expect this business to thrive, because (i) the attack capacity is continuously increasing due to the expansion of the IoT devices

[27] Need explanation here, observing the stream and do anomaly detection requires some costs.

and network capacity, and (ii) the monitoring cost is expected to be lowered by increased computing power.

3.6 Concluding Remarks

Cybersecurity is recognized as one of the most critical societal challenges as the society heavily relies on the cyber infrastructure. We argue that this suboptimal cybersecurity issue can be addressed by enhancing our understanding on the strategic interactions among the stakeholders. In the cybersecurity context, we develop a game model between a cyber attacker and defender, and fully analyze the equilibrium interaction between the players. Our game model is the first to include the optimal termination of the game, with asymmetric information and continuous time Bayesian updates. We find that, in case the cyber defender does not properly cope with strategic attackers, the defender's expected cost can explode as the attack-intensity bound rapidly increases. This observation suggests that the defender's strategic role of blocking suspicious users is necessary for the viability of the Internet-based society. We provide a method to empirically calibrate an important model parameter – initial suspicion level – and claim that ISPs can effectively perform this task as cyber defenders. Extending the model with a monitoring cost, we provide sufficient conditions that MSSW business model becomes profitable for ISPs.

In a broader context, our research can contribute to the cyber insurance market. Unlike traditional insurances (e.g., auto insurance or health

insurance), cyber insurance has a unique nature in that there exists deliberate, evolving adversaries. Moreover, due to lack of proper data sharing policies, it is hard to find comprehensive historical data on cyber risk. As a result of these factors, it is extremely challenging to construct proper cyber insurance policies. We suggest that the MSSW providers can actively monitor their customers' network activity, and assess cyber risk by computing the expected cost.[28][29] In addition, the MSSW providers can create synergistic values by tightly combining the roles of protecting customers and lowering the associated risk, which third party insurance providers may not achieve. In this cyber insurance framework, our future direction is to generalize our cybersecurity game model by incorporating time-dependent noise size (periodic patterns of noise) and multi-dimensional signal processes (traffic from multiple channels).

[28]Such individual monitoring is widely used by auto insurance companies. **Threewitt, Cherise and Vincent, John M.** 2018. "How Do Those Car Insurance Tracking Devices Work?" *U.S. News.* February 26. https://cars.usnews.com/cars-trucks/how-do-those-car-insurance-tracking-devices-work.

[29]**Symantec Corporation**. 2016. "Internet Security Threat Report." Volume 21, April, https://www.symantec.com/content/dam/symantec/docs/reports/istr-21-2016-en.pdf.

Chapter 4

Cyber Incident Prediction Using Public Cyber Risk Data and Disclosed Risk Factors

4.1 Introduction

Security incidents do not happen every day or every month, but it can cause a critical damage once it happens. Research conducted by the National Cyber Security Alliance[1] found that 60% of SMBs went out of business in 6 months after a data breach. So it is important to predict future security incidents and take proactive actions. However, predicting an organization's future data breach is a very challenging problem. While organizations can use their internal monitoring systems to detect and prevent cyber risks, it is hard for external entities such as investors or insurers to predict such incidents. Also, there are many small businesses and non-IT companies who are not capable of measuring their own security risks.[2] Given the background, the goal of this research is to provide a cyber breach prediction method leveraging relatively easily accessible datasets without the need of internal data. To build such model, it is necessary to have a data to measure security status and previous

[1] www.staysafeonline.org

[2] https://www.theguardian.com/small-business-network/2016/feb/08/huge-rise-hack-attacks-cyber-criminals-target-small-businesses

data breach history [52]. In addition to these security related information, it would be beneficial to have statistical, financial information of a firm and its own understanding of existing risks. Thus, we use four different types of datasets, which are (1) malicious activity data such as spam/phishing/botnet activity, (2) security breach record data, (3) disclosed annual risk reports from organizations, and (4) Compustat dataset[3] for public firms. Note that data sets (1) and (2) are available to public and private firms, as well as non-profit organizations, and (3), (4) are only available for public firms.

We collect data on malicious activities such as spam, phishing and botnet activity which can be used as a proxy of security posture. We use this data as a proxy of defense level for the target organizations. Since these activities are relatively easy to mitigate once noticed,[4] we assume that companies with large outgoing spam or phishing activities for a long period may have lower security protection level, or less investment in externality issues.[5] In addition, we use cybersecurity incident records from Privacy Rights Clearinghouse (PRC),[6] VERIS[7] Community Database (VCDB),[8] and Hackmageddon. PRC data includes confirmed data breach incidents recorded in state government notifications and various media sources. VCDB is a public repository of breach data including community shared security incident information. Lastly,

[3]Standard and Poor's Dataset. (2011)

[4]Simple malware removal steps provided by Norton:
https://us.norton.com/internetsecurity-malware-how-to-remove-malware.html

[5]More details are available in Chapter 2.

[6]https://www.privacyrights.org/

[7]Vocabulary for Event Recording and Incident Sharing

[8]http://veriscommunity.net/vcdb.html

Hackmageddon provides more hacking oriented records.

Also, we gather 10-K annual report from the Securities and Exchange Commission (SEC), and Compustat dataset from S&P Global Market Intelligence.[9] Based on economics and accounting literature, investors have been using public financial statements for firm valuation [9, 11]. Also, studies found that acquired data through the statements reduces information asymmetries between firms and investors in the stock market [49]. In the recent years, due to a number of tremendous data breach events, investors started having great interests in the security area.[10] Moreover, in 2011, the SEC notified public companies that cybersecurity incidents and security risks in their IT systems should be reported through public disclosures in a section called Item 1A.[11] We analyze the risk documents using latent Dirichlet allocation (LDA) topic modeling to extract a probabilistic distribution over a set of underlying topics of the statement [69, 70]. To explore the effectiveness of these multidimensional data features, we construct various classification models using support vector machine (SVM) [41], k-nearest neighbors (kNN) [27], random forests (RF) [51], deep feed forward neural networks (FFNN) [75], and feed forward neural network with dropout (dropoutNN) [72]. The best resulting classifier using dropoutNN marks an AUC score of 76.04.

[9] Both datasets are publicly available.

[10] https://finance.yahoo.com/news/investors-pouring-money-cyber-security-125500156.html

[11] General Item 1A was formalized in 2005: https://www.sec.gov/fast-answers/answersreada10khtm.html, and the security risk became mandated from 2011: https://www.sec.gov/divisions/corpfin/guidance/cfguidance-topic2.htm

There are a few related works using machine learning algorithms to predict security incidents. Wang et al [82] introduced using text mining on public filings to predict future breach notifications and stock market reactions. Soska et al. [71] introduce website breach prediction based on textual and structural data, including web contents and HTML tags. They use C4.5 decision tree algorithm to predict whether a website will turn malicious in the future. Liu et al. [52] present a way of predicting organizational level breaches from security posture data and incident data using random forest method. The result shows that security features by themselves are not as powerful as collective feature set in the prediction. The authors reported overall accuracy of 90% using RF model, but the result is based on modified training/validation set by undersampling the dataset to artificially generate balanced dataset with equal number of positive and negative class, which is not exactly projecting the real world distribution. We propose an extension of this work by extending total period of data (1.5 years vs. 7 years), adding text analysis data from 10-K reports, and using sophisticated deep learning algorithms. We use more realistic evaluation methods and test sets to make our study closely reflect the real world problem. Under our setting, Liu et al [52] method gives an AUC score of 64. On the other hand, one of our prediction models using dropoutNN marks an AUC score of 76, which is a big improvement over the previous work.

