Information Disclosure and Security Vulnerability Awareness: A Large-Scale Randomized Field Experiment in Pan-Asia

Yunhui Zhuang
College of Business
City University of Hong Kong
yhzhuang2-c@my.cityu.edu.hk

Yunsik Choi
Department of Computer Science
University of Texas at Austin
yunsik@cs.utexas.edu

Shu He
School of Business
University of Connecticut
shu.he@uconn.edu

Alvin Chung Man Leung
College of Business
City University of Hong Kong
acmleung@cityu.edu.hk

Gene Moo Lee
Sauder School of Business
University of British Columbia
gene.lee@sauder.ubc.ca

Andrew B. Whinston
McCombs School of Business
University of Texas at Austin
abw@uts.cc.utexas.edu

Abstract

This paper investigates how the disclosure of a security vulnerability index based on outgoing spams and phishing website hosting, which may serve as an indicator of a firm's inadequate security controls, affects companies' security protection strategy. Our core objective is to study whether firms improve their security when they become aware of their vulnerabilities and such information is publicized. To achieve this goal, we conduct a randomized field experiment on 1,262 firms in six Pan-Asian countries and regions. For the treatment group of 631 firms, we alert them of their security vulnerability index and ranking over time, and their relative performance compared to their peers via emails and a public advisory website. Compared with the control group without being informed of their security vulnerability index, the treatment group improved their security over time, with a significant reduction of outgoing spam volume. A marginally significant improvement in reducing phishing hosting websites is also observed among non-web hosting firms in the treatment group. The security improvement may be attributed to firms' proactive reaction to the security vulnerability information. Our study provides cybersecurity policy makers with useful insights on how to motivate firms to adopt better security measures.

1. Introduction

Cyberattacks impose serious threats to individuals, firms, and our society at large. Even with technological advances in security software and hardware, we are still experiencing an ever-increasing number of cyberattacks [1, 2] Although firms are aware of cybersecurity issues, they are still reluctant to adopt adequate measures to prevent the spread of cyberattacks. Such a problem is partly due to negative externalities, information asymmetry, and misaligned incentives [3]. Negative externalities refer to the phenomenon that firms in a network have a higher incentive to wait than to adopt a security technology immediately because the cost of the technology adoption is greater than its initial benefit until a minimum number of network players adopt it.[1] Previous research shows that such a wait-and-see approach is widely used by senior managers and it may lead to the ubiquitous security breaches in the US [4, 5]. In addition, due to the quality uncertainty of security technologies available in the market[2] and misaligned incentives of for-profit firms, firms may deprioritize security issues when related security problems are less likely to directly harm themselves, even though the issues create negative externalities to other firms and the general public at large [3, 7, 8]. In this paper, we investigate potential measures which can be effective in increasing firms' awareness of cyber security and internalizing the externalities to develop more secure cyber environments. This study also echoes the origin[3] responsibility principle of the research framework of the Bright ICT initiative [9, 10] by proposing a new security vulnerability index to incentivize firms to behave as good citizens and take a proactive approach to prevent the widespread use of undesirable content over the Internet [11].

Similar to the idea of Moody's and Standard and Poor's credit ratings, our proposed security vulnerability index may reflect an organization's vulnerabilities to cybercrime and its adequacy to prevent the spread of unsolicited online content. The index is constructed by processing large-scale, real-time cyber incident data from spam emission[4] (sources: CBL[5] and PSBL[6]) and phishing website hosting[7] activities (sources:

[1] An analogy to adoption of security technology is vaccinating children against a contagious disease. A parent may choose not to vaccinate their children and freeride on others in the same community who have already done so [3].

[2] Such uncertainty may lead to the problem of "market for lemons" or information asymmetry [6].

[3] Origin refers to firms whose servers may be compromised to send undesired content to the Internet and the company owners may or may not be aware of such a problem and have control of it [9].

[4] Note that the term "spam mail" in this paper includes advertisement, phishing mail, and malware attached email.

[5] Composite Block List: `https://www.abuseat.org/`

[6] Passive Spam Block List: `https://psbl.org/about/`

[7] Note that phishing, in this paper, exclusively refers to website-related incidents, and we only focus on the firms who are actually hosting the phishing websites on their own server. All email-related attacks including phishing emails are included in our spam data.

URI: https://hdl.handle.net/10125/64481
978-0-9981331-3-3

APWG[8] and OpenPhish[9]). We choose spam and phishing as data sources because they are the most commonly seen undesirable content on the Internet. Firms' computers with inadequate preventive security measures may be easily controlled by their adversaries via bots to send spams or host phishing websites. As a result, the outgoing spam volumes and phishing websites hosted may be indicative of the security vulnerabilities of a firm. We are interested to test whether informing and publicizing individual firms of their security vulnerability index may motivate them to adopt better security measures over time. To evaluate the effectiveness of such approach, we conduct a large-scale randomized field experiment (RFE) in Pan Asia, which is characterized by blooming e-commerce markets and heterogeneous juridical systems. Furthermore, our research addresses several limitations of a similar study by [12]. First, to our knowledge, we are among the first to implement RFE in Pan Asia, which is not restricted by one single jurisdiction on cybercrime. Second, because e-commerce is blooming in Pan Asia, it is different from the U.S., where the sense of awareness of cybersecurity is stronger. Therefore, our RFE treatment effects are less likely to be influenced by external factors (e.g., stricter laws and stronger sense of awareness on cybersecurity). Third, instead of restricting our study to spam collected from a single data source (i.e., CBL), we diversify the data sources and also include data on phishing website hosting. The diversification of data may increase the robustness of our proposed security vulnerability index. Fourth, He et al. [12] has a relatively short treatment window (from January to March 2014) and analyzed the pre- and post-treatment in a 6-month window. We send out treatment emails three times (July, September, and November 2017) and use a more concise window of one month to measure the gradual security performance changes of firms over time prior to and after individual treatments. In addition to the treatment emails, we develop a public website, cybeRatings, for the treatment firms and the general public to search for and read more details on individual firms' security vulnerabilities. Fifth, to ensure the treatment compliance, we adopt email and web analytic tools to check whether the treated firms have received our treatments properly and to tightly monitor firms' reactions to our treatments. Finally, we implement more robust statistical analyses. Apart from the difference-in-difference (DID) model, we also analyze the heterogeneous treatment effect.

