JOURNAL OF CYBERSECURITY

Journal of Cybersecurity, 2(1), 2016, 99–118
doi: 10.1093/cybsec/tyw011
Research paper

Research paper

How would information disclosure influence organizations' outbound spam volume? Evidence from a field experiment

Shu He,[1] Gene Moo Lee,[2,*] Sukjin Han[3] and Andrew B. Whinston[4]

[1]Department of Operations and Information Management, University of Connecticut, Storrs, CT 06268, USA; [2]Department of Information Systems and Operations Management, University of Texas at Arlington, Arlington, TX 76019, USA; [3]Department of Economics, University of Texas at Austin, Austin, TX 78712, USA and [4]Department of Information, Risk, and Operations Management, University of Texas at Austin, Austin, TX 78712, USA

*Correspondence address. Department of Information Systems and Operations Management, University of Texas at Arlington, Arlington, TX 76019, USA. Tel: + 1-817-272-3084; Fax: + 1-817-272-5801; E-mail: gene.lee@uta.edu

Received 10 November 2015; revised 1 July 2016; accepted 25 August 2016

Abstract

Cyber-insecurity is a serious threat in the digital world. In the present paper, we argue that a suboptimal cybersecurity environment is partly due to organizations' underinvestment on security and a lack of suitable policies. The motivation for this paper stems from a related policy question: how to design policies for governments and other organizations that can ensure a sufficient level of cybersecurity. We address the question by exploring a policy devised to alleviate information asymmetry and to achieve transparency in cybersecurity information sharing practice. We propose a cybersecurity evaluation agency along with regulations on information disclosure. To empirically evaluate the effectiveness of such an institution, we conduct a large-scale randomized field experiment on 7919 US organizations. Specifically, we generate organizations' security reports based on their outbound spam relative to the industry peers, then share the reports with the subjects in either private or public ways. Using models for heterogeneous treatment effects and machine learning techniques, we find evidence from this experiment that the security information sharing combined with publicity treatment has significant effects on spam reduction for original large spammers. Moreover, significant peer effects are observed among industry peers after the experiment.

Key words: cybersecurity; policy design; randomized field experiments; information asymmetry; peer effects; regression tree; random forest; heterogeneous treatment effects

Introduction

Cybersecurity has become a vital issue: our daily lives, businesses, governments and society at large heavily rely on the Internet. In recent years, the threat from cyber-attacks has been increasingly witnessed around the world. Data from 2013 show that the average cost of security breaches can be as much as 3.5 million—an increase of 15% compared with that in the previous year (Data source: 2014 Cost of Data Breach Study: Global Analysis by Ponemon Institute LLC). According to PWC's global state of information security report, the number of detected incidents increased by 25% in 2013 (http://www.whitehouse.gov/issues/foreign-policy/cybersecurity/eo-13636). In addition, a popular book, "Spam Nation", reported that anti-virus companies are fighting an average of 82,000 new attacks every day [1]. McAfee—which is now Intel Security Group—detected 14 million new pieces of malware in the first quarter of 2013 alone. One conspicuous example that has brought wide public attention is Target Corporation's data breach, which affected 2.6 million consumers during the holiday season in 2013 (The data come from the announcement of Target. The incident caused a significant amount of business and reputation loss to the company.

Target has taken a series of security measures after the data breach, including promoting its system, issuing more secure chip-and-PIN cards, and deploying more advanced technology). The ever-rising trend of cybersecurity incidents—such as data breaches in retail, financial services, and health service companies—which is parallel to the emergence of ever-sophisticated cyber-attacks, calls for more efficient solutions to the problem.

A large body of literature has investigated the root causes and countermeasures to mitigate the cybersecurity issues from technical and economic perspectives. In the economics literature, there is a general consensus that cyber-insecurity is partly due to organizations' underinvestment on security as a result of distorted incentives by asymmetric information, network externalities, and moral hazard [2, 3]. Organizations strategically choose a security protection level that minimizes their private costs, whereas a social planner is motivated to minimize the social cost. Without adequate policy interventions, a socially suboptimal cybersecurity level is thus achieved. One potential solution to this cybersecurity issue is to introduce policies that bolster the overall security of the defender side, namely, organizational-level security protection, and this is the present paper's point of departure. Evidently, given the same level of attacks, a defending organization with stronger protection is less likely to be compromised by the attackers. Although Internet Service Providers (ISPs) are suitable for protection from various malicious cyber-activities [4, 5], many organizations in various sectors, both private and public, do not rely on ISPs and manage their own on-premises infrastructures. Thus, it is particularly important for these organizations to be prepared themselves against cyber-attacks.

The motivation for this paper stems from the following policy question: How can we design policies for governments and other organizations that can ensure a sufficient level of cybersecurity? We argue that current policies or regulations are not sufficient to effectively encourage or force organizations to protect their systems and information. Taking the US federal government as an example, there are three main cybersecurity regulations: the Health Insurance Portability and Accountability Act in 1996, the Gramm-Leach-Bliley Act in 1999, and the Homeland Security Act in 2002. These regulations require health care organizations, financial institutions, and federal agencies to take security measures to protect their systems and data. However, there are no regulations on other sectors, especially for high-technology industries, in which companies manage large amount of valuable data. More recently, President Obama signed Executive Order 13636, "Improving Critical Infrastructure Cybersecurity," which emphasizes the importance of information sharing and cybersecurity framework development (http://www.whitehouse.gov/issues/foreign-policy/cybersecurity/eo-13636). However, programs initiated according to the Executive Order are mostly voluntary (For instance, according to the Executive Order, the US Department of Homeland Security will promote the adoption of cybersecurity framework through a voluntary program with the help of security experts from the private sector), and thus the success of the programs highly depends on how to incentivize organizations to participate in them. More importantly, most government regulations that have been recently introduced tend to be knee-jerk reactions to cyber-threats, under which compromised companies are arbitrarily fined. Such a lump sum fine is not effective, because it would have no binding power to bigger companies and would completely wipe out the profits of smaller companies (A recent article from the Financial Times delivers a similar point: http://www.ft.com/intl/cms/s/0/817f4146-7e4e-11e5-a1fe-567b37f80b64.html).

One method to provide sufficient motivation for organizations is to introduce information disclosure in order to alleviate the cybersecurity information asymmetry [2]. Previous works such as Moore and Claytone [6] and Tang *et al.* [7] found evidence that security information publication helps improve cybersecurity conditions on the country level [6, 7]. In the present paper, we design a fine-grained policy to incentivize security level improvement at the organizational level. Specifically, we propose a nationwide cybersecurity evaluation agency: (i) that monitors and evaluates the security performance of all organizations on the Internet using various data including spam, phishing, and distributed denial of service (DDoS) attacks and (ii) that publishes organizations' cybersecurity evaluation reports to the public. In this way, the evaluation institution would work the same way as Moody's or S&P do for bonds.

The rationale behind this institution is as follows. First, the information disclosure helps reduce the information asymmetry issue within an organization. Due to insufficient internal resources and policies, some organizations with budget constraints may not have a full understanding of their security problems [8]. The proposed institution can alleviate this problem with evaluation reports. Second, 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. Publication of a cybersecurity evaluation report can force organizations to raise their security bars for fear of losing customers from their competitors [7, 9]. Third, a peer-ranking system can allow organizations to make direct comparisons with their industry peers, so that peer pressures can induce overall security improvements.