Table 4.1: Summary statistics of the malicious activity data between 2012 and 2017. (unit of observation: firm-year)

		PSBL			CBL			bots
		volume	host	host_dates	volume	host	host_dates	
total	mean	2828.57	312.58	12.70	5097976.61	10449.86	334.29	5.28
	std	34911.46	2798.74	39.60	89317563.98	96822.41	1587.36	15.53
	max	770134.00	52302.00	361.00	2282611911.00	1604688.00	21671.00	175.00
	min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	#	854						
breached	mean	1539.86	312.69	18.30	360467.96	5719.99	378.23	6.03
	std	9497.02	1834.60	41.26	2646971.70	34595.45	1649.86	17.18
	max	23772.00	14480.00	232.00	22557562.00	327450.00	15504.00	16.00
	min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	#	110						
non breached	mean	3019.11	312.56	11.87	5798414.72	11149.17	327.79	5.17
	std	37220.88	2914.35	39.28	95667498.60	102858.43	1577.80	15.27
	max	770134.00	52302.00	361.00	2282611911.00	1604688.00	21671.00	175.00
	min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	#	744						

Table 4.2: Summary statistics of log transformed ($log(num + 1)$) malicious activity data between 2012 and 2017. (unit of observation: firm-year)

		PSBL			CBL			bots
		volume	host	host_dates	volume	host	host_dates	
total	mean	1.34	1.07	0.91	3.34	3.58	3.34	0.80
	std	2.32	1.92	1.47	3.95	2.60	2.17	1.02
	max	13.55	10.86	5.89	21.55	14.29	9.98	5.16
	min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	#	854						
breached	mean	1.91	1.60	1.35	3.49	3.70	3.44	0.83
	std	2.60	2.22	1.70	3.72	2.68	2.26	1.11
	max	11.28	9.58	5.45	16.93	12.70	9.65	4.97
	min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	#	110						
non breached	mean	1.25	0.99	0.84	3.32	3.56	3.33	0.79
	std	2.27	1.86	1.43	3.99	2.59	2.16	1.01
	max	13.55	10.86	5.89	21.55	14.29	9.98	5.16
	min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	#	744						

Table 4.3: Number of firm-years and IP host counts of top 10 botnets.

botnet	firm-year	# of IP addresses
Cutwail	164	9524
Kelihos	119	7100
Conficker	116	9634
lh	89	8033
Gozi	83	8316
Darkmailer2	64	6273
Sendsafe	51	7042
Tinba	51	5512
Necurs	46	4179
KINS	47	3896

4.2 Data

4.2.1 Security Posture Data

We use spam/phishing mail emissions and botnet activity data from organization owned IP addresses as a measure of security mismanagement. This kind of carelessness could be a starting point of a serious data breach in the near future. About 23% of data breach incidents started from a piece of malware, 92% of those were from email attachments in 2017.[12] In addition, more than 90% spam mails are sent from bot-compromised hosts [46], which are undetected by the host owners. Since most of the spambots can be detected and exterminated by anti-malware software,[13] continuous spam emission and/or botnet existence could be a sign of security mismanagement [**?**].

[12] 2018 Verizon Enterprises annual Data Breach Investigations Report (DBIR) https://enterprise.verizon.com/resources/reports/dbir/

[13] https://usa.kaspersky.com/resource-center/preemptive-safety/antivirus-malware-detection

From CBL and PSBL raw data, we extract total outbound spam mails, number of days with at least one spam mail, and number of IP addresses (host counts) involved in spam activities from a firm. Also, CBL provides botnet information which were used to send out malicious emails. From January 2011 to December 2018, we collected a total of 369,585,209,511 email records from CBL, and 492,792,524 from PSBL with origin IP addresses. Those IP addresses are mapped to ASNs, then to the owner organizations using WHOIS queries at our data processing stage. Note that companies who had zero spam/phishing emission in both CBL and PSBL are not included in this study. As a result, we identified a total of 10,473 U.S. organizations who experienced episodes of spam and phishing issues. In addition to the malicious email data, we use botnet information to see if particular bots are more correlated with security breach events. Table 4.1 shows detailed summary statistics of CBL and PSBL data.

A botnet is a network of Internet connected machines infected by one or more malware called bots, controlled by a botmaster. Botmasters use command and control server to send out attack orders to botnet hosts to send out spam/phishing emails, DDoS attacks, and infecting other machines to increase the botnet. CBL provides botnet information along with spam data if identifiable. Table 4.3 shows top 10 botnet information from CBL dataset. Among all, Cutwail, Kelihos, and Conflicker botnets are the most notorious botnets. Cutwail botnet is a spam-botnet founded around 2007, infected by Pushdo Trojan malware. In 2009, based on analysis of MessageLabs, it was capable of

sending out 51 million spam mails per minute, which was 46.5% of world wide spam volume [74]. In 2010, Cutwail was used for DDoS attacks on 300 major sites including CIA, FBI, Twitter, and PayPal.[14] A notorious Kelihos botnet was first discovered around December 2010. It is mainly used for spreading spams, but has various versions capable of email harvesting, DDoS attack, click fraud, and Bitcoin mining. The operator of Kelihos botnet was arrested on 2017 by Spanish authorities.[15] Conflicker is a computer worm targeting Microsoft Windows systems, first discovered in November 2008.

Table 4.4: Summary tables of Privacy Rights Clearing House data.

type of organization	counts	%
healthcare/medical	4045	49.5
educational	818	10.0
government	775	9.5
financial/Insurance	750	9.2
retail/merchant	618	7.6
nonprofits	118	1.4
others	1045	12.8
total	8169	100

(a) type of organizations

type of breach	counts	%
hacking or malware	2432	29.8
unintended disclosure	1704	20.9
physical loss	1687	20.7
portable device	1172	14.3
insider	608	7.4
stationary device	249	3.0
card fraud	68	0.1
unknown	249	3.0
total	8169	100

(b) breach types

incidents	total records	mean	std	max	min
8170	9,691,000,571	1,186,314.18	32,084,227.89	2,147,483,647	1

(c) summary statistics

[14]https://www.itbusiness.ca/news/pushdo-botnet-pummels-more-than-300-web-sites/12341

[15]https://www.justice.gov/opa/pr/alleged-operator-kelihos-botnet-extradited-spain

4.2.2 Security Incident Data

Based on security incidents that can serve as symptoms of security mismanagement, our goal is to predict more critical cyber incidents such as a large scale data leakage or taking down servers due to denial of service attacks. These problems can damage victims on both financialand non-financial aspects such as bad reputationand legal issues [63, 66]. State level data security breach notification law has been active since 2002 starting from California.[16] The Personal Data Notification & Protection Act proposed by Barak Obama in 2015 includes national data breach notification standard, which requires all U.S organizations to establish a 30-day notification from the discovery of a breach. Privacy rights clearing house (PRC) collects state/federal level security breach reports and publicize them in their website. We collected 8169 unique data breach incidents which involve 9.7 billion individual records from 2005 to 2018. Table 4.4 shows type of breach, type of organization, and summary statistics of PRC data. It shows about 49.5% of incidents are from healthcare and medical field, and about 30% of incidents are from hacking or malware infection.

VERIS Community Database (VCDB) includes a broad range of public community effort to gather cyber incident reports, maintained by one of the biggest telecommunication/network company Verizon.

Hackmageddon is an independent security web-blog which collects various cyber incident reports. From the website, we collected 1966 incidents from

[16]California Security Breach Information Act (SB-1386)

Table 4.5: Type of breaches of Hackmageddon data.

type	counts	%
malware	1001	50.1
account	292	14.9
targeted attack	264	13.4
DDoS	108	5.5
defacement	78	4.0
SQL injection	73	3.7
DNS hijacking	15	0.8
brute force	14	0.7
others	121	6.2
total	1966	100

September 2015 to December 2017. In Table 4.5 shows detailed incidents and counts for each category. Comparing to PRC data, Hackmageddon specifically focuses on "hacking or malware" category with more depth. There are 2 overlapping incidents between PRC and Hackmageddon, and 18 between PRC and VCDB. These are only counted once per each.

4.2.3 Annual Risk Report

For better information security protection, it is important for organizations to understand their own risks and properly address the issues. U.S. Securities and Exchange Commission (SEC) mandates all public firms to submit annual Form 10-K report. The report contains a comprehensive overview of the company's business and audited financial condition. It has fifteen sections called "Items" and we are particularly interested in Item 1A "Risk Factors,"

mandated from 2005. [17] This item is where the filing companies describe their own risks. The description only focus on the risk itself, not how the companies handle it. Our hypothesis is that an organization's own descriptions of the risks may reveal their focus and direction in the risk management, and their awareness of the problem. For example, if more security issues are described in the risk report, they may put more resources to overcome the risk. On the other hand, the risk report may reveal they have more weaknesses on the system and attract attackers.