Our empirical results show that the treatments (i.e., emails and visits to the advisory website) induced a significant reduction of outbound spam volume. Our dynamic analysis shows that there is a significant decline in CBL spam volume after the first two batches of our treatment emails. Interestingly, although we do not observe overall treatment effect on the phishing website hosting, an extended analysis shows that our treatments had marginally significant effects on phishing website reduction for the firms that are neither Internet service providers nor web hosting providers. Finally, we analyze overall security performance by Borda counts which aggregate spam and phishing volumes from different sources. The results show that the treatments can increase country-level security vulnerability rank, which suggests improved performance among peers in the same countries. In sum, our research findings show that firms have different incentives when dealing with phishing website hosting compared to spam emission.

This study contributes to the cybersecurity literature in multiple ways. First, we develop a novel security vulnerability index based on outgoing spam volume and phishing website hosting. Second, we implement a large-scale information system to alert treatment firms of their security vulnerabilities by emails and to publish their vulnerability details on a public advisory website. Third, our study comprehensively covers *all* firms in the six targeted countries and regions in Pan Asia with at least one Autonomous System Number (ASN) and one valid contact email address. Through the large-scale field experiment, we show that firms improve their internal system security over time when they learn the information about their security vulnerabilities by emails and our advisory website in the short term. We also show disparate firm behaviors on disclosure of spam and phishing vulnerabilities. Our research can provide useful insights to cybersecurity policy-makers. Instead of using penalties to make firms comply, they may use publicized security information to incentivize firms to adopt better preventive measures to mitigate the widespread use of undesired content. Our study also responds to the call of the Bright ICT Initiative by developing an incentive mechanism to promote origin responsibilities.

2. Theoretical Background

Researchers from information systems, computer science, and economics are actively seeking the most efficient solutions to contain widespread cybersecurity threats. To thwart cybercrime, prevention and protection tools are equally important. Whilst existing research primarily focuses on protective solutions, for example, spam filtering [13, 14], intrusion detection systems [15, 16, 17], and digital forensics [18, 19], few discuss preventive measures, which include rules and reminders regarding best safety practices. In fact, prevention comes before protection; only when prevention fails does protection take place [20]. With note of the inadequate preventive measures to curb the spread of undesired online content, the Bright ICT Initiative has established the origin responsibility principle [9, 10]. How to motivate firms to adopt better preventive security measures seems to be an important research question to be addressed. In this section, we discuss related literature at the boundary of security and economics that investigates the relationship among incentives, externalities, and security investment.

2.1. Misaligned Incentives and Security Underinvestment

Anderson and Moore [3] showed that economic incentives are as important as technical designs in information security solutions. Senior management is willing to invest in protective security that can safeguard their *internal* corporate assets

[8] Anti-Phishing Working Group: `https://apwg.org/`

[9] OpenPhish Phishing Intelligence: `https://openphish.com/`

from cyberattacks. In contrast, their incentive to invest deteriorates when the underlying technologies are to protect assets of *external* entities. Because they bear no financial benefits from such an investment, misaligned incentives may lead to underinvestment in information security [21]. As a result, managers may simply adopt the minimal security measures to protect their own assets, rather than a comprehensive security solution that can prevent widespread cyberattacks to the general public.

2.2. Heterogeneity Defense

Sharman et al. [22] show that the diversity in security investment is important to deter cybercrime. Such a strategy is known as "functionality defense by heterogeneity." If a firm only focuses on one type of security (e.g., protection) with negligence on another (e.g., prevention), the security solutions are not considered to be thorough. A firm that fails to invest in comprehensive security not only increases the probability of its own security risks but also increases the likelihood that such risks will spill over to other firms [23]. Instead, diversification in security investment may help prevent correlated failures (e.g., shared vulnerabilities due to homogeneous security investment and loss of availability of connected company networks) [24].

2.3. Information Disclosure as Externality

Network externalities may provide some explanation regarding the reluctance of firms to adopt adequate security measures. Kunreuther and Heal [25] demonstrate that the security of a group of people often leans on each of its members. As one user in the system takes more precautions to protect his/her computers, the less the others in the same group will be infected or intruded upon. Such a setting leads to the classic *free-rider problem* that each user in the system lacks the incentive to adequately protect themselves against attacks or viruses because the cost of the spread of the attacks or viruses is borne by other users. Therefore, in the absence of a market for appropriate incentives, individuals will choose less security than the social optimal level. In other words, firms may deprioritize IT security problems when they are less likely to directly harm themselves, even though they create negative externalities to others (e.g., spam and phishing attacks initiated by their compromised computers).

To combat the problems brought by the negative externalities of security underinvestment, one approach can be to alert firms of their security vulnerabilities and the associated loss due to such insecurity. To quantify such loss, previous research proposes the use of a "vulnerability matrix" [26] and "node failure correlation matrix" [24]. Besides enhancing awareness, public disclosure of attack incidents may help defenders get prepared against cybercrime [27]. In the same token, public disclosure of a firm's vulnerabilities in spam may make the firm take more proactive security action to salvage its public reputation [28]. Such disclosure may also alleviate the information asymmetry issue and allow firms to better understand their security weakness [12]. Furthermore, social comparison and peer pressure may incentivize firms even more to adopt better countermeasures [29]. It may also serve as an additional externality to raise firms' cybersecurity awareness due to the fear of losing customers to their competitors [30].