It is important to evaluate the effectiveness of a proposed public policy, and we present a randomized field experiment (RFE) in the present paper. RFEs are regarded as the gold standard to estimate the causal treatment effects of proposed policy, since, with careful experiment design, it can exclude other confounding factors [10, 11, 12]. Although a series of studies measure the impact of cybersecurity information disclosure on remedies and countermeasures, they are not based on rigorous randomization; thus, it may be hard to claim causal effects of the information disclosure [6, 13–17]. In our experiment, we use outbound spam data as a proxy to estimate the latent security levels of 7919 US organizations (We note that spam is one out of many ways of evaluating security and that other metrics can be used as alternatives, provided they can be externally observable by outside researchers without internal audits. More discussions about the security indicator are in the "Data collection" section). With careful randomization, we divide the subjects into three groups. The first is the control group, to which we take no action. The second is the private treatment group, to which we provide with exclusive security reports via emails. The last is the public treatment group, to which we provide emails with security reports including explicit information that the report is publicly available on our experiment website. The private treatment is to measure the information awareness treatment effects, and the public treatment is to estimate the publicity effects. Our empirical results show that the combination of information and publicity successfully reduces the outbound spam volume of large unwitting intermediaries (i.e. compromised organizations). However, the data show that security information disclosure by itself does not have a significant average treatment effect. To evaluate the heterogeneous policy impacts, we analyze treatment effects for different subgroups using causal tree and causal forest [18, 19]. Furthermore, with the peer effect analysis, we find evidence that organizations' security decisions are influenced by the average outcome of their peers. This interesting finding gives us confidence that our peer ranking system is effective in spam reduction.

The present paper has important contributions to the literature and provides practical implications for policy makers. Our study contributes to the literature by extending prior work on the effects of security information disclosure and by providing potential policies to mitigate Internet insecurity problems. More importantly, our experiment design and policy evaluation analysis could be a road map for public policy evaluation in the cybersecurity area, which can be generalized and extended to other potential security remedies in different environments. Since our current experimental universe includes only US organizations, the conclusions in this paper may not be sufficiently applicable to organizations in other countries with different economic and cultural environments. Researchers and government staff in other countries can follow our large-scale field experiment supported by the cloud computing to design effective policies for their own countries. Finally, with the constructed security metrics, we can potentially set up cybersecurity insurance premiums for cyber risks.

The remainder of the paper is organized as follows. In the "Experiment design and implementation" section, we describe the experimental design for our RFEs, followed by hypotheses development in the "Hypotheses development" section. In the "Empirical analysis" and the "Robustness check" section, we deliver the empirical analysis of our paper. In the "Extending experiments" section, we discuss the future research direction and conclude the paper.

Experiment design and implementation

As discussed in the "introduction" section, one potential solution to the security problem is to alleviate the information asymmetry of cybersecurity. Following previous works such as Moore and Clayton [6] and Tang *et al.* [7], we propose a cybersecurity evaluation agency that actively monitors organizational security levels and shares the evaluation reports to the focal organization and the public [6, 7], thus reducing the cybersecurity information asymmetry problem. Ideally, the institution would monitor all organizations' security performances using externally observable data such as spam, phishing, and DDoS attacks and publish them on its public website. Since the institution evaluates and publicizes the latent security condition for each organization, consumers and investors can make informed decisions by incorporating the available security information.

This proposed institution could be quite costly, considering the large number of involved organizations. Thus, a preliminary evaluation of the proposition's effectiveness is prudent. We conducted a large-scale RFE from January 2014 to March 2014 on 7919 US organizations to see the treatment effects of information sharing and publication on spam reduction, although the potential effectiveness of our experiment would not be so remarkable compared with the "real" proposed institution. To be more specific, we had three treatment groups with two different information disclosure methods to distinguish publicity effect from information notification effect. The whole experiment can be summarized in Fig. 1.

Randomization

Rigorous randomization is needed to extract causality from our experiment. We divide organizations into three equally sized groups—the control group, the private treatment group, and the public treatment group—using a stratified, match-pair randomization [20]. Specifically, we first define 195 subgroups by Standard Industrial Classification (SIC) codes (39 industry sectors) and number of IP addresses (five segments). The detailed groups based on industry sectors and numbers of IP addresses are listed in Tables A1 and A2 in the Appendix. Then, we find clusters of three organizations within each subgroup that minimize the sum of three pairwise differences among them. After that, we randomly assign organizations in each cluster into the control or treatment groups. Finally, we check the distances between the control group and two treated groups with respect to companies' various characteristics (For more detailed information, see the Appendix).

Data collection

As an organization's cybersecurity level is a latent variable, we need to find a good proxy that can be externally observable without any internal audits. Our approach is to use outbound email spam generated from each organization's autonomous system (AS). Spam is defined as unsolicited bulk emails (https://www.spamhaus.org/consumer/definition/). As in Rao and Reiley [21] and Moore and Clayton [6], most spam (over 90%) is sent from botnets, which are the networks of virus-infected computers [6, 21]. These compromised computers may also be used for even worse cyber-criminal activities such as identity thefts, blackmails, and DDoS attacks. Thus, we argue that organizations' large outbound spam volumes are an important indicator of their weak security levels.

Many security organizations maintain spam blocklists to blacklist IP addresses that actively engage in spam emission. These organizations install various spamtraps, which are email-receiving machines without valid users. Thus, any computers that attempt to conduct email activities with the spamtraps are suspicious spammers. Since these monitoring activities are done in the transmission

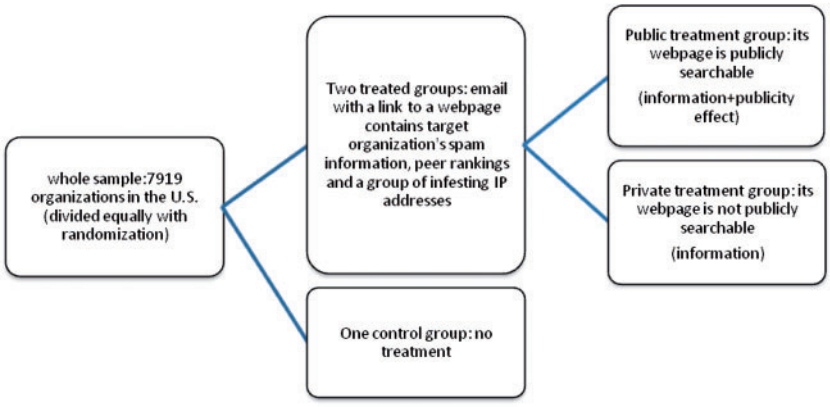

Figure 1. Design of randomized field experiment.

control protocol level, spammer IP addresses are verified. In other words, spam-sending IP addresses are not spoofed. In order to overcome the potential biases on heterogeneous spamtrap settings [22], we use two independent spam data feeds: Composite Block List (CBL) (http://cbl.abuseat.org) and Passive Spam Block List (PSBL) (http://psbl.org). Both blocklists provide us the list of spamming IP addresses and their associated spam volumes on a daily basis.