From publicly available EDGAR[18] database, we were able to collect 10-K filings from 6,144 unique firms across 26,197 firm-years beginning of 2006. The filings include 1.73 million risk factors in 64,385 annual reports.[19] Among them, there are 38,622 reports discussing IT-related risks. We analyze 10-K documents using LDA topic modeling, which is an unsupervised learning algorithm for natural language processing. LDA discovers latent topics and related keywords for each topic from a large collection of documents. A document may be viewed as a mixture of various topics based on the including words. Figure 4.1 visualizes popular topics in our dataset. Positive counts are the number of breached firm-year observations which has a particular topic in their risk report, and negative counts are for the non-breach observations. Keywords distributed in each topic model is described in Table 4.6. Among

[17] https://www.sec.gov/Archives/edgar/data/1289850/000110465906017294/a06-2620_110k.htm

[18] https://www.sec.gov/edgar.shtml

[19] Processed data provided from [70]

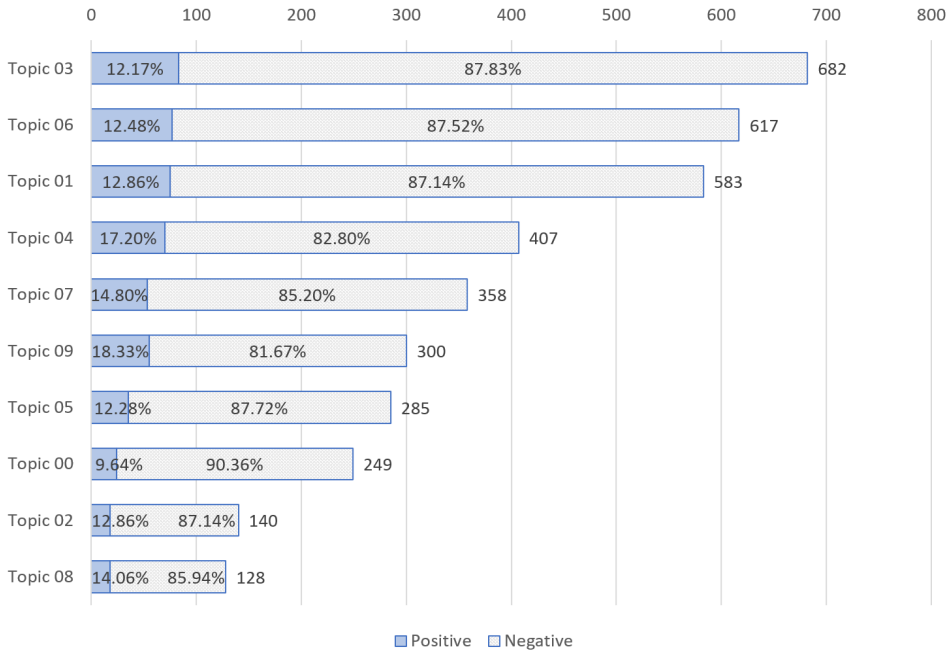


Figure 4.1: Popular topics in 10-K item 1A between 2012 and 2017 in 10 topic model.

854 total samples, 682 describes technology related risks (topic 3) in their Item 1A, followed by data/network services related risk (topic 6) and foreign tax/regulations (topic1). Topic 9 (medical topic) has the highest data breach rate closely followed by topic 4 (financial topic).

4.2.4 Compustat Data

While 10-K data contains text based description, Standard & Poor's Compustat data provides financial, statistical, and market information on firms publicly traded in U.S. stock markets. This set is heavily used in finan-

Table 4.6: Top keywords in each topic (10 Topics)

topic	description	keywords
topic 00	income/tax/investment	income, investment, properties, tax, property, assets
topic 01	laws/regulations	foreign, tax, government, laws, regulations, contracts
topic 02	oil/natural gas	gas, oil, natural, production, prices, regulations
topic 03	technology/system management	technology, revenue, customer, property, management, systems
topic 04	credit/financial	loans, interest, losses, insurance, risk, credit
topic 05	stock/share	shares, securities, trading, directors, investment
topic 06	security/data network	service, data, internet, information, laws, network
topic 07	credit/market	credit, material, condition, effect, impact, conditions
topic 08	drug/FDA	clinical, candidates, regulatory, development, approval
topic 09	medical/healthcare	health, healthcare, state, federal, care, laws, medical

cial researches, such as predicting business failure [12, 14], bankruptcy [44], and stock-market returns [32]. Our intuition is that it is possible for profit oriented hackers may use some financial information to select a target for profit maximization for the same amount of effort. Among 645 total variables in Compustat, we extract 583 variables with numerical values which are not date or ID numbers, and SIC code dummy variable at the end. SIC code enables us to analyze organizations based on industry types. Table 4.7 shows detailed industry distributions, and proportions of breached and non-breached firms. Blue colored cells indicate medicare/healthcare related industry codes, and gray ones indicate software/data/network related industry codes. We notice a clear separation of these two categories, former has much higher breach rates compare to the latter.

4.2.5 Combining Datasources

We combine the individual datasets base on the firm-year index.[20] First, we use a company identification code called DossierID for data sources with IP addresses which can be mapped to ASN, and organizational level (CBL, PSBL). All firms included in this study has at least one spam or phishing emission record. Next, for datasource without organization identifications, we use string comparison algorithm called Fuzzy matching [8] to match records. All mappings are manually verified after automated matching. As a

[20]The reasons for using yearly data are as follows: 1. Annual 10-K reports include more comprehensive information compare to quarterly 10-Q reports, 2.most of discovery dates have gaps with actual breach date, and 3. low sparsity of security incidents data points.

Table 4.7: Number of breached and non-breached firm-year for each SIC code.

SIC	description	breach	non breach	total	%
8082	SERVICES-HOME HEALTH CARE SERVICES	3	1	4	75.00
7841	SERVICES-VIDEO TAPE RENTAL	3	1	4	75.00
7200	SERVICES-PERSONAL SERVICES	2	2	4	50.00
6324	HOSPITAL & MEDICAL SERVICE PLANS	4	6	10	40.00
6141	PERSONAL CREDIT INSTITUTIONS	5	8	13	38.46
8090	SERVICES-MISC HEALTH & ALLIED SERVICES, NEC	5	8	13	38.46
5912	RETAIL-DRUG STORES AND PROPRIETARY STORES	1	2	3	33.33
5700	RETAIL-HOME FURNITURE, FURNISHINGS & EQUIPMENT STORES	1	2	3	33.33
6361	TITLE INSURANCE	2	4	6	33.33
4953	REFUSE SYSTEMS	2	5	7	28.57
5331	RETAIL-VARIETY STORES	2	6	8	25.00
7997	SERVICES-MEMBERSHIP SPORTS & RECREATION CLUBS	1	3	4	25.00
5141	WHOLESALE-GROCERIES, GENERAL LINE	2	6	8	25.00
6022	STATE COMMERCIAL BANKS	4	14	18	22.22
2300	APPAREL & OTHER FINISHD PRODS OF FABRICS & SIMILAR MATL	1	4	5	20.00
5500	RETAIL-AUTO DEALERS & GASOLINE STATIONS	2	8	10	20.00
8711	SERVICES-ENGINEERING SERVICES	2	8	10	20.00
8071	SERVICES-MEDICAL LABORATORIES	1	4	5	20.00
4833	TELEVISION BROADCASTING STATIONS	1	4	5	20.00
5651	RETAIL-FAMILY CLOTHING STORES	3	13	16	18.75
3826	LABORATORY ANALYTICAL INSTRUMENTS	1	5	6	16.67
6311	LIFE INSURANCE	4	21	25	16.00
7374	SERVICES-COMPUTER PROCESSING & DATA PREPARATION	5	29	34	14.71
8200	SERVICES-EDUCATIONAL SERVICES	3	19	22	13.64
8062	SERVICES-GENERAL MEDICAL & SURGICAL HOSPITALS, NEC	2	13	15	13.33
4931	ELECTRIC & OTHER SERVICES COMBINED	2	14	16	12.50
3845	ELECTROMEDICAL & ELECTROTHERAPEUTIC APPARATUS	2	14	16	12.50
4512	AIR TRANSPORTATION, SCHEDULED	2	15	17	11.76
2834	PHARMACEUTICAL PREPARATIONS	4	30	34	11.76
4841	CABLE & OTHER PAY TELEVISION SERVICES	3	27	30	10.00
3576	COMPUTER COMMUNICATIONS EQUIPMENT	3	27	30	10.00
5812	RETAIL-EATING PLACES	1	9	10	10.00
5311	RETAIL-DEPARTMENT STORES	1	10	11	9.09
7372	SERVICES-PREPACKAGED SOFTWARE	11	119	130	8.46
2836	BIOLOGICAL PRODUCTS, (NO DISGNOSTIC SUBSTANCES)	2	22	24	8.33
7373	SERVICES-COMPUTER INTEGRATED SYSTEMS DESIGN	2	24	26	7.69
6798	REAL ESTATE INVESTMENT TRUSTS	3	37	40	7.50
7370	SERVICES-COMPUTER PROGRAMMING, DATA PROCESSING, ETC.	10	132	142	7.04
3711	MOTOR VEHICLES & PASSENGER CAR BODIES	1	15	16	6.25
3674	SEMICONDUCTORS & RELATED DEVICES	1	53	54	1.85