Based on the literature review, we find that security awareness enhancement can be an effective mechanism to motivate firms to adopt an optimal level of security solutions and prevent the wide spread of undesired online content. To achieve this aim, we can inform firms of their security vulnerabilities. To amplify the effect, we can use the method of public disclosure and facilitate firms to compare their security performance with that of their peers. In the next section, we will discuss how to evaluate the effectiveness of such a design through a randomized field experiment.

3. Experimental Design and Implementation

3.1. Development of Security Vulnerability Index

An organization's Internet security condition is a latent variable that cannot be directly measured. However, one way to estimate it is by the use of perceptible data. Security attacks originating from a corporate network may be a good indicator of weak security infrastructure. To estimate the number of attacks, we can use outbound spam volume and phishing websites as proxies. According to Symantec's MessageLabs, over 50% of spam is sent by botnets [31]. These infected computers and servers may be used by adversaries as media for even more serious cyberattacks, for example, distributed denial of service (DDoS) attacks, identity thefts, hacking, data breaches, and cyber vandalism. In this research, we use (1) outbound spam volume generated from a corporate network and (2) number of phishing websites hosted in the corporate network to construct a comprehensive security vulnerability index.

To construct a composite ranking from four constituent rankings from each data source (i.e., CBL, PSBL, APWG, and OpenPhish), we use Borda count [32]. First, we extract the ranking for each of the five combinations of data sources and metrics (CBL Volume, PSBL Volume, APWG volume, OpenPhish volume, and HSIC) with worse performance being ranked higher in terms of spam or phishing volume. Next, we can construct the composite Borda ranking by taking a firm's rank k for a given ranking and grant that firm a point of $(n + 1 - k)$ for that ranking, where n is the total number of firms in that ranking. Finally, we sum these points for the individual rankings to produce the Borda count for each firm. Firms with higher Borda counts get higher composite Borda ranks, which indicate worse performance. Firms with the best security level (e.g., no spam and phishing volume) are ranked equally the lowest.

3.2. Randomized Field Experiment (RFE)

To causally test whether publicized security information increases firms' awareness and eventually improves their security over time, we employ randomized field experiment

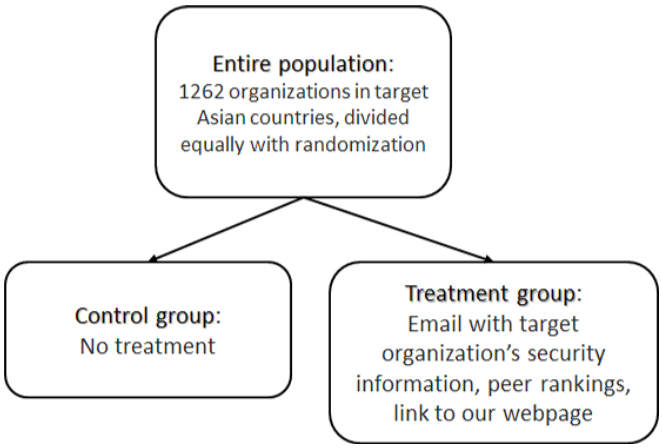

Figure 1: Design of the Randomized Field Experiment

(RFE) along with econometric analysis as the main evaluation methodology. RFE, also referred to as a randomized controlled trial, is a well-established evaluation methodology in the social sciences for policy interventions, in which the findings can be explained by different factors associated with the interventions [33]. The main advantage of this methodology is its capability of detecting a causal relationship in a naturally occurring environment.

The firms in this experiment were split into two equally sized, statistically homogeneous groups by stratified and match-pair randomization [34]. The grouping is summarized in Figure 1. In the treatment group, advisory emails with security evaluation reports were sent to relevant contacts within each organization in three different time periods. Each treatment email included (i) the organization's spam and phishing data, such as total spam mail and phishing website hosting volume, (ii) peer rankings in the corresponding industry sectors or certain region, and (iii) a hyperlink to a designated advisory webpage for the treated firm. The webpage also facilitated peer search of security vulnerability reports over time. In the control group, there was no such treatment.

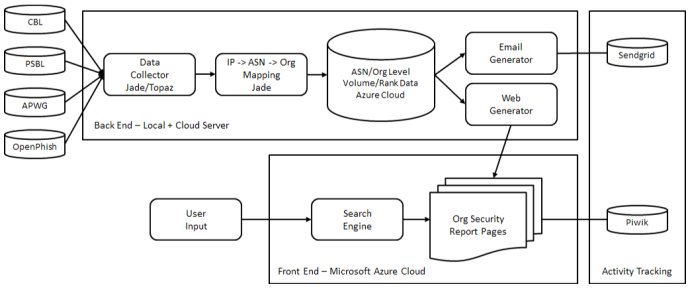

Figure 2: System Architecture

3.3. Data

Firstly, we collected a full list of 1,930 registered ASN information from the target Pan-Asian countries and regions via the WHOIS database[10]. After mapping the ASNs to registered company names, we created a list of 1,293 firms who own at least one ASN. Lastly, we manually collected and validated corporate email addresses from those firms and finalized a list of 1,262 firms. It is important to point out that our field experiment was conducted with a "full population" of firms who own at least one registered ASN and a valid email address in six Pan-Asian countries and regions. Table 1 shows the number of firms in each country. Figure 2 illustrates the architecture of the entire experimental system. The system is concurrently hosted by two authors' research centers.