From this raw IP-level data, we need to construct organization-level data to evaluate each organization's security condition. To do that, we need three levels of mapping: from an IP to a netblock, to an autonomous system number (ASN), then finally to an organization. The data are based on the IPv4 address space, which uses classless inter-domain routing (CIDR) (http://tools.ietf.org/html/rfc1519). Thus, the IP to netblock mapping, also known as IP lookup, is a longest prefix matching problem, which has a well-known efficient algorithm [23]. For netblock to ASN mappings, we receive daily netblock-ASN data feeds from Team Cymru through remote synchronization (rsync) (http://www.team-cymru.org). On average, we have 584,000 netblocks and 49,000 ASNs in our mapping data. Finally, for ASN to organization mappings, we group ASNs by the operating organization with manual inspections. Note that we process only ASNs located in the USA for this experiment. For each ASN group, we then find the corresponding organization by searching the LexisNexis database (http://www.lexisnexis.com). As a result, 7919 US organizations are identified for the experiment.

One may argue that spam is not a major cybersecurity threat for organizations and that it can be largely solved using filtering and blocklists. Also, there can be questions about whether outbound spam volume is a comprehensive index of organizations' security level. However, we want to emphasize that, in our current experiment, spam information is only a tool we apply to see how organizations' security strategies will respond to our interventions. Given the fact that spam is perceived as a less dangerous risk for those organizations, we would expect that the treatment effects will be larger and more significant if the security evaluation reports include other cyber-attacks such as phishing and DDoS attacks.

In addition to the outbound spam data for each organization, we collect organizations' characteristics from the LexisNexis database, including the industry codes (SIC and NAICS) and whether the organization is publicly traded.

Experiment treatments

Among the three groups in our experiment, the first one is the control group, to which we do not apply any treatments. For the two treatment groups, we send treatment emails to relevant contacts in various departments (from marketing to IT) within each organization to inform them of their security evaluation reports. Treatment emails were sent at the end of January and March 2014. Each treatment email included (1) the organization's spam volume, (2) peer rankings, (3) a partial list of spamming IP addresses, and (4) a hyperlink to a designated web page for the treated organization. The difference between the private and public treatment groups is whether the information of the focal organization is publicly searchable on our treatment website. For the publicly treated organizations, the emails clearly mention that the spam information is publicized on our treatment website (The link of our website: http://cloud.spamrankings.net/). On the other hand, the privately treated organizations are notified that the web page directed by the link in the email is not publicly available. With this setting, the difference of average spam volumes between the control and the private treatment groups is due to the information awareness effect. Similarly, we can estimate the publicity treatment effect with the difference between the private and public treatment groups.

Spam ranking for peer effects

In the security evaluation reports provided by the treatment emails and website, we have our own peer ranking, which is different from security rankings on other existing security evaluation websites. Essentially, organizations within an industry sector are ranked according to their spam metrics (Sectors are defined by the two digits in two industry codes: SIC and North American Industry Classification System [NAICS]. Note that high ranks indicate a low security level and that all of the organizations with no spam will be ranked equally with the lowest rank).

Currently, there are only a handful of websites that publish spam information such as CBL, Spamhaus, and Cisco. These rankings provide information only for "top spammers." And most of their information is based on the unit of AS rather than organization (Classic.SpamRankings.net presents the top 10 spammers per country [http://www.spamrankings.net/classic/]. Spamhaus posts top 10 spam-producing countries, ISPs, and spammers each day [http://www.spamhaus.org/statistics/countries/]. Cisco, on the other hand, has at most the top 100 spam senders by IP, network owners, and country [http://www.senderbase.org/static/spam/]). Furthermore, most companies are more likely to reactively disclose information security issues in case of compromised customer information. This may lead to underestimated information risk. Most importantly, existing websites do not provide industry rankings. In other words, an organization cannot directly compare its performance with its close competitors. This lack of comparative information may weaken peer effects.

Our peer ranking helps an organization to better evaluate its security performance against its competitors. The rationale of constructing peer ranking is as follows. First, there is substantial heterogeneity across different industry sectors. For example, companies in financial and health sectors may have more sensitive customer information that attracts more threats from cyber-attackers. More importantly, it is well known that individuals' and organizations' behaviors are likely to be influenced by their peers [24]. Our peer ranking can potentially provide a channel to enhance the peer effects among organizations in the same industry sector, and can further have an impact on organizational behavior. We demonstrate the existence of peer effects in the following empirical analysis.

Hypotheses development

Information disclosure effect

The information disclosure effect refers to the treatment effect of spam information provided in our treatment emails for organizations that previously neglected the importance or did not have a full understanding of the security conditions due to insufficient internal resources and policies [8]. In our present experiment, we send out organization-specific spam report via email to each organization in our treated groups. The detailed spam information includes spam volumes, number of spamming hosts, specific infested IP addresses, compositions of spam volumes over time, as well as its relative performance (peer ranking) compared with close competitors within the same industry. After receiving our emails, organizations without good prior knowledge of their security levels can be better informed. In addition, they also get information (e.g. infested IP addresses) that helps them quickly isolate the problems. If our email treatment with security information is helpful to treated organizations, we

would expect the spam volume of organizations in the private treatment group, who only receive private emails from the researcher, will decrease, when compared with that in the control group. Hence, we hypothesize:

> **Hypothesis 1.** *There will be a significant decrease of spam volumes after the experiment for organizations in the private treatment group as compared with those in the control group due to the spam information disclosure in the email treatment.*

Publicity effect

The publicity effect refers to the treatment effect on the public treatment group by publishing security evaluation reports on our public treatment website. Due to security information asymmetry [2], it is difficult for customers and investors to get relevant security information for a focal organization. Thus, organizations may lack motivation to make sufficient investment in cybersecurity, especially when the cost of cybersecurity improvement is relatively higher than the expected cost of data breaches. Security information publication can alleviate the information asymmetry problem since we provide the public access to more detailed security information. In this way, customers and investors can reevaluate their choices with more transparent information; furthermore, our treatment website creates extra cost to the organizations through the threat of reputation damage and customer loss [7, 9]. If our publicity treatment is effective, we would expect to see a greater decrease of spam volume for organizations in the public treatment group, who receive both information sharing and publicity treatments, than for those in the private treatment group. We therefore propose the following hypothesis:

> **Hypothesis 2.** *After the experiment, there will be a significantdecrease of spam volumes for organizations in the public treatment group as compared with those in the private treatment group, due to the spam information publicity treatment.*

In addition, organizations with large outbound spam volumes may have security problems that are relatively easy to isolate and resolve. From the reputational aspect, organizations with different ranks may have heterogeneous levels of pressures from the publicity treatment. Low-ranked organizations can be more embarrassed with the publicity, whereas highly ranked organizations may view it as praise, even though they still have positive outbound spam volumes. These possible interpretations may lead to increased motivation for large, unwitting intermediaries for spammers. Thus, our policy interference may be more effective for organizations with larger pre-experimental spam volumes.

> **Hypothesis 3.** *Organizations in the public treatment group with higher spam volumes will have larger spam volume drops after the experiment.*

Peer effects

A peer effect refers to the change of an organization's security level that is influenced by its peer organizations' performances. Theoretically, a peer effect is driven by reputational concerns, observational learning, and other factors [25]. For example, organizations in the same industry may have technical knowledge exchange among their employers. Researchers have investigated peer effects in wide variety of individual and corporate outcomes, including academic achievement [26], product adoption [27], stock market behavior [28], dividend payment [24], and managerial decision making [29]. In our case, organizations' security strategies can also be influenced by their peers.