result, 1060 incidents from PRC, 636 incidents from VCDB, and 79 incidents from Hackmageddon are mapped. 20 overlapping incidents are counted once. Lastly, DossierID - CIK (Central Index Key) pairs are mapped based on official organization name using Fuzzy matching again to unify all data together. Note that the CIK is used on the SEC's computer systems to identify corporations and individual people who have filed disclosure with the SEC[21]. Lastly, from 1809 total combined samples, we select organizations with Standard Industrial Classification (SIC)[22] codes which experienced security incident at least once since 2012. As a result, our final dataset contains 854 firm-year observations: 110 positive samples (with data breaches) and 744 negative samples (no breaches).

4.3 Prediction model

4.3.1 Feature set construction

We use four different sets of data features. As a primary set, we use security mismanagement data which includes annual spam/phishing volume, number of IP addresses with spam emission, total spam dates, and total number of bots detected. The second set is our unique feature, the risk topic model score. From the risk report, we use LDA topic model to learn what kind of risks are discussed and focused by an organization. For a k-topic model, k number of features are generated through LDA. We use 10, 20, 30, 50, 100

[21]https://www.sec.gov/edgar/searchedgar/cik.htm
[22]https://www.sec.gov/info/edgar/siccodes.htm

Table 4.8: Description of features used in the prediction model.

dataset	dimension	features
Security posture	7	CBL volume, CBL hosts, CBL dates, PSBL volume, PSBL hosts, PSBL dates, Number of identified bots
10-K risk factor	10, 20, 30, 50, 100	LDA topic modeling scores
Compustat	624	Numeric data + Industry code dummy
Botnet	626	botnet ID dummy

Table 4.9: Specifications of hyperparameters

model	hyperparameter	values
SVM	kernel function	radial basis function, linear, polynomial, sigmoid
kNN	# of neighbors	1, 3, 5, 7, 9, 11
RF	# of trees	10, 20, 30, 40, 50, 100
	max tree depth	1, 2, 4, 8, 16, 32, 64, 128
FFNN	# of layers	2, 3, 4, 5, 6, 7, 8
	# of nodes	8, 16, 32, 64, 128, 256, 512
DropoutNN	# of layers	2, 3, 4, 5, 6, 7, 8
	# of nodes	8, 16, 32, 64, 128, 256, 512
	dropout ratio	0.20

for k, and select best performing model at the training stage. The third feature set is the Compustat data which provides financial, statistical data. The fourth feature set provides the information about detected botnets and total counts in certain firm-year.

4.3.2 Prediction and Validation Settings

We use open source deep learning library Keras[23] for neural network models (FFNN, dropoutNN), and scikit-learn library[24] for other classification algorithms (SVM, Logit, RF, kNN) in Python. The dataset is split into two using stratified randomization, where the first portion is used for training (80%) and the other for testing (20%). In the training set, we use 5-fold cross validation to find optimal hyperparameter for each considered model. Table 4.9 describes the specifications of hyperparameters we tested.[25] To reduce possible bias in data splitting, we report the average test results over more than 20 different training/test splitsf. Note that we use stratified random sampling to maintain the same distributions of positive and negative samples. For the prediction evaluation metrics, we use precision ($\frac{TP}{TP+FP}$), recall ($\frac{TP}{TP+FN}$), f1 score ($2 \times \frac{precision \times recall}{precision+recall}$), accuracy ($\frac{FP+FN}{TP+TN+FP+FN}$), and AUC (area under roc curve) score.[26] In cybersecurity area and medical area [59], false negatives are much more expensive than false positives. For example, when a company is predicted to have a data breach in the future, more security investments will be made to prevent a future incident. On the other hand, when the predictor says there will be no breach in the company but a breach actually happens (false negative), there will be a tremendous loss. Therefore, we put more emphasis

[23]https://keras.io/
[24]https://scikit-learn.org/stable/
[25]Note that logit model does not require hyperparamteter tuning.
[26]TP: true positive, TN: true negative, FP: false positive, FN: false negative

on higher recall value[27] over precision score while maintaining highest possible AUC and f1 score.

Since our data is highly imbalanced (110 positive and 744 negative samples) with more importance on the minority class, we use two methods to overcome the issue [24, 39, 60] and properly train our models. First, we use a well known upsampling method called Synthetic Minority Oversampling Technique (SMOTE) [23], which is based on kNN algorithm[28] to generate artificial samples. SMOTE is applied at the training stage only on minority class (positive), and make balanced samples across the two classes. Second, for neural network models (FFNN, dropoutNN), we use weight balancing when computing the loss. In this way, the prediction model could learn better about minor, but more important class. We report that 0.82% weight on positive class and 0.18% on negative class produced the highest results.[29]

4.4 Results

4.4.1 Topic model vs. non-topic (temp title)

Figures 4.2 and 4.3 visualize prediction score (f1 and AUC) comparisons between the model with 10-K item 1A risk factor topic features and the model without topic features. The X axis represents model labels, and the Y axis

[27] We set the minimum threshold for the recall to be 60. For example, we ignored the highest f1 score with very high precision, but with less than 50 recall value. This is a common issue for imbalanced binary classification problems.

[28] There are a few different variations available using different internal algorithms [36, 37, 67].

[29] We tried 0.5, 0.6, 0.7, 0.8, and 0.9, and then 0.75, 0.78, 0.82, and 0.85.

Table 4.10: Comparisons of the best results for different prediction models with risk topic model.

model	compustat	bot	precision	recall	f1	AUC
Logit	no	no	21.92	60.91	32.17	65.43
SVM	no	yes	23.73	60.0	33.82	69.22
RF	yes	yes	24.11	62.73	34.64	71.33
kNN	no	no	30.83	63.64	41.50	71.64
FFNN	no	no	24.03	61.82	33.63	66.54
DropoutNN	no	no	32.03	61.82	42.20	76.03
Random	-	-	12.86	50.0	20.46	50.0

represents score. Brighter bars are results without risk topic features, and darker bars are scores with risk topic features. Numbers by the darker bars are the improvements for using the risk topic features. Note that all models are tuned with best performing hyperparameters in this comparison. Every single model is improved without exception, ranging 0.31 to 6.93 in f1 and 0.31 to 10.31 points in AUC. The result significantly improved in dropoutNN model (6.93 in f1 and 10.01 in AUC), followed by kNN (6.17 and 4.89) and SVM(4.62 and 7.33). Thus our prediction experiments show that the descriptions of risk factor in 10K Item 1A greatly help the cyber breach prediction results. We also find that there is a substitution effect with Compustat features. When risk topic features are used (Table 4.10), except RF model, Compustat is not used in training. However, when risk topic features are removed, results with Compustat data shows the best performance in Table 4.11. One possible reason is that the industry information appears both Compustat and risk reports as we can observe in Tables 4.6 and 4.7.