Table 1: Number of firms for each country and district

Countries and Districts	Number of Firms	Control Group	Treatment Group
Hong Kong	309	631	631
Mainland China	309		
Singapore	264		
Malaysia	171		
Taiwan	138		
Macau	4		
Others	67		
Total	1,262	1,262	

4. Empirical Analysis

Our data were taken from 1,262 firms from six Pan-Asian countries and regions: Hong Kong, Mainland China, Singapore, Malaysia, Taiwan, and Macau. Among them, 631 firms were randomly selected for the treatment group and the rest were placed in the control group. Once we received approval from the human research ethics committees of the authors' universities to implement this research, we contacted firms in the treatment group to provide them with the opportunity to opt out of the experiment and three firms choose to opt out. Starting in July 2017, we sent out a batch of security information emails to firms in the treatment group every two months, for a total of three batches. Overall, 565 out of 631 treatment firms successfully received at least one treatment email. As a result, we used these 565 firms and their corresponding 565 firms in the control group as our empirical analysis data set, for a total of 1,130 firms. Table 2 contains summary statistics for the main variables in our empirical analysis. We collected each firm's number of IP addresses from Team Cymru.[11] Note that the Team Cymru does not have IP address information for a small number of firms in our dataset.

To evaluate whether the security performance of the firms in the treatment group had improved after our intervention, we compared treatment firms' outbound spam and phishing volume prior to and after our experimental intervention with those from the control group. Since the first batch of emails was sent in July 2017, we used 6-month average spam and phishing volume between January 2017 and June 2017 as firms' pre-experiment security measures. To check the internal validity of our randomized field experiment, we used t-tests and Kolmogorov-Smirnov test (KS-test) to examine

[10]WHOIS database: `https://whois.icann.org/en`

[11]Team Cymru: `https://www.team-cymru.com/`

Table 2: Summary statistics

Variable	Variable description	Mean	S.D.	Max	Min
Log(cv+1)	Log transformed CBL volume	2.099	3.667	18.420	0
Log(pv+1)	Log transformed PSBL volume	0.393	1.339	11.969	0
Log(av+1)	Log transformed APWG volume	0.0340	0.275	6.125	0
Log(ov+1)	Log transformed OpenPhish volume	0.0678	0.388	4.663	0
Number of IP addresses	Total number of IP addresses owned by each firm	352,038.3	3,936,268	141 million	0
If has social media account	If the company has at least one social media account	0.7035	0.4569	1	0
If has opened treatment emails	If a firm has opened a treatment email on or before this month	0.2062	0.4048	1	0
If has visited treatment website	If a firm has visited our advisory website on or before this month	0.07080	0.2566	1	0

Table 3: Baseline comparison for internal validity

Variable	Mean Difference	t-statistics	K-S prob (P value)
ln(CV)	-0.05986	-0.2962	0.909
CV	-38.25	-0.3251	0.796
ln(PV)	-0.03817	-0.5162	1.000
PV	0.03769	0.2849	1.000
ln(OV)	-0.02108	-1.0929	1.000
OV	-0.0001904	-1.3810	1.000
ln(AV)	-0.001803	-0.2708	1.000
AV	-0.0000164	-0.5911	1.000
ln(number of IP addresses)	0.1673	0.7346	0.751
number of IP addresses	101837.1	0.3094	0.751
If has social media account	-0.1815	0.8560	1.000
HSIC (first 2 digits)			1.000

whether firms in the treatment group were statistically equivalent to those in the control group. The results are shown in Table 3. We observed that the differences of the average characteristics and the distributions between the treatment and control groups were marginal, and none of them were statistically significant. Therefore, our randomization satisfies the assumption of exogeneity.

4.1. Difference-in-Differences (DID) Analysis

We face a non-compliance issue as some firms might not receive or actually open our treatment emails. Thus we started with an intention-to-treat (ITT) analysis. We used a company's spam volume and phishing website count from July 2017 to December 2017 as its security performance (i.e., dependent variable) after our experimental intervention. If a firm's security condition improved, we would expect a reduction of spam emission and phishing hosting compared with those of the control group after our treatment. For the panel data set of firms' spam and phishing information from January 2017 to December 2017, we applied a DID model to estimate the average treatment effect of our intervention. In particular, the email treatment dummy variable $email\_treat_{it}$ was set equals to 1 if a firm i was in the treatment group and had successfully received the treatment email in month t. Specifically, the ordinary least squares (OLS) regression function is as follows:

$$y_{it} = \alpha_0 + \alpha_1 email\_treat_{it} + \theta_i + \sigma_t + \epsilon_{it} \qquad (1)$$

where y_{it} is one of the four security performance measures in our data set. From Table 2, we can see that the distributions of all main variables are highly skewed, thus we used log transformed spam or phishing volume as our dependent variables.[12] Moreover, in our data set, 20.62% of treated firms have opened our treatment emails. In other words, about 41.5% of the treated organizations who received the treatment emails have opened them. In Equation 1, α_1 is our main variable of interest. If α_1 is negative and statistically significant, then compared with firms in the control group, the security performance of those in the treatment group has improved after our intervention. To control for an organization's time-invariant unobservable characteristics and temporal variation, we also included organization-specific (θ_i) and month (σ_t) fixed effects in our regression.

The main results are reported in Table 4. The results show that among different security performance measures, the treatment had significantly effect on firms' outbound spam volume as measured by CBL. The estimated treatment effect on PSBL spam volume is negative but not statistically significant. On the other hand, for phishing information, there is no evidence showing that our intervention motivates firms to correct their phishing website hosting behavior. The results support our proposition that firms will have different responses to spam and phishing information. While firms care about their own internal security issues (i.e., their own computers being compromised), it seems that they are reluctant to solve negative externality issues (i.e., hosting phishing websites) [3].

Since three different batches of emails were sent, we can evaluate how the treatment effects evolve from the first batch to the third one. If firms put emphasis on security, they would respond to our emails consistently over time. On the other

[12] Specifically, using CBL spam volume as an example, the dependent variable used in the analysis is $ln(CV) = log(CV + 1)$.