In the security evaluation report, we try to induce peer effects by providing industry rankings in addition to general spam metrics. With the industry rankings, organizations and their customers can make direct comparisons with competitors. Hence, an organization may change its cybersecurity strategies in response to its peer organizations' security performance. Therefore, we hypothesize:

> **Hypothesis 4.** *Organizations' outbound spam volumes areinfluenced by their peers' performance after the experiment.*

Empirically identifying the existence of a peer effect is important in understanding the mechanism by which our treatment influences organizations' security strategies. Although we do not have an advanced experiment design of randomization on peer ranking, if publicity's only effect is to embarrass the focal organizations, then peer ranking may not be necessary for an effective policy. In addition, the outcome of the treatment may be different. With the existence of peer effect, organizations' security protection levels may tend to converge to the center. In other words, organizations with the best initial security levels may lower their guard after the publicity. Also, our treatment effects estimated from the experiment will be those for the treated organizations. An ideal way to check the effect of our peer ranking would be an experiment with random treatment of peer ranking. However, due to the interactions among organizations, it is difficult to design such an experiment.

Empirical analysis

Descriptive statistics

Changes in the outbound spam are the basis of our experiment, but the spam volumes fluctuate dramatically from month to month. Although the most relevant reason for the outbound spam volume changes is the change of organizational security levels, there could be alternative reasons, such as the change of spam demand in the black market and botnets' strategic change of target victims to avoid being detected. Thus, we use the average spam volumes over multiple months in the statistical analysis. Our data show that more than half of the organizations with positive spam volumes have experienced one or two spamming episodes a year. Therefore, we use the 6-month average spam volumes right before the experiment started as the pre-experimental spam volumes. Since our experiment started at the end of January 2014, we regard the time frame between July 2013 and December 2013 to be the pre-experimental period, and the one between February 2014 and July 2014 to be the post-experimental period.

We use the natural logarithm transform for the outcome variables (monthly spam volumes and spam hosts) and the covariate (number of IP addresses). This is because the distributions of these outcome variables and the covariate are highly positively skewed, as shown in Fig. 2. The power of the experiment has significantly improved with the natural logarithm transform.

From the experimental data, we observe that the spam volume of all organizations in our sample decreases on average after the experiment. This may be due to the rapid increase of data breach announcements at the end of 2013. These incidents attracted much attention from the public, so organizations generally became more cautious about cybersecurity. In addition, the difference between pre- and post-experimental spam volumes is quite heterogeneous across organizations.

From Fig. 3, we see that organizations with zero (quantile 1) or small initial spam volumes (quantile 2) have more outbound spam after the experiment started, whereas top 25% spammers' (quantile 4)

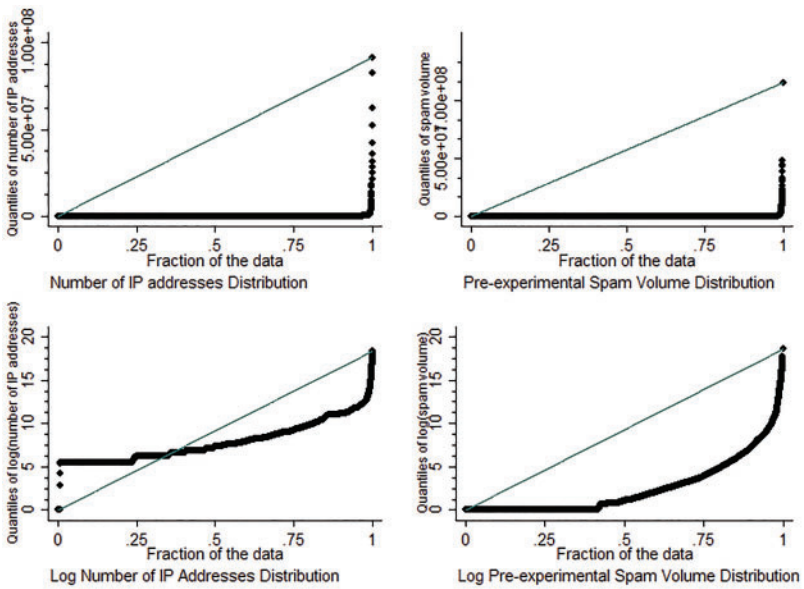

Figure 2. Distributions of spam volumes and numbers of IP addresses.

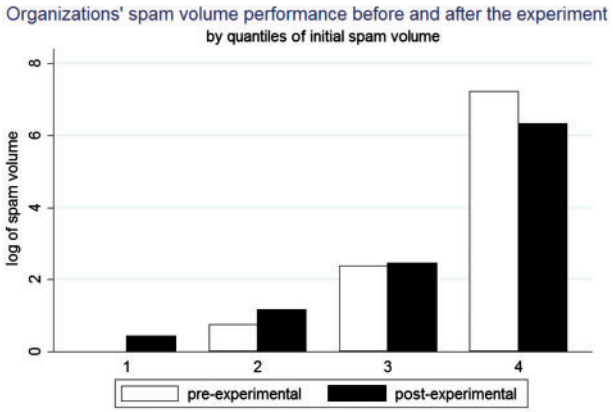

Figure 3. Spam performance within each quantile for all organizations.

outbound spam volumes have decreased. This result may be due to the fact that small spammers, especially the organizations with zero spam volumes, could hardly improve their security condition any more. On the other hand, large spamming organizations in the treatment groups will face the risk of losing customers and investors with our experiment, leading to more cautious cyber security. We also observe that the spam performance of organizations varies among different industry groups, as shown in Fig. 4. This finding can be explained by the distinct business models and characteristics of different industries.

The summary statistics of the related variables are listed in Table 1.

Balance on observables

The advantage of a RFE is that the random assignment ensures the exogeneity of the treatments and the exclusion of selection bias [10]. In the randomization process, each organization has the same probability to be in one of the three groups; hence, on average, organizations in the control and treated groups have homogeneous characteristics. However, it is well known that a pure random assignment may have a probability of imbalance along some dimensions [30].

To ensure that our randomization successfully balances on observables, we conduct two tests for validity. With the RFE setting in the "Experiment design and implementation" section, we have three groups G_i based on two treatments (T_{1i} and T_{2i}) as follows:

$$G_i \begin{cases} 1 \text{ if } T_{1i} = 0 \text{ and } T_{2i} = 0 \\ 2 \text{ if } T_{1i} = 1 \text{ and } T_{2i} = 0 \\ 3 \text{ if } T_{1i} = 1 \text{ and } T_{2i} = 1 \end{cases}, \tag{1}$$

where $T_{1i} = 1$ indicates that organization i receives treatment emails, and $T_{2i} = 1$ indicates that organization i's security evaluation report is publicized in the treatment website. We run regressions of pre-experimental characteristics of organizations on the treatment assignments using the following formula:

$$X_i = \theta_0 + \theta_1 T_{1i} + \theta_2 T_{2i} + \phi_i, \quad \phi_i \sim N(0, \sigma^2), \tag{2}$$

where X_i represents organization i's characteristics (listed below) before the experiment, T_{1i} is a dummy variable indicating whether organization i is privately treated or not, and T_{2i} is the public treatment dummy. We also apply a Kolmogorov–Smirnov (K–S) test and calculate the difference in the normalized standard deviation to check the balance based on the whole distribution of X_i. The results are shown in Table 2. We see that the differences of the average characteristics between the treatment and control groups are marginal, and none of them is statistically significant. Therefore, our randomization groups are balanced.