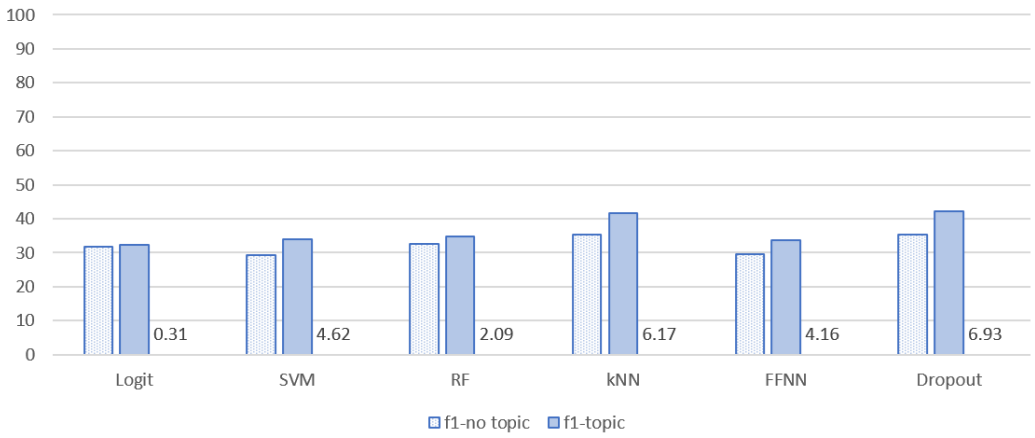


Figure 4.2: F1 score comparison between results with topic vs. without topic. Numbers represent improved scores by including risk topic features.

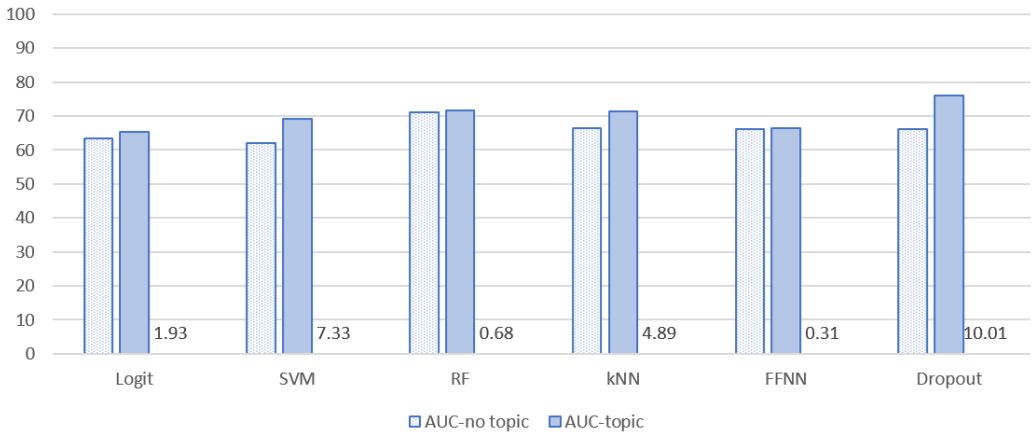


Figure 4.3: AUC score comparison between results with topic vs. without topic. Numbers represent improved scores by including risk topic features.

Table 4.11: Comparisons of the best results for different prediction models without risk factor topic model.

model	compustat	bot	precision	recall	f1	AUC
Logit	yes	no	24.39	46.36	31.86	63.5
SVM	yes	no	22.61	41.82	29.2	61.89
RF	yes	no	23.86	51.82	32.55	70.96
kNN	yes	no	27.29	50.91	35.33	66.44
FFNN	yes	no	19.18	63.64	29.47	66.23
DropoutNN	yes	no	27.01	51.82	35.27	66.02
Random	-	-	12.86	50.0	20.46	50.0

4.4.2 Feature importance

Now we investigate the relative importance of individual topics and other variables for the prediction performance. For each feature score measurement, we use 5-fold cross validation with same exact parameter (k=7) and fixed random seeds for test/validation splits, and report averaged prediction score. If the AUC goes lower without certain feature, it means that the feature plays a positive role for the prediction model. So the feature importance is defined as the difference of AUC score of the model with the target feature and the score of the model without the feature. In Figures 4.4 and 4.5, variables with higher scores are more important features for our prediction model. We find that the topic 09 (healthcare/medical) and psbl host dates mark the highest feature score. In addition, lower score of topic 03 (technology/system management) matches with the industry code distribution. It could be related to the industry code distribution of Table 4.7, as healthcare/medicare related SICs have high tendency of data breaches, and computer related SICs

have lower tendency. It provides extra supports for the substitution effects in Section 4.4.1.

4.4.3 Model comparisons

Now we analyze performance difference of prediction models. Table 4.10 reports the best results from each model. As baseline prediction models, we use support vector machine (SVM) and logit. Among the different models, dropoutNN model yields the best f1 and AUC scores (42.20 and 76.03)followed by kNN (41.50 and 71.64) and RF (34.64 and 71.33). Dropout is a generalization technique which can be applied to deep neural network models[72]. At each stage, individual nodes are randomly dropped out with probability p, so that a reduced network is left. This helps reducing interdependent learning amongst the neurons, which causes overfitting of training data. We measure generalized results by repeated cross validation, and dropoutNN clearly outperforms FFNN by 8.57 in f1 score, and 9.49 in AUC.

In addition to the prediction score, we measure running time between two best performing models, which are kNN and dropoutNN. In Table 4.12, itemized processing times are reported. Since kNN does not require any training time, we put the tuning time to find best k in the training/tuning stage. For the marginal performance improvement, the processing time of dropoutNN takes about 850 times more than the kNN. However, for security related problems, it is much more important to have a higher prediction score.

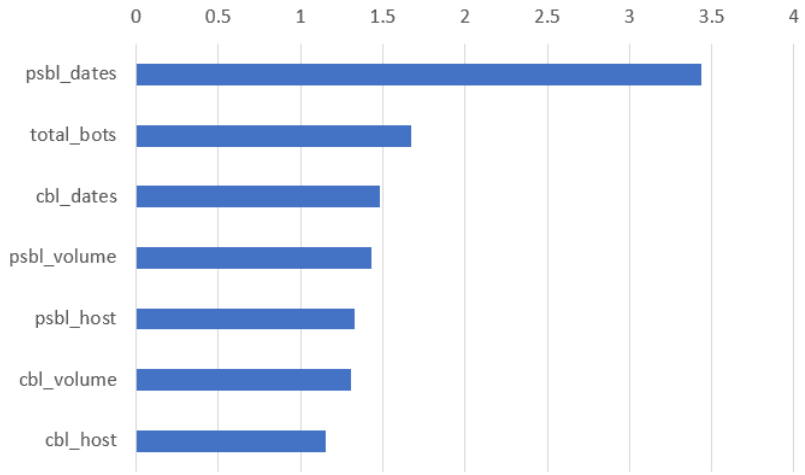


Figure 4.4: Security feature importance score on kNN model.

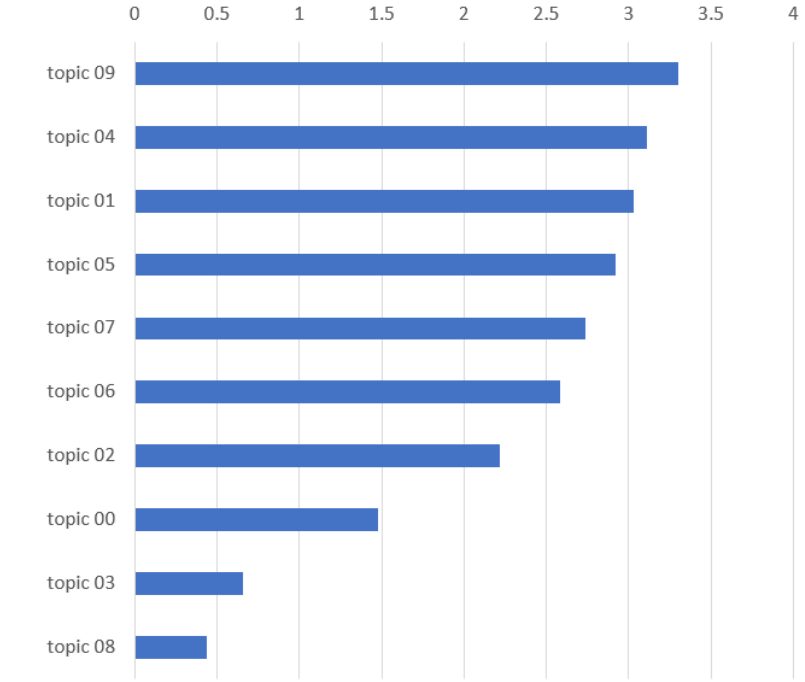


Figure 4.5: Topic importance score on kNN model.