Table 4: DID analysis on monthly security measures

	ln(CV)	ln(PV)	ln(AV)	ln(OV)
	(1)	(2)	(3)	(4)
email_treat	-0.135**	-0.000842	0.00974	-0.00766
	(0.0682)	(0.0338)	(0.0114)	(0.0121)
Organization fixed effects	yes	yes	yes	yes
Month fixed effects	yes	yes	yes	yes
Constant	1.893***	0.287***	0.0417***	0.0779***
	(0.0341)	(0.0166)	(0.00522)	(0.00698)
Number of observations	13,560	13,560	13,560	13,560
Number of organizations	1,130	1,130	1,130	1,130
R-squared	0.014	0.053	0.012	0.004

Note: Clustered standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

hand, they may put lower priority on our emails after a few months. To empirically test this, we have included three interaction terms representing each round of emails separately. The results are reported in Table 5, which show that the first two emails have significantly reduced firms' outbound CBL spam volume, while the last one's impact is quite marginal. Considering the fact that the outcomes of firms' security protect measures may not show up until a few months later, the significant effect in the second round might be partially due to the influence from the first email. This result can be an evidence of our hypothesis that firms do not pay enough attentions to security problems, as they stop responding to our treatments after a few months.

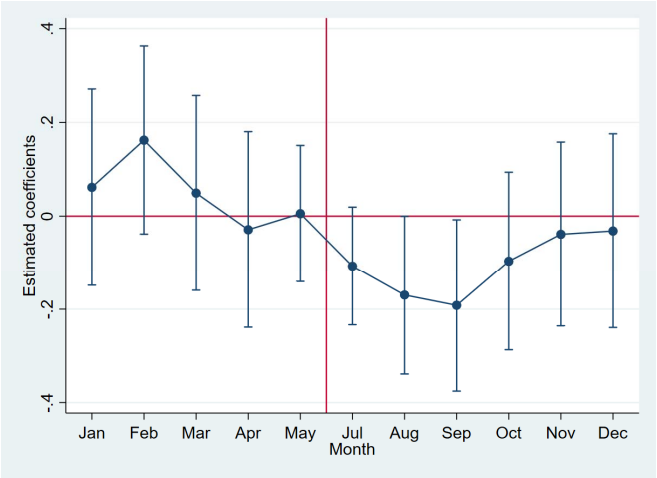

Figure 3: Monthly interaction coefficients for the DID trend test on CV

One common assumption of the DID model is the parallel trend assumption, which means that in the absence of treatment, the difference between the control and treatment groups is constant over time [35, 36]. Violation of this assumption

Table 5: Treatment effects of three batches of emails

	ln(CV)	ln(PV)	ln(AV)	ln(OV)
	(1)	(2)	(3)	(4)
email_interaction1	-0.157**	0.0230	0.00287	-0.0105
	(0.0788)	(0.0485)	(0.00810)	(0.0126)
email_interaction2	-0.179**	-0.00936	0.0252	-0.000239
	(0.0801)	(0.0340)	(0.0155)	(0.0156)
email_interaction3	-0.0676	0.0145	0.00182	-0.00525
	(0.0833)	(0.0441)	(0.0175)	(0.0154)
Organization fixed effects	yes	yes	yes	yes
Month fixed effects	yes	yes	yes	yes
Constant	1.893***	0.287***	0.0417***	0.0779***
	(0.0341)	(0.0166)	(0.00522)	(0.00698)
Number of observations	13,560	13,560	13,560	13,560
Number of organizations	1,130	1,130	1,130	1,130
R-squared	0.014	0.053	0.012	0.004

Note: Clustered standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

can lead to biased estimates. Though our analysis is based on a randomized field experiment, we still include the leads and lags of the treatment effect [37]. Specifically, we add interactions between the treatment dummy and the monthly dummies, and use the interaction with June as the baseline. Figure 3 illustrates the estimated coefficients of these interactions. It is clear that none of the pre-treatment interactions is significant, which shows that the parallel trend assumption is met.

4.2. Treatment Effects on Firms with Security Issues

One possible reason of the insignificant results of PSBL and phishing volume models in Table 4 is that many firms did not have security issues (i.e., zero spam volume or no phishing website) during the period of our experiment. Because security condition is a relatively hard characteristic to observe, it is possible that our existing security measures doe not evaluate all firms' cyber security conditions in a highly accurate manner. Although these firms' security protection levels may have changed, we may lack the ability to precisely measure the difference in our current experiment. Table 6 confirms this. It shows that approximately 40% of all firms in our data showed a positive spam volume based on CBL. However, only approximately 22% of them had a positive spam volume based on PSBL. For the two phishing volume measures, only approximately 5% and 8% of firms had a positive volume based on APWG and OpenPhish, respectively.

For the reasons discussed above, we repeat the main analysis using a subset of firms which have positive outbound spam volume or phishing website counts, respectively. If our treatment emails are effective, we should observe that spam volume or phishing website count from those firms have a larger reduction after the intervention. The results are reported in Table 7. For the first two columns, we only use data from

Table 6: Number of firms in control and treatment groups with positive spam or phishing volume

	Number of firms	Number of firms with positive volume before experiment (treatment)	Number of firms with positive volume before experiment (control)
CV	1,130	228	230
PV	1,130	131	120
AV	1,130	31	27
OV	1,130	46	43

treated firms with either positive CBL or PSBL volumes and their matched control ones. For columns 3 and 4, we only use data from treated firms with phishing websites in either APWG or OpenPhish and their matched control ones. Compared with the data in Table 4, we found that the magnitude of the treatment effect for CBL spam volume is larger. More importantly, the treatment effect for PSBL spam volume is significantly negative at 10% level. This result further indicates that our email treatment will motivate firms to improve their security protection, leading to less outbound spam volume. However, for the phishing performance, we still could not find evidence of a reduction in phishing volume. A possible reason may be the small sample sizes in phishing website hosting data.