Average treatment effect analysis

First of all, we would like to see the average treatment for all companies in our data set. We use the linear model to estimate the coefficients in our model as follows:

$$Y_i = \alpha_0 + \alpha_1 T_{1i} + \alpha_2 T_{2i} + \alpha_3 X_i + \epsilon_{1i}, \ \epsilon_{1i} \sim N(0, \sigma_1^2), \tag{3}$$

where Y_i is the spam volume for organization i post-experiment, X_i is the k-dimensional vector that represents organization i's characteristics, such as pre-experimental spam volume, pre-experimental number of spamming IP addresses, number of IP addresses, number

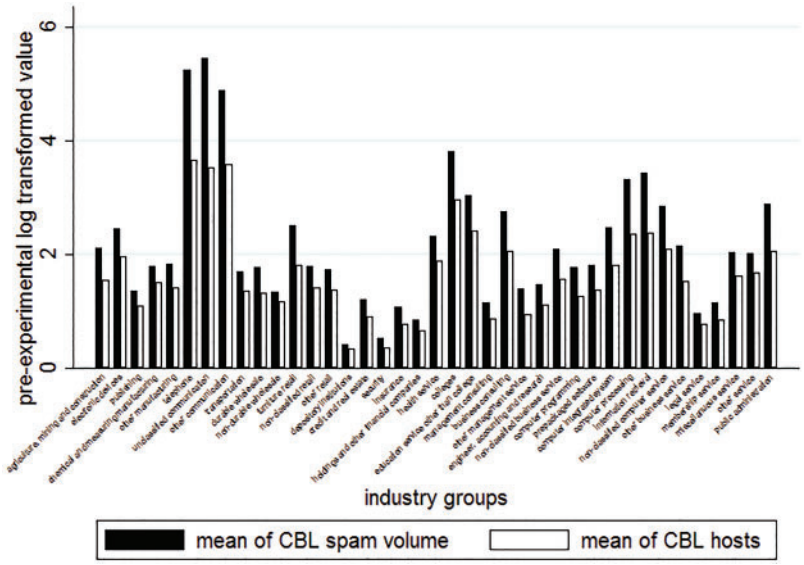

Figure 4. Spam performance in different industry groups.

Table 1. Summary statistics

Variable	Observation	Mean	Standard deviation	Min	Max
log (Post-experimental spam volume+1)	7919	2.469	3.139	0	17.913
log (Post-experimental spam host+1)	7919	1.830	2.072	0	12.134
log (Pre-experimental spam volume+1)	7919	2.474	3.258	0	18.566
log (Pre-experimental spam host+1)	7919	1.738	2.064	0	12.261
log (Number of IP addresses)	7919	7.807	2.289	0	18.333
Number of infesting botnets	7919	1.175	2.677	0	40
Publicly traded or not	7919	0.0885	0.284	0	1
log (Number of employees)	7021	1.410	0.605	0	2.860

of IP addresses squared, whether the organization is publicly traded or not, number of observed botnets, and industry dummies.

The results are reported in Table 3. Columns 1 and 2 present the results from the regression model. As expected, all treatment effects are negative, and the magnitude of public treatment effect is larger than that of private treatment effect. However, the estimated treatment effects lack statistical significance.

Most coefficients of the control variables are significant with expected signs. Organizations with more pre-experimental spam volumes and larger number of botnets generally have more post-experimental spam volumes, which can be evidence that the spam volumes are a consistent indicator for the organizational latent security level.

Another interesting finding is that the relationship between spam volumes and numbers of IP addresses is concave: As the number of IP addresses increases, the spam volume first increases, then it decreases. The estimated largest spammer will have about 60,0000 IP addresses. This phenomenon can be explained by two opposing forces. On the one hand, organizations with large numbers of IP addresses have wider attack surfaces because (i) institutions with large IP counts generally have more potential targets for bot herders (Bot herders are hackers who install malwares on victims' computers to gain unauthorized controls. https://www.fbi.gov/news/stories/2007/june/botnet_061307/) and (ii) it costs more to maintain and protect the system. On the other hand, larger organizations may have stronger security protection since many of them are high-tech companies that have more resources for security investment. We also estimated the regression with industry dummies defined by two-digit SIC codes.

Heterogeneous treatment effects

Due to the heterogeneity and diversity of organizations in our sample, our interference may result in different outcomes among organizations. In addition to the overall policy impact estimation, it is also important and significant for us to see which organizations have been mostly influenced by the policy. As we observed in Fig. 3, only large spammers tend to have reduced spam volumes after our intervention. Spam volumes for smaller spammers—especially for the initially "clean companies" actually increased. To see how the treatment effects vary among different organizations, we use non-parametric causal tree [18] and causal forest [19] to estimate the heterogeneous treatment effects. We first use our results from regression tree as a guidance to split the sample within which we calculate the heterogeneous treatment effects presented in columns 3-6 of Table 3. One common concern of exploring the heterogeneous treatment effects is that researchers will arbitrarily divide samples into subgroups, searching for extreme subsample results. The new methods we use are data-driven that do not need any subjective restriction or judgment. We also use random forest to reduce variance of the key treatment effect estimates. The models are revised regression tree and revised random forest with a modified criterion for splitting the data set. Rather than using the dependent variable as the

Table 2. Baseline comparison for internal validity

	No control		Industry fixed effects		K–S prob.		Ln(St/Sc)	
Dependent variables	Private	Public	Private	Public	Private	Public	Private	Public
Pre-experimental spam volume	−0.00016	−0.005447	−0.00476	−0.005676	1.000	0.998	0.008146	0.01713
	(0.03833)	(0.03646)	(0.03542)	(0.03357)				
Pre-experimental number of infested IP addresses	−0.00548	0.0009089	−0.00996	−0.000657	1.000	0.997	−0.01454	0.01798
	(0.02933)	(0.01993)	(0.02791)	(0.01989)				
Number of IP addresses	−0.03488	0.02178	−0.04225	0.01687	0.891	0.997	−0.02775	−0.03337
	(0.04453)	(0.04150)	(0.04199)	(0.03893)				
Number of botnets	0.003922	−0.005299	0.001906	−0.003606	1.000	1.000	−0.00112	0.02757
	(0.04076)	(0.03692)	(0.03943)	(0.03397)				
Publicly traded or not (=1 if yes)	−0.00145	−0.006752	−0.00206	−0.007416	1.000	1.000	−0.01439	−0.06947
	(0.00707)	(0.00705)	(0.00702)	(0.00716)				

Notes: This table presents comparisons of organizations' characteristics in the control and treatment groups. Columns 1 and 3 contain estimates of the average differences in characteristics between the control and private treatment organizations, without controls and with industry group fixed effects. Columns 2 and 4 contain estimates of the average differences in characteristics between the control and public treatment organizations, without controls and with industry group fixed effects. Columns 5 and 6 contain statistics from Kolmogorov–Smirnov test. Columns 7 and 8 contain the differences in normalized standard deviations between the treatment and control groups. Standard errors are clustered by industry group and shown in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 3. Treatment effect estimation