Table 4.12: Process time comparison between kNN and dropoutNN model.

	average training/tuning time (sec)	average testing time (sec)
kNN	2.50	0.04
dropontNN	2113.72	12.10

4.5 Conclusion

Predicting future security incidents is a challenging problem. In this chapter, we introduced a novel cyber incident prediction method using a deep learning model. We constructed a large collection of organizational data feature sets including security postures and risk factor reports. Using these features to train a Dropout Neural Network classifier, it is shown that we can achieve an AUC score of 76. We also analyzed the relative importance of the security posture and topics features. Our work has significant practical implications for organizations, investors, and general public by providing reliable prediction results and topic analysis of the risk reports. We expect further performance improvement along with more enriched dataset and enhanced machine learning algorithms in the future.

Chapter 5

Conclusion

In this dissertation, we studied series of ideas for helping organizational security problems using various research methods. First, we conducted a randomized field experiment in Asian countries and actually improved security level of our subject companies, and found causal relationship between security awareness and protection level. Second, we introduced a game theory model between unknown attacker and an ISP as a defender. By finding the Bayesian Nash equilibrium, we proved that the defenders can limit the attacker's attack intensity by optimal blocking threshold. Third, we proposed a machine learning model to predict data breach incidents. We combined numerical security posture data and text analysis results on risk reports to train the model, and produced meaningful prediction results that can be used in the real world. An interesting direction for work is applying our game theory model in the actual network packet analysis combined with incident prediction model for real time threat detection at the ISP layer.

Appendix: Proofs

For convenience, we define functions φ, y and constants a, b, c, q^* as follows:

$$\begin{aligned}
\varphi(x) &:= \tfrac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \\
a &:= \tfrac{1}{2}\Big(\sqrt{1 + \tfrac{8r\sigma^2}{M^2}} - 1\Big) \\
b &:= \tfrac{(1-a)M}{2\sigma\sqrt{r}} \\
c &:= \tfrac{2\sigma\sqrt{r}e^{-b^2}}{M\sqrt{\pi}} + \varphi(b) \\
q^* &:= \tfrac{p(c-\varphi(b))}{c-p\,\varphi(b)} \\
y(x) &:= \varphi^{-1}\big(\tfrac{c(p-x)}{p(1-x)}\big)
\end{aligned} \tag{5.0.1}$$

In Theorem 3.2.2, the equilibrium attack strategy α, block threshold p, the corresponding attacker's expected profit V_e and the defender's expected cost C_e are following: If $\frac{r\sigma^2}{M^2} \geq 1$, then

$$\begin{aligned}
p &= \tfrac{(1+a)r\,l_f}{(1+a)r\,l_f + aM} \\
\alpha(q) &= M, \quad \text{if} \;\; q \in [0,1] \\
V_e(q) &= \begin{cases} \frac{M}{r}\big(1 - \big(\frac{1-p}{p}\big)^a \big(\frac{q}{1-q}\big)^a\big), & \text{if} \;\; q \in [0,p) \\ 0, & \text{if} \;\; q \in [p,1] \end{cases} \\
C_e(q) &= \begin{cases} q\Big(\frac{M}{r} - \big(\frac{M}{r} - \frac{(1-p)l_f}{p}\big)\big(\frac{1-p}{p}\big)^a \big(\frac{q}{1-q}\big)^a\Big), & \text{if} \;\; q \in [0,p) \\ (1-q)l_f, & \text{if} \;\; q \in [p,1] \end{cases}
\end{aligned} \tag{5.0.2}$$

If $\frac{r\sigma^2}{M^2} < 1$, then

$$
\begin{aligned}
p &= \tfrac{c l_f \sqrt{\pi r}}{c l_f \sqrt{\pi r} + \sigma} \\
\alpha(q) &= \begin{cases} M, & \text{if } \ q \in [0, q^*] \\ \frac{2p(1-q)\sigma\sqrt{r}}{c\sqrt{\pi}(1-p)q}\, e^{-y(q)^2}, & \text{if } \ q \in (q^*, p) \\ \frac{2\sigma\sqrt{r}}{c\sqrt{\pi}}, & \text{if } \ q \in [p, 1] \end{cases} \\
V_e(q) &= \begin{cases} \frac{M}{r} - \frac{\sigma^2}{aM}\big(\frac{1-q^*}{q^*}\big)^a \big(\frac{q}{1-q}\big)^a, & \text{if } \ q \in [0, q^*] \\ \frac{\sigma}{\sqrt{r}}\, y(q), & \text{if } \ q \in (q^*, p) \\ 0, & \text{if } \ q \in [p, 1] \end{cases} \\
C_e(q) &= \begin{cases} q\Big(\frac{M}{r} - \big(\frac{\sigma^2}{aM} - \frac{c\sqrt{\pi r}(1-p) l_f \sigma}{(1+a)Mp}\big)\big(\frac{1-q^*}{q^*}\big)^a \big(\frac{q}{1-q}\big)^a\Big), & \text{if } \ q \in [0, q^*] \\ \frac{\sigma}{\sqrt{r}}\, q\, y(q) + l_f(1-q)\big(e^{-y(q)^2} - \frac{c\sqrt{\pi}(1-p) q\, y(q)}{p(1-q)}\big), & \text{if } \ q \in (q^*, p) \\ (1-q) l_f, & \text{if } \ q \in [p, 1] \end{cases}
\end{aligned}
\tag{5.0.3}
$$

In Proposition 3.3.1, the block threshold $\tilde{p}$, and the corresponding attacker's expected profit V_2 and the defender's expected cost C_2 are following:

$$
\begin{aligned}
\tilde{p} &= \tfrac{(1+a) r\, l_f}{(1+a) r\, l_f + aM} \\
C_2(q_0) &= \begin{cases} q_0\Big(\frac{M}{r} - \big(\frac{M}{r} - \frac{(1-\tilde{p}) l_f}{\tilde{p}}\big)\big(\frac{1-\tilde{p}}{\tilde{p}}\big)^a \big(\frac{q_0}{1-q_0}\big)^a\Big), & \text{if } \ q_0 \in [0, \tilde{p}) \\ (1-q_0) l_f, & \text{if } \ q_0 \in [\tilde{p}, 1] \end{cases} \\
V_2(q_0) &= \begin{cases} \frac{M}{r}\big(1 - \big(\frac{1-\tilde{p}}{\tilde{p}}\big)^a \big(\frac{q_0}{1-q_0}\big)^a\big), & \text{if } \ q_0 \in [0, \tilde{p}) \\ 0, & \text{if } \ q_0 \in [\tilde{p}, 1] \end{cases}
\end{aligned}
\tag{5.0.4}
$$

In Proposition 3.3.2, the block threshold $\tilde{p}$, the attack intensity $\tilde{\alpha}$, the corresponding attacker's expected profit V_3 and the defender's expected cost

C_3 are following: If $\frac{r\sigma^2}{M^2} \geq 1$, then

$$\begin{aligned}
\tilde{p} &= \tfrac{(1+a)r\,l_f}{(1+a)r\,l_f+aM} \\
\tilde{\alpha}(q) &= M, \quad \text{if} \;\; q \in [0,1] \\
V_3(q_0) &= \begin{cases} \frac{M}{r}\big(1-(\frac{1-\tilde{p}}{\tilde{p}})^a(\frac{q_0}{1-q_0})^a\big), & \text{if} \;\; q_0 \in [0,\tilde{p}) \\ 0, & \text{if} \;\; q \in [\tilde{p},1] \end{cases} \\
C_3(q_0) &= \begin{cases} q_0\,V_3(q_0) + l_f(1-q_0)(\frac{1-\tilde{p}}{\tilde{p}})^{a+1}(\frac{q_0}{1-q_0})^{a+1}, & \text{if} \;\; q_0 \in [0,\tilde{p}) \\ l_f(1-q_0), & \text{if} \;\; q_0 \in [\tilde{p},1] \end{cases}
\end{aligned} \tag{5.0.5}$$