4.3. Hosting and Non-hosting Firms' Phishing Websites

There were multiple potential reasons for the overall insignificant treatment effects on phishing website hosting behavior. First, there were only a small number of firms with phishing websites during our study time period. In total, we had 124 firms (in either control or treatment group) which had at least one phishing website in any month. In addition, our phishing measure evaluated the number of phishing websites hosted by the focal firm, and the websites were targeting external entities. In that sense, there may be an externality issue where the associated risk did not directly harm the focal firm. For the hosting service providers, phishing website owners could be considered to be legitimate customers. As a result, web hosting firms might not have a strong incentive to take down the websites in question owned by their legitimate customers. To testify this proposition, we further divide firms into two groups: Group 1 consisted of only Internet service providers and web hosting firms, with the rest being regarded as Group 2. The results support our conjecture that the intervention had a marginally significant effect in phishing website reduction for the firms in Group 2 and that had no effect on the firms in Group 1.

4.4. Overall security performance

So far, we have investigated how firms' security protection evolves after our treatments based on each individual security measure. Another important question to explore is how the treated organizations' overall security conditions change after our interventions. In order to combine the four different security measures from both spam and phishing perspectives, we utilize the ranking data based on Borda count, which we reported both in the treatment emails and on our website.

Table 7: Analysis on subset firms with positive security measures before the experiment

	Sample of positive spam volume		Sample with phishing websites	
	ln(CV)	ln(PV)	ln(AV)	ln(OV)
	(1)	(2)	(3)	(4)
email_treat	-0.430***	-0.128*	0.178*	-0.138
	(0.138)	(0.0708)	(0.107)	(0.120)
Organization fixed effects	yes	yes	yes	yes
Month fixed effects	yes	yes	yes	yes
Constant	3.255***	0.538***	0.335***	0.697***
	(0.0700)	(0.0340)	(0.0471)	(0.0697)
Observations	5,544	5,544	1,200	1,200
Number of organizations	462	462	100	100
R-squared	0.033	0.091	0.109	0.038

Note: Clustered standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

After we have created the Borda count for each firm-month observation, we rank all firms based on the value, by each country or by each industry. Then, we use the rank information as the dependent variable and repeat the DID analysis. The results are reported in the Table 8. The results show that after our experiment, the treated firms' relative security ranking has significantly improved (lower ranking means better security performance). For the industry level ranking, the results are not statistically significant. It seems that compared with other firms in the same country, treated ones have taken measures to improve their security level. In addition, based on the results of monthly interactions in Figure 4, we can see that the main effect happened after the first batch of emails. This also echoes the results in the main analysis that our treatment effects last in a short time period.

5. Research Discussion

In our experiment, we used outbound spam volume and phishing websites as two distinct perceptible cyberattack data sources to measure the pre- and post-experimental cybersecurity risk level of the firms. Security rankings were published on our cybeRatings website[13] to not only enhance the security awareness of the general public, but also to increase economic motivations for firms. From a series of regression analyses on two different types of cyberattacks, we found evidence that the security report publication has a statistically significant effect in reducing spam volume in a short time period. The results showed that publicized security information may compel firms to adopt better preventive measures against spam emission.

[13]cybeRatings: `https://cyberatings.is.cityu.edu.hk`

Table 8: Analysis on firms' security rankings

	Full sample		Sample w/ security incidents	
	Country rank	Industry rank	Country rank	Industry rank
	(1)	(2)	(3)	(4)
email_treat	0.563**	0.177	1.304**	0.447
	(0.276)	(0.231)	(0.532)	(0.459)
Organization fixed effects	yes	yes	yes	yes
Month fixed effects	yes	yes	yes	yes
Constant	35.44***	22.59***	33.74***	22.68***
	(0.181)	(0.144)	(0.313)	(0.255)
Observations	13,560	13,560	5,472	5,472
Number of organizations	0.261	0.135	0.153	0.071
R-squared	1,130	1,130	456	456

Note: Clustered standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

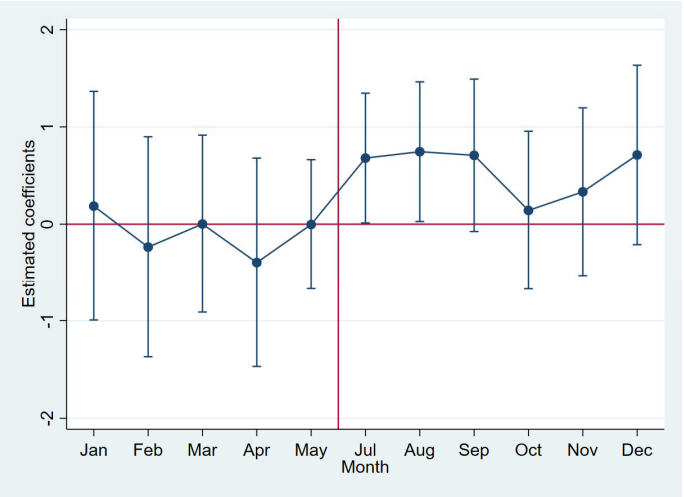

Figure 4: Monthly interaction coefficients for the DID trend test on country level ranking

Interestingly, we did not find a statistically significant effect on mitigating phishing website hosting behavior. There are two possible explanations for this: First, web hosting firms do not have economic incentives to eliminate phishing websites because they are legitimate customers of the hosting services. This can be considered to be a negative externality issue. Second, due to a lack of phishing-related laws and policies, the malicious entities and telecommunication firms face less liability risks for the phishing attacks and resulting damages. Following this line, some ISPs and hosting services may indirectly pass the responsibilities onto their customers. Third, web hosting firms adhere to the non-self-censorship principle. Therefore, they do not content filter web materials uploaded by their customers, allowing phishing website owners to abuse the firms' web hosting services for malicious activities.