	Avg. treatment effects		Heterogeneous treatment effects			
	Overall		Private versus control		Public versus control	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Private treatment	−0.0154	−0.00811	0.258	0.149		
	(0.0868)	(0.0336)	(0.2440)	(0.1010)		
Public treatment	−0.0607	−0.061			0.310*	0.259***
	(0.0862)	(0.0408)			(0.1740)	(0.0663)
Indicator			−6.887***	−0.688***	−5.154***	−0.539**
			(0.1780)	(0.1670)	(0.1270)	(0.2180)
Private treatment × indicator			−0.265	−0.173		
			(0.2530)	(0.1050)		
Public treatment × indicator					0.309*	0.265***
					(0.1820)	(0.0755)
Pre-experimental spam volume		0.547***		0.542***		0.493***
		(0.0230)		(0.0174)		(0.0169)
Number of IP addresses		0.393***		0.333***		0.386***
		(0.0552)		(0.0636)		(0.0644)
(Number of IP addresses)^2		−0.0144***		−0.0112***		−0.0140***
		(0.0036)		(0.0042)		(0.0040)
Number of botnets		0.282***		0.231***		0.289***
		(0.0406)		(0.0368)		(0.0330)
Stock		0.0724		0.189***		0.0285
		(0.0671)		(0.0579)		(0.0945)
Intercept	2.494***	−1.639***	8.696***	−0.639**	6.277***	−1.103***
	(0.0615)	(0.2010)	(0.1720)	(0.2930)	(0.1220)	(0.2430)
Industry	No	Two-digit SIC	No	Two-digit SIC	No	Two-digit SIC
P-value for H_0: $a_1 = a_2$	0.7363	0.322				
Observations	7919	7919	5280	5280	5280	5280
R-squared	0	0.744	0.434	0.751	0.501	0.745

Notes: This table displays the estimated private and public treatment effects with OLS model. Columns 1 and 2 report the estimates of the differences between the spam volume of treatment groups and control controlling for pre-experimental spam volume, number of pre-experimental IP addresses, number of pre-experimental IP addresses squared, number of pre-experimental infesting botnets, whether or not publicly traded, and industry fixed effects. Columns 3 and 4 report the estimates of the heterogeneous treatment effects with organizations in control and public treatment group. The indicator equals to 1 if organization's log pre-experimental spam volume is less than 3.6. Columns 5 and 6 report the estimates of the heterogeneous treatment effects with organizations in control and private treatment group. The indicator equals to 1 if organization's log pre-experimental botnet is less than 3.4. Standard errors are clustered by industry codes and shown in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

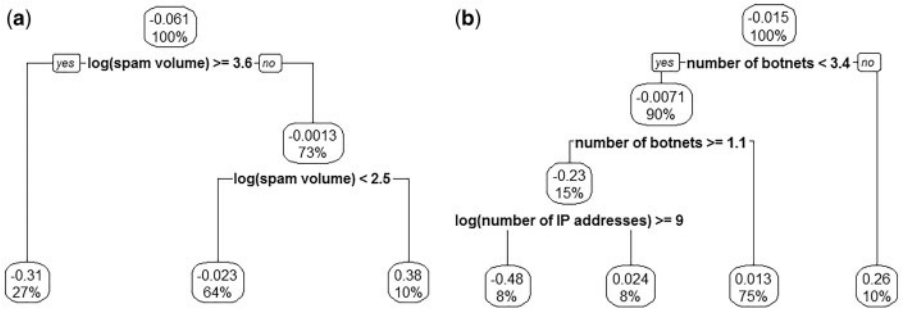

Figure 5. Result from causal tree for organizations in our experiment. (a) Public versus control and (b) private versus control.

criterion, the causal tree and causal forest use the average treatment effect within each leaf. The confidence interval for the causal forest estimator could be calculated following an infinitesimal jackknife procedure [31]. These models allow us to estimate the conditional average treatment effect (CATE) for subset sample (We would like to thank Susan Athey for the R codes of causal forest and causal tree).

Causal tree

In traditional machine learning techniques, a regression tree is a predictive model that maps observations to an item's target value where the target value is continuous. Fundamentally, a tree is learned by splitting the source set into subsets in a recursive way. This process will complete when splitting will not add value to the prediction. Athey and Imbens [18] revised the criteria function to find that in splitting the whole sample, the treatment effect of each observation within each "leaf" would be the same.

The results of a causal tree for our whole sample are reported in Fig. 5. We can see that for the subsample with organizations only in the control and public groups, the main characteristics that may influence the treatment effect are organizations' pre-experimental spam volume and number of IP addresses. For the subsample with only the private and control groups, the essential variables that may lead to heterogeneous treatment effects are the pre-experimental number of botnet and number of IP addresses.

With the causal tree splitting available, for each leaf corresponding to subset X_m, the average treatment effect within the leaf is:

$$\tau|_{X_m} = E[Y_i(1) - Y_i(0)|x_i \in X_m]. \quad (4)$$

As in our experiment randomization design, we did the pair-wise matching on organizations' pre-experimental spam volumes. Companies that have similar pre-experimental spam volumes have the same probability to be in the control group or in any of the two treated groups. As a result, when we divide the whole sample based on pre-experimental spam volume, we can directly use the linear model to compare the outcomes of control and treated organizations in each leaf to estimate the CATE. In particular, we add a dummy representing whether one organization's log pre-experimental spam volume is less than 3.6, which is the cutoff resulted from the regression tree, to show whether we would find heterogeneous treatment effects among different subgroups. The estimation results are reported in Table 3. Columns 3 and 4 show the estimation results with the subsample of organizations in the control and private groups, Columns 5 and 6 show the estimation results with the subsample of organizations in the control and public groups. As we can see, for the subsample of organizations whose log pre-experimental spam volumes are larger than 3.6, public treatment significantly reduces organizations' outbound spam volume, an indicator of security condition improvement. However, for the subsample of organizations in the control and private groups, we do not observe significant CATE after controlling other variables.

To ensure the causality, it would be more reliable to get the treatment estimator using honest splitting, which means to get the splitting and treatment effect from different samples. Specifically, we could randomly choose a subsample from the data set to find the optimal splitting first. Then, this splitting could be applied to estimate the heterogeneous treatment effects from another subsample, without the concern of sample randomness [18]. The main issue with this method is that it is hard for us to choose the optimal training and estimation subsamples.

Causal forest

Unlike one single causal tree, a causal forest is composed of B such trees, which is only trained by a random subsample of organizations. If each causal tree, indexed by b, gives us an estimate of the CATE at x as $\hat{\tau}_b(x)$, we could calculate the random forest CATE at x by averaging the treatment effect over B causal trees: $\hat{\tau}_b(x) = B^{-1}\sum_{b=1}^{B}\hat{\tau}_b(x)$. According to Breiman [32], this aggregation process over many trees helps reduce the variance of the estimates.