If $\frac{r\sigma^2}{M^2} < 1$, then

$$\begin{aligned}
\tilde{p} &= \tfrac{(1+a)r\,l_f}{(1+a)r\,l_f+aM} \\
\tilde{\alpha}(q) &= \begin{cases} M, & \text{if} \;\; q \in [0,\tilde{q}^*] \\ 0, & \text{if} \;\; q \in (\tilde{q}^*,1] \end{cases} \\
V_3(q_0) &= \begin{cases} \frac{M}{r}\big(1-(\frac{a+1}{2})(\frac{1-\tilde{q}^*}{\tilde{q}^*})^a(\frac{q_0}{1-q_0})^a\big), & \text{if} \;\; q_0 \in [0,\tilde{q}^*] \\ \frac{a^2M}{r(1+2a)}(\frac{(1-\tilde{q}^*)q_0}{\tilde{q}^*(1-q_0)})^{a+1}\big((\frac{\tilde{p}(1-q_0)}{(1-\tilde{p})q_0})^{2a+1}-1\big), & \text{if} \;\; q_0 \in (\tilde{q}^*,\tilde{p}) \\ 0, & \text{if} \;\; q \in [\tilde{p},1] \end{cases} \\
C_3(q_0) &= \begin{cases} q_0\,V_3(q_0) + l_f(1-q_0)(\frac{1-\tilde{p}}{\tilde{p}})^{a+1}(\frac{q_0}{1-q_0})^{a+1}, & \text{if} \;\; q_0 \in [0,\tilde{p}) \\ l_f(1-q_0), & \text{if} \;\; q_0 \in [\tilde{p},1] \end{cases}
\end{aligned} \tag{5.0.6}$$

where the constant a is defined in (5.0.1) and

$$\tilde{q}^* := \frac{1}{1+(\frac{1+a}{2a^2})^{\frac{1}{1+2a}}(\frac{1-\tilde{p}}{\tilde{p}})}. \tag{5.0.7}$$

Bibliography

[1] Rony M Adelsman and Andrew B Whinston. Sophisticated voting with information for two voting functions. *Journal of Economic Theory*, 15(1):145–159, 1977.

[2] Axel Anderson and Lones Smith. Dynamic deception. *American Economic Review*, 103(7):2811–47, 2013.

[3] R. Anderson. Why information security is hard - an economic perspective. In *Computer Security Applications Conference, 2001. ACSAC 2001. Proceedings 17th Annual*, pages 358–365, Dec 2001.

[4] Ross Anderson and Tyler Moore. The economics of information security. *Science*, 314(5799):610–613, 2006.

[5] Iván Arce. The weakest link revisited [information security]. *IEEE Security & Privacy*, 99(2):72–76, 2003.

[6] Robert J Aumann and M Maschler. Game theoretic aspects of gradual disarmament. *Report of the US Arms Control and Disarmament Agency*, 80:1–55, 1966.

[7] Kerry Back and Shmuel Baruch. Information in securities markets: Kyle meets glosten and milgrom. *Econometrica*, 72(2):433–465, 2004.

[8] Ricardo Baeza-Yates, Berthier de Araújo Neto Ribeiro, et al. *Modern information retrieval*. New York: ACM Press; Harlow, England: Addison-Wesley,, 2011.

[9] Ray Ball and Philip Brown. An empirical evaluation of accounting income numbers. *Journal of accounting research*, pages 159–178, 1968.

[10] Johannes M Bauer and Michel JG Van Eeten. Cybersecurity: Stakeholder incentives, externalities, and policy options. *Telecommunications Policy*, 33(10):706–719, 2009.

[11] William H Beaver. The information content of annual earnings announcements. *Journal of accounting research*, pages 67–92, 1968.

[12] Shyam B Bhandari and Rajesh Iyer. Predicting business failure using cash flow statement based measures. *Managerial Finance*, 39(7):667–676, 2013.

[13] Karl Borch. Equilibrium in a reinsurance market. In *Foundations of Insurance Economics*, pages 230–250. Springer, 1992.

[14] J Efrim Boritz and Duane B Kennedy. Effectiveness of neural network types for prediction of business failure. *Expert Systems with Applications*, 9(4):503–512, 1995.

[15] Indranil Bose and Alvin Chung Man Leung. Assessing anti-phishing preparedness: a study of online banks in hong kong. *Decision Support Systems*, 45(4):897–912, 2008.

[16] Indranil Bose and Alvin Chung Man Leung. Technical opinion what drives the adoption of antiphishing measures by hong kong banks? *Communications of the ACM*, 52(8):141–143, 2009.

[17] Indranil Bose and Alvin Chung Man Leung. The impact of adoption of identity theft countermeasures on firm value. *Decision Support Systems*, 55(3):753–763, 2013.

[18] Andrej Bratko, Gordon V Cormack, Bogdan Filipič, Thomas R Lynam, and Blaž Zupan. Spam filtering using statistical data compression models. *Journal of machine learning research*, 7(Dec):2673–2698, 2006.

[19] Ismail Butun, Salvatore D Morgera, and Ravi Sankar. A survey of intrusion detection systems in wireless sensor networks. *IEEE communications surveys & tutorials*, 16(1):266–282, 2014.

[20] Ran Canetti. Universally composable security: A new paradigm for cryptographic protocols. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 136–145. IEEE, 2001.

[21] Eoghan Casey. *Digital evidence and computer crime: Forensic science, computers, and the internet*. Academic press, 2011.

[22] Virginia Cerullo and Michael J Cerullo. Business continuity planning: a comprehensive approach. *Information Systems Management*, 21(3):70–78, 2004.

[23] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[24] Nitesh V Chawla, Nathalie Japkowicz, and Aleksander Kotcz. Special issue on learning from imbalanced data sets. *ACM Sigkdd Explorations Newsletter*, 6(1):1–6, 2004.

[25] Hsinchun Chen and Fei-Yue Wang. Guest editors' introduction: Artificial intelligence for homeland security. *IEEE Intelligent Systems*, 20(5):12–16, 2005.

[26] Gordon V Cormack and Thomas R Lynam. Online supervised spam filter evaluation. *ACM Transactions on Information Systems (TOIS)*, 25(3):11, 2007.

[27] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

[28] Vincent P Crawford. Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *The American economic review*, 93(1):133–149, 2003.

[29] John D'Arcy, Anat Hovav, and Dennis Galletta. User awareness of security countermeasures and its impact on information systems misuse: A deterrence approach. *Information Systems Research*, 20(1):79–98, 2009.

[30] Dorothy E Denning. An intrusion-detection model. *IEEE Transactions on software engineering*, (2):222–232, 1987.

[31] Cuong T. Do, Nguyen H. Tran, Choongseon Hong, Charles A. Kamhoua, Kevin A. Kwiat, Erik Blasch, Shaolei Ren, Niki Pissinou, and Sundaraja Sitharama Iyengar. Game theory for cyber security and privacy. *ACM Comput. Surv.*, 50(2):30:1–30:37, May 2017.

[32] Darrell Duffie, Leandro Saita, and Ke Wang. Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3):635–665, 2007.

[33] Esther Gal-Or and Anindya Ghose. The economic incentives for sharing security information. *Information Systems Research*, 16(2):186–208, 2005.

[34] Lawrence A Gordon, Martin P Loeb, and William Lucyshyn. Information security expenditures and real options: A wait-and-see approach. 2003.

[35] Samuel Greengard. Cybersecurity gets smart. *Communications of the ACM*, 59(5):29–31, 2016.

[36] Qiong Gu, Xian-Ming Wang, Zhao Wu, Bing Ning, and Chun-Sheng Xin. An improved smote algorithm based on genetic algorithm for imbalanced data classification. *Journal of Digital Information Management*, 14(2):92–103, 2016.