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 firms 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 firms who hosted phishing websites, four fully addressed the issues. This result may suggest that the provided information was appreciated and induced a certain level of improvement in the subject's information security condition.

To summarize, our results from the empirical analysis suggest that information security monitoring websites, such as cybeRatings, can be effective in reducing botnet activities represented by outgoing spam volume. Meanwhile, we observed that firms have different incentives in terms of managing phishing attacks. This work may have policy implications in that stronger regulations may be required to internalize the negative externalities resulting from phishing websites hosted by malicious entities.

Apart from legislation, our analysis also shows that public disclosure of information security performance may be an alternative approach to encourage firms to invest in security improvement and adopt better security measures. The primary reason for such an improvement is that by alerting firms of their security vulnerabilities, they are under significant pressure with respect to losing their customers and being surpassed by their peers in the same industries; thus, they are willing to substantially invest in security improvements to prevent future attacks and are more proactive in information security so as to create a better corporate social responsibility image. With all these reasons, public disclosure of information security performance may have direct and indirect effects to encourage firms to invest in information security over time.

6. Limitations and Future Research

One limitation of our current experiment is that the communication channel to subjects was only emails. The emails may only be received by operating staff, rather than customers or investors of the focal firms. As a result, the publicity effect may be limited. As a future direction, we plan to expand our communication channels to social media platforms (e.g., Twitter, Facebook, Linkedin, Weibo, and WeChat). Apart from IT staff members, the social media followers may also be informed of the security evaluation reports with the treated firms. 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 firms.

Another limitation of our study is the focus of the firms in six Pan-Asian countries and regions. A possible extension is to expand the scope of the experiment to firms in other countries. Because our data sources include phishing and spam data from more than 200 countries worldwide, we plan to generate and publicize security reports for other regions. With a

larger sample size, we may be able to test the efficacy of different treatment contents (e.g., security vulnerability index, the index with a list of IP addresses involved in cybercrime, and index with possible countermeasures).

Finally, our security performance measures only include outgoing spam and phishing website hosting. Performance against other common cybercrime (e.g., DDoS) can also be analyzed. In fact, there is a possible spillover effect in our treatment group. Emails alerting firms of potential spam and phishing problems may make them aware of other cybercrime and improve overall security levels to deter other cybercrime as well. These areas may be further studied in future research.

7. Conclusion

The US Department of States and European Commission advocate the use of 3Ps, namely, prevention, protection, and prosecution, to combat crime (e.g., human trafficking and domestic violence). Despite the wisdom contained in the idiom "prevention is better than cure", it is also the weakest link in preventing the wide spread of cybercrime. Due to negative externalities and misaligned incentives, firms may simply choose not to adopt any preventive solutions. To some extent, it is very similar to air pollution in that people who connect insecure computers to the network do not bear the full consequences of their actions and make a poor security investment [3, 21]. In this paper, we show that publicizing a vulnerability index may rectify such misaligned incentives. To some extent, such an approach may achieve similar results to other measures such as legislation [38], subsidy on self-protection [39] and penalties/taxes for non-compliance [9] to heighten public awareness to related cybercrime. Besides, our suggested approach requires lower processing costs (e.g., time and effort to collect evidence for prosecution) and may incentivize firms to uphold origin responsibility, which is one of the four principles of the Bright ICT Initiatives.

8. Acknowledgment

The work described in this paper was fully supported by grants from US National Science Foundation (NSF Award Number: 1718360) and the Public Policy Research Funding Scheme (Project Number: 2015.A1.030.16A) from the Policy Innovation and Coordination Office of the Hong Kong Special Administrative Region Government.

References

[1] M. Sethumadhavan, R. Santanam, and M. Virendra, "Cyber security, cyber crime and cyber forensics - applications and perspectives." Hershey, United States: IGI Global, 2011.

[2] Symantec, "Internet security threat report 2017," https://www.symantec.com/content/dam/symantec/docs/reports/istr-22-2017-en.pdf, accessed: 16-06-2019.

[3] R. Anderson and T. Moore, "The economics of information security," *Science*, vol. 314, no. 5799, pp. 610–613, 2006.

[4] S. Ba, A. B. Whinston, and H. Zhang, "The dynamics of the electronic market: An evolutionary game approach," *Information Systems Frontiers*, vol. 2, no. 1, pp. 31–40, Jan 2000.

[5] L. A. Gordon, M. Loeb, and W. Lucyshyn, "Information security expenditures and real options: A wait-and-see approach," *Computer Security Journal*, vol. 19, 05 2003.

[6] G. A. Akerlof, "The market for lemons: Quality uncertainty and the market mechanism," in *Uncertainty in Economics*, P. D. M. Rothschild, Ed. Academic Press, 1978, pp. 235 – 251.

[7] M. J. G. van Eeten and J. M. Bauer, "Economics of malware," 2008. [Online]. Available: https://www.oecd-ilibrary.org/content/paper/241440230621

[8] N. Shetty, G. Schwartz, and J. Walrand, "Can competitive insurers improve network security?" in *Trust and Trustworthy Computing*, A. Acquisti, S. W. Smith, and A.-R. Sadeghi, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 308–322.

[9] J. K. Lee, "Research framework for ais grand vision of the bright ict initiative," *MIS Quarterly.*, vol. 39, no. 2, pp. iii–xii, Jun. 2015.

[10] J. Lee, D. Cho, and G. Lim, "Design and validation of the bright internet," *Journal of the Association of Information Systems*, vol. 19, no. 2, pp. 63–85, 2 2018.