In our present example, we set the number of trees to be 2000, and estimate the treatment effect separately for the public and private treatment groups. The estimated treatment effects and T-values are reported in Figs 6 and 7, respectively. We find interesting patterns for the heterogeneous treatment effects. For the public treated group, we see positive treatment effects for organizations with an initial low outbound spam volume, especially for those whose log average pre-experimental outbound spam volume is less than about 3.6 (as in our result in the "Causal tree" section). For organizations with larger initial outbound spam volume, the majorities have presented negative treatment effects. The results support Hypotheses 2 and 3. For the private treatment group, we do not observe significant heterogeneity in treatment effects among organizations with various pre-experimental spam volumes.

There can be multiple reasons why large spammers in the public group will send out smaller outbound spam volumes after receiving our treatment. First, these organizations may have larger but shallower problems (e.g. compromised computers) with their information systems. As a result, large spammers have a lower cost to improve their security conditions than their counterparts with subtle security issues. Second, these organizations may be more embarrassed when their poor security performance reports are publicly announced online. In addition, large spammers face more pressure from their close competitors. Considering their

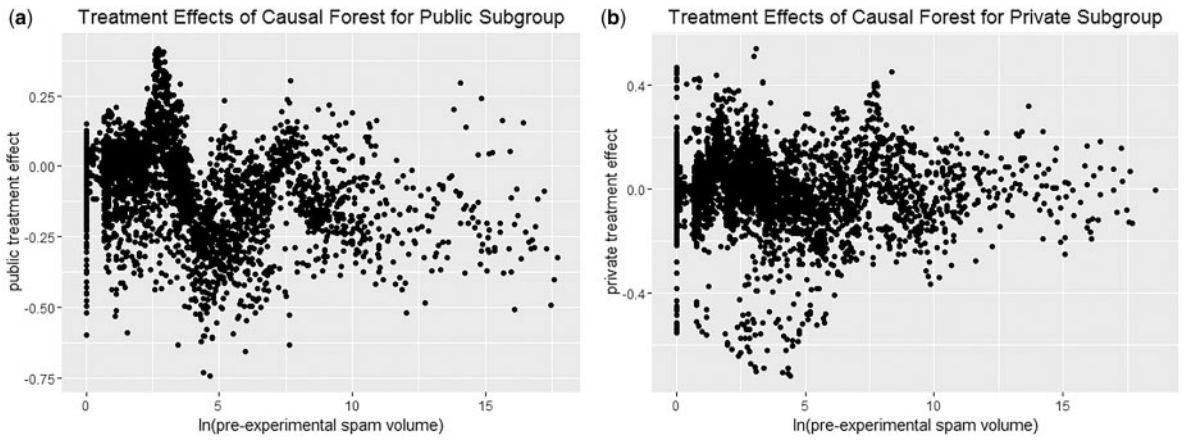

Figure 6. Results of treatment effects from causal forest for organizations in our experiment. (a) Public versus control and (b) private versus control

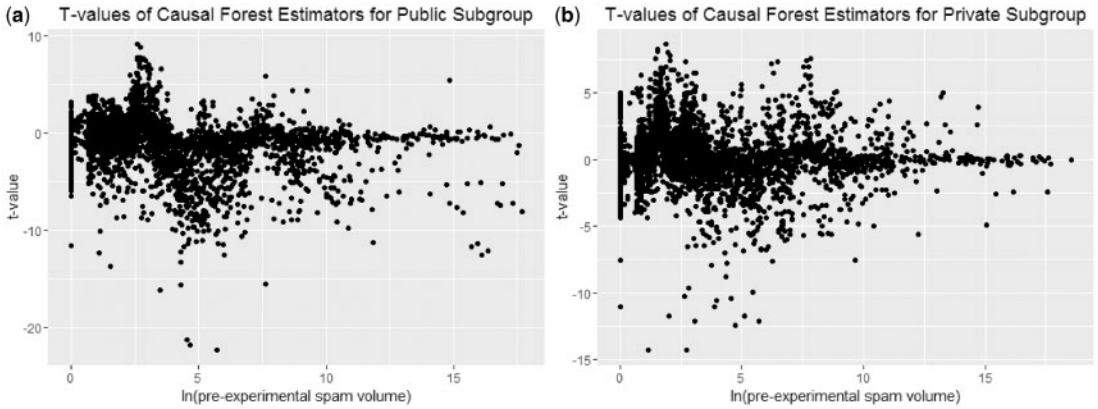

Figure 7. Results of T-values from causal forest for organizations in our experiment. (a) Public versus control and (b) private versus control

relatively worse security levels, it is likely their customers will shift their business to their competitors. We do not observe similar results from the private treatment effect, indicating that sending treatment emails by itself lacks effectiveness in spam reduction and changing organizational behavior on security measures. Our results show that the combination of information sharing and public announcement provides more economic motivation to the organizations.

Discussions on empirical results

Although our work is a major first step forward to test the idea of a security evaluation agency, our experiment has some limitations that may undermine the treatment effects. However, the discovered treatment effects can be further amplified if we mitigate the existing shortcomings with the proposed nationwide independent institution. The first limitation is the security evaluation metric; we only consider outbound spam as a proxy for the latent security level. However, organizations may pay more attention to other types of cyber-attacks, such as phishing and DDoS attacks, since these are more dangerous to organizations' cybersecurity. Although the existing results with outbound spam are still valid, we expect organizations to be more responsive to our intervention if more comprehensive security reports are provided in the experiment. The second limitation of the experiment is the visibility of our website. We had limited time to promote our website to attract attention, which may have undermined our treatment effect given the importance of the reputation effect. Also, some organizations may not have paid enough attention to our emails. Fortunately, this limitation will be largely alleviated if the website is sponsored by the government. We want to note that the estimated treatment effects are based on just two waves of treatment emails at the end of January and March 2014. With constant and longtime notifications, the influence of our treatment may increase over time.

To sum up, the data from the experiment show that large spammers in the public treatment group sent out significantly smaller outbound spam volumes after exposure than those in the control group. We expect the reason is that organizations in the public treatment group have taken active measures to clean up existing malwares and to improve their security protection for new malware prevention. Our proposed policy with security information evaluation and publicity effectively improve organizations' security protection with respect to the outbound spam volumes.

Peer effect analysis

From the previous results, we see that security information publicity induced organizational security improvement, which is measured by the reduced outbound spam volumes. We do further analysis to recover the underlying mechanisms of organizations' security strategies. Organizations may improve their security protection due to the shame of being spammers. On the other hand, they may also change their strategies due to the peer pressure from their close competitors. If customers and investors of an organization are aware of

Table 4. Peer effect analysis on spam volume and number of spamming hosts

	Spam volume		Number of spamming hosts	
	SIC	NAICS	SIC	NAICS
γ^2	2.021	2.179	2.095	2.682
P-value for H_0: $\gamma^2 = 1$	0.0002***	0.0084***	0.0000***	0.0004***
γ	1.422	1.476	1.447	1.638
Organization-specific covariates	Yes	Yes	Yes	Yes
Peer organization average covariates	Yes	Yes	Yes	Yes
Observations	7919	7919	7919	7919

Notes: This table displays the estimated peer effects using excess variance approach. Columns 1 and 2 represent the results using outbound spam volume. Columns 3 and 4 represent the results using number of spamming hosts. We use two-digit SIC and NAICS codes to define peer groups. We use bootstrap to test the null hypothesis of no peer effects for 5000 samples. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

competing companies' having better security levels, the poorly performing organizations may experience churns.