[37] Hui Han, Wen-Yuan Wang, and Bing-Huan Mao. Borderline-smote: a new over-sampling method in imbalanced data sets learning. In *International conference on intelligent computing*, pages 878–887. Springer, 2005.

[38] William Hardy, Lingwei Chen, Shifu Hou, Yanfang Ye, and Xin Li. Dl4md: A deep learning framework for intelligent malware detection. In *Proceedings of the International Conference on Data Mining (DMIN)*, page 61. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2016.

[39] Haibo He and Edwardo A Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge & Data Engineering*, (9):1263–1284, 2008.

[40] Shu He, Gene Moo Lee, Sukjin Han, and Andrew B Whinston. How would information disclosure influence organizations' outbound spam volume? Evidence from a field experiment. *Journal of Cybersecurity*, 2(1):99–118, 2016.

[41] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

[42] James J Heckman and Jeffrey A. Smith. Assessing the case for social experiments. *The Journal of Economic Perspectives*, 9(2):85–110, 1995.

[43] Kenneth Hendricks and R Preston McAfee. Feints. *Journal of Economics & Management Strategy*, 15(2):431–456, 2006.

[44] Stephen A Hillegeist, Elizabeth K Keating, Donald P Cram, and Kyle G Lundstedt. Assessing the probability of bankruptcy. *Review of accounting studies*, 9(1):5–34, 2004.

[45] Marek Kaluszka. Optimal reinsurance under mean-variance premium principles. *Insurance: Mathematics and Economics*, 28(1):61–67, 2001.

[46] Wazir Zada Khan, Muhammad Khurram Khan, Fahad T Bin Muhaya, Mohammed Y Aalsalem, and Han-Chieh Chao. A comprehensive study of email spam botnet detection. *IEEE Communications Surveys & Tutorials*, 17(4):2271–2295, 2015.

[47] Logan Kugler. Online privacy: regional differences. *Communications of the ACM*, 58(2):18–20, 2015.

[48] Albert S Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335, 1985.

[49] Charles MC Lee, Belinda Mucklow, and Mark J Ready. Spreads, depths, and the impact of earnings information: An intraday analysis. *The Review of Financial Studies*, 6(2):345–374, 1993.

[50] Wenke Lee, Salvatore J Stolfo, et al. Data mining approaches for intrusion detection. In *USENIX Security Symposium*, pages 79–93. San Antonio, TX, 1998.

[51] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[52] Yang Liu, Armin Sarabi, Jing Zhang, Parinaz Naghizadeh, Manish Karir, Michael Bailey, and Mingyan Liu. Cloudy with a chance of breach: Forecasting cyber security incidents. In *USENIX Security Symposium*, pages 1009–1024, 2015.

[53] Semyon Malamud, Huaxia Rui, and Andrew Whinston. Optimal reinsurance with multiple tranches. *Journal of Mathematical Economics*, 65:71–82, 2016.

[54] Mohammad Hossein Manshaei, Quanyan Zhu, Tansu Alpcan, Tamer Bacşar, and Jean-Pierre Hubaux. Game theory meets network security and privacy. *ACM Computing Surveys (CSUR)*, 45(3):25, 2013.

[55] Tyler Moore and Richard Clayton. The impact of public information on phishing attack and defense. 2011.

[56] Tyler Moore, Richard Clayton, and Ross Anderson. The economics of online crime. *Journal of Economic Perspectives*, 23(3):3–20, 2009.

[57] Kari Lock Morgan, Donald B Rubin, et al. Rerandomization to improve covariate balance in experiments. *The Annals of Statistics*, 40(2):1263–1282, 2012.

[58] Leonid Portnoy, Eleazar Eskin, and Sal Stolfo. Intrusion detection with unlabeled data using clustering. In *In Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001*. Citeseer, 2001.

[59] Cristhian Potes, Saman Parvaneh, Asif Rahman, and Bryan Conroy. Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds. In *Computing in Cardiology Conference (CinC), 2016*, pages 621–624. IEEE, 2016.

[60] Foster Provost. Machine learning from imbalanced data sets 101. In *Proceedings of the AAAI2000 workshop on imbalanced data sets*, pages 1–3, 2000.

[61] John Quarterman, Leigh Linden, Qian Tang, Gene Moo Lee, and Andrew Whinston. Spam and botnet reputation randomized control trials and policy. 2013.

[62] Martin Roesch et al. Snort: Lightweight intrusion detection for networks. In *Lisa*, volume 99, pages 229–238, 1999.

[63] Sasha Romanosky. Examining the costs and causes of cyber incidents. *Journal of Cybersecurity*, 2(2):121–135, 2016.

[64] Sankardas Roy, Charles Ellis, Sajjan Shiva, Dipankar Dasgupta, Vivek Shandilya, and Qishi Wu. A survey of game theory as applied to network security. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, pages 1–10. IEEE, 2010.

[65] Sushmita Ruj, Milos Stojmenovic, and Amiya Nayak. Decentralized access control with anonymous authentication of data stored in clouds. *IEEE transactions on parallel and distributed systems*, 25(2):384–394, 2014.

[66] Ravi Sen and Sharad Borle. Estimating the contextual risk of data breach: An empirical approach. *Journal of Management Information Systems*, 32(2):314–341, 2015.

[67] Sima Sharifirad, Azra Nazari, and Mehdi Ghatee. Modified smote using mutual information and different sorts of entropies. *arXiv preprint arXiv:1803.11002*, 2018.

[68] Nikhil Shetty, Galina Schwartz, and Jean Walrand. Can competitive insurers improve network security? In *International Conference on Trust and Trustworthy Computing*, pages 308–322. Springer, 2010.

[69] Zhan Shi, Gene Moo Lee, and Andrew B Whinston. Toward a better measure of business proximity: Topic modeling for industry intelligence. *MIS quarterly*, 40(4), 2016.

[70] Victor Song, Hasan Cavusoglu, Gene Moo Lee, and Mary Li Zhi Ma. It risk factor disclosure and stock price crash risk. 2019.

[71] Kyle Soska and Nicolas Christin. Automatically detecting vulnerable websites before they turn malicious. In *USENIX Security Symposium*, pages 625–640, 2014.

[72] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[73] Brett Stone-Gross, Marco Cova, Lorenzo Cavallaro, Bob Gilbert, Martin Szydlowski, Richard Kemmerer, Christopher Kruegel, and Giovanni Vigna. Your botnet is my botnet: analysis of a botnet takeover. In *Proceedings of the 16th ACM conference on Computer and communications security*, pages 635–647. ACM, 2009.

[74] Brett Stone-Gross, Thorsten Holz, Gianluca Stringhini, and Giovanni Vigna. The underground economy of spam: A botmaster's perspective of coordinating large-scale spam campaigns. *LEET*, 11:4–4, 2011.

[75] Daniel Svozil, Vladimir Kvasnicka, and Jiri Pospichal. Introduction to multi-layer feed-forward neural networks. *Chemometrics and intelligent laboratory systems*, 39(1):43–62, 1997.

[76] Qian Tang, Leigh Linden, John S Quarterman, and Andrew B Whinston. Improving internet security through social information and social comparison: A field quasi-experiment. *WEIS 2013*, 2013.

[77] Robert W Taylor, Eric J Fritsch, and John Liederbach. *Digital crime and digital terrorism*. Prentice Hall Press, 2014.

[78] Michel J van Eeten and Johannes M Bauer. Economics of malware: Security decisions, incentives and externalities. *OECD Science, Technology and Industry Working Papers*, 2008(1):0_1, 2008.

[79] Henk CA Van Tilborg and Sushil Jajodia. *Encyclopedia of cryptography and security*. Springer Science & Business Media, 2014.

[80] Erik Verhoef. Externalities. Serie Research Memoranda 0031, VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics, 1997.

[81] Jay Rodney Walton and Sanjiv Nanda. High speed media access control and direct link protocol, June 30 2015. US Patent 9,072,101.

[82] Tawei Wang, Karthik N Kannan, and Jackie Rees Ulmer. The association between the disclosure and the realization of information security risk factors. *Information Systems Research*, 24(2):201–218, 2013.

[83] Christopher W Zobel and Lara Khansa. Quantifying cyberinfrastructure resilience against multi-event attacks. *Decision Sciences*, 43(4):687–710, 2012.