[11] R. T. Wright, M. L. Jensen, J. B. Thatcher, M. Dinger, and K. Marett, "Research noteinfluence techniques in phishing attacks: An examination of vulnerability and resistance," *Information Systems Research*, vol. 25, no. 2, pp. 385–400, 2014.

[12] S. He, G. M. Lee, S. Han, and A. B. Whinston, "How would information disclosure influence organizations outbound spam volume? Evidence from a field experiment," *Journal of Cybersecurity*, vol. 2, no. 1, pp. 99–118, 12 2016.

[13] A. Bratko, B. Filipič, G. V. Cormack, T. R. Lynam, and B. Zupan, "Spam filtering using statistical data compression models," *Journal Machine Learning Research*, vol. 7, pp. 2673–2698, Dec. 2006.

[14] G. V. Cormack and T. R. Lynam, "Online supervised spam filter evaluation," *ACM Transactions on Information Systems*, vol. 25, no. 3, Jul. 2007.

[15] D. E. Denning, "An intrusion-detection model," *IEEE Transactions on Software Engineering*, vol. SE-13, no. 2, pp. 222–232, Feb 1987.

[16] W. Lee and S. J. Stolfo, "Data mining approaches for intrusion detection," in *Proceedings of the 7th Conference on USENIX Security Symposium - Volume 7*, ser. SSYM'98. Berkeley, CA, USA: USENIX Association, 1998, pp. 6–6.

[17] M. Roesch, "Snort - lightweight intrusion detection for networks," in *Proceedings of the 13th USENIX Conference on System Administration*, ser. LISA '99. Berkeley, CA, USA: USENIX Association, 1999, pp. 229–238.

[18] E. Casey, *Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet*, 3rd ed. Orlando, FL, USA: Academic Press, Inc., 2011.

[19] R. W. Taylor, E. J. Fritsch, and J. Liederbach, *Digital Crime and Digital Terrorism*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall Press, 2014.

[20] S. Kane, "Business security: The difference between protection and prevention," https://bit.ly/2Ij9AeP, 2018, accessed: 16-06-2019.

[21] R. Anderson, "Why information security is hard-an economic perspective," in *Proceedings of the 17th Annual Computer Security Applications Conference*, ser. ACSAC '01. Washington, DC, USA: IEEE Computer Society, 2001, pp. 358–.

[22] R. Sharman, H. R. Rao, S. Upadhyaya, P. Khot, S. Manocha, and S. Ganguly, "Functionality defense by heterogeneity: a new paradigm for securing systems," in *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*, Jan 2004.

[23] D. Acemoglu, A. Malekian, and A. Ozdaglar, "Network security and contagion," *Journal of Economic Theory*, vol. 166, pp. 536 – 585, 2016.

[24] P.-Y. Chen, G. Kataria, and R. Krishnan, "Correlated failures, diversification, and information security risk management," *MIS Quarterly*, vol. 35, no. 2, pp. 397–422, Jun. 2011.

[25] H. Kunreuther and G. Heal, "Interdependent security," *Journal of Risk and Uncertainty*, vol. 26, no. 2, pp. 231–249, Mar 2003.

[26] I. Park, R. Sharman, H. R. Rao, and S. Upadhyaya, "Short term and total life impact analysis of email worms in computer systems," *Decision Support System*, vol. 43, no. 3, pp. 827–841, Apr. 2007.

[27] T. Moore and R. Clayton, "The Impact of Public Information on Phishing Attack and Defense," *Communications & Strategies*, vol. 1, no. 81, pp. 45–68, 2011.

[28] J. Quarterman, L. L. Linden, Q. Tang, and A. Whinston, "Spam and botnet reputation randomized control trials and policy," *TPRC 41: The 41st Research Conference on Communication, Information and Internet Policy*, 2013.

[29] Q. Tang, L. L. Linden, J. S. Quarterman, and A. B. Whinston, "Improving internet security through social information and social comparison: A field quasi-experiment," in *Workshop on the Economics of Information Security (WEIS)*, 2013.

[30] E. Gal-Or and A. Ghose, "The economic incentives for sharing security information," *Information Systems Research*, vol. 16, no. 2, pp. 186–208, 2005.

[31] Symantec, "Internet security threat report 2016," https://www.symantec.com/content/dam/symantec/docs/reports/istr-21-2016-en.pdf, accessed: 16-06-2019.

[32] R. M. Adelsman and A. B. Whinston, "Sophisticated voting with information for two voting functions," *Journal of Economic Theory*, vol. 15, no. 1, pp. 145 – 159, 1977.

[33] J. J. Heckman and J. A. Smith, "Assessing the case for social experiments," *Journal of Economic Perspectives*, vol. 9, no. 2, pp. 85–110, June 1995.

[34] K. L. Morgan and D. B. Rubin, "Rerandomization to improve covariate balance in experiments," *Ann. Statist.*, vol. 40, no. 2, pp. 1263–1282, 04 2012.

[35] A. Abadie, "Semiparametric Difference-in-Differences Estimators," *The Review of Economic Studies*, vol. 72, no. 1, pp. 1–19, 01 2005.

[36] J. Angrist and J.-S. Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, 1st ed. Princeton University Press, 2008.

[37] D. H. Autor, "Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing," *Journal of Labor Economics*, vol. 21, no. 1, pp. 1–42, 2003.

[38] J. D'Arcy, A. Hovav, and D. Galletta, "User awareness of security countermeasures and its impact on information systems misuse: A deterrence approach," *Information Systems Research*, vol. 20, no. 1, pp. 79–98, 2009.

[39] H. Ogut, S. Raghunathan, and N. Menon, "Cyber security risk management: Public policy implications of correlated risk, imperfect ability to prove loss, and observability of self-protection," *Risk Analysis*, vol. 31, no. 3, pp. 497–512, 2011.