With our peer ranking information available, we provide organizations a convenient way to compare their security levels with those of their peers, thus enhancing the peer influence in security management. The existence of peer effects is important in understanding organizations' security strategies. If peer effects are important, then providing more comparisons between peers may be more effective in pushing organizations to invest resources on their security protection. At the same time, organizations with strong security protection may lack motivation to correct existing problems since they are already in the lead.

Peer effects exist if organizations' behaviors are influenced by their peers' mean outcomes, which, in our context, represent the industry sector's average security level. The identification of peer effects is difficult due to the reflection problem, unobservable variables, and selection problem [33]. To overcome the difficulties, we implement the excess-variance approach identification strategy.

Excess-variance approach

The identification strategy we use to analyze the existence and magnitude of peer effects is the excess-variance approach [24, 34]. The main idea of this method is to take advantage of various sizes for each industry group and the mathematical identity that the variances and sizes do not change in the same proportion. The intuition is as follows. The unconditional between-group variance is equal to the sum of (1) the variance of group-level heterogeneity (different industrial characteristics), (2) between-group variance of individual-level heterogeneity (average organizations' characteristics), and (3) the strength of peer effects. With different sizes of industries, although the distribution of group-level heterogeneity is the same, we can use a method similar to difference-in-differences to compare the between- and within-group variance from different-sized industries to estimate the peer effects. Since organizations are not randomly assigned to different industry groups, the main issue in applying this identification method is that the results may be biased if self-selection also makes the variance change disproportionately to group size. We believe that it is not a main issue to be considered since cybersecurity is not a major factor to consider in making a decision to enter the market. Moreover, it is hard to imagine that organizations will sort them into peer groups differently, based on the group sizes. For example, financial services, retail organizations, and ISP will face a high risk of potential cyber-attacks, but the sizes of the three industry groups vary a lot, as shown in Fig. A1 in the Appendix.

With the typical linear-in-means model [33], organization i's spam behavior from industry j, D_{ij}, will be:

$$D_{ij} = \alpha_j + (\gamma - 1)\bar{\epsilon_j} + \bar{\epsilon_{ij}}, \tag{5}$$

where α_j represents industry-level heterogeneity, ϵ_{ij} represents organization-level heterogeneity, and $\bar{\epsilon_j}$ represents the industry mean of the firm-level heterogeneity. So γ is the peer effect parameter to be estimated. If $\gamma > 1$, then organizations' Internet security levels are influenced by their peers. As in Graham [34] and Popadak [24], the square of the peer influence, γ^2, can be identified as follows:

$$\gamma^2 = \frac{E\left[V_j^b | S_j = 1\right] - E\left[V_j^b | S_j = 0\right]}{E\left[V_j^w | S_j = 1\right] - E\left[V_j^w | S_j = 0\right]}, \tag{6}$$

where S_j indicates industry j's size (large or small), and V_j^b and V_j^w represent the between-group variance and within-group variance for industry j, respectively. In the empirical analysis, we define $S_j = 1$ if the size of industry j is equal to or larger than the median size (26 for two-digit SIC code and 149 for two-digit NAICS code) of all industries in our data set, and $S_j = 0$ otherwise. To exclude other characteristics, the variation attributed to other organizational and industry-level average characteristics is removed.

The results from the excess-variance approach for spam volumes are listed in Table 4. Since we report the peer rankings in the treatments (emails and website) using the peer group defined by the two-digit SIC and NAICS industry codes, we define organizations sharing the same two-digit SIC and NAICS industry codes to be in the same peer group. The estimated γ^2 is about 2, which is statistically different from 1 using bootstrap, rejecting the null hypothesis that there is no peer effect. Our results support Hypothesis 5—that there are peer effects among organizations within the same industry group.

Robustness check

Our estimates are based on a large-scale RFE, which helps us exclude potential problems of omitted variables. But we conduct multiple robustness checks to provide more reliability of our estimates.

Tobit model

The distribution of the dependent variable—the outbound spam volumes of each organization—is censored at 0, since 40% of the organizations did not emit any spam in the observed time period. This may influence our estimation results in an average treatment effect in linear regressions. We include a robustness check with Tobit model, and the results are reported in Table A3 in the Appendix. We verify that the results are quite consistent with those we find in the main results section.

Placebo test

Our present experiment started at the end of January 2014. To demonstrate the robustness of our estimated results, we assume that our experiment started at the end of June in 2013 and re-estimate the treatment effects. To be specific, we still use the 6-month average spam volume before and after the assumed experimental start time as the pre- and post-experimental metrics. For the analysis that started at the end of June 2013, the post-experimental period will be from July 2013 to December 2013. We should not find any significant effect. The results are shown in Table A4 in the Appendix. We can see that, when the assumed start time is closer to the actual experiment start time, the treatment effects get larger (the magnitude of the public treatment coefficient is larger). The results support the proposition that the spam reduction is actually due to our intervention.

Alternative pre-experimental spam measure

In our experiment design and empirical analysis, we use the 6-month (from June 2013 to December 2013) average spam volume right before the start of the experiment (January 2014) as the control of the organization's original security condition. To test the robustness of our results, we re-run the regression with 2-month and 4-month average spam volumes as the pre-experimental security levels. The results are presented in Table A5 in the Appendix. We find similar treatment effects, although the magnitudes of the public treatment effects are smaller. This may be due to the fluctuation of spam volume over time.

Differences-in-differences analysis

In our treatment effect analysis in the "Empirical analysis" section, there can be unobserved characteristics that can be correlated with other main variables in our regression. To address this potential problem, we also apply a difference-in-differences approach to find the treatment effect. To test the robustness of the results, we use both the Tobit model and linear model. The results are shown in Table A6 in the Appendix. We can see that the results are consistent with those in Table 3. As compared with the organizations in the control group, large spammers tend to send out about 30% less spam after the experiment.

Alternative security measure

We have multiple spam volume variables in our data set, with which we can do further analysis to demonstrate the robustness of our main results with CBL volume data. In addition to the volume data (CBL volume), which measures the total number of emitted spam, we also have host data that count the number of IP addresses with positive spam volume (CBL host). Furthermore, we have the spam volume measure from another spam data feed: Spamikaze's PSBL (volume). The estimation results using CBL host and PSBL volume are presented in Table A7 in the Appendix. We observe, with both dependent variables, that large spammers in the public treatment group achieve a large spam reduction when compared with those in the control group.

Data without ISPs

Since ISPs usually serve residential and business customers, they generally have different security policies and capabilities than organizations that independently operate their own Internet infrastructures. For example, ISPs have less control over their customers' behavior on the Internet. Intuitively, we would expect them to be less responsive to our treatments. In our data set, there are three industry groups that are related to ISPs: telephone (group 6), unclassified communication (group 7), and other communication (group 8). We re-estimate the regressions using observations without those three industry groups and the results are listed in Table A8 in the Appendix. We can see that, as expected, the magnitude of the public treatment effects is larger.

Extending experiments

We are currently pursuing possible extensions of our present experiment. For example, we are now collaborating with local researchers at City University of Hong Kong and KAIST to collect Asian organizational data. Using several other ASN lookup services on Google, we have manually identified 2706 valid Asian organizations in China, Hong Kong, South Korea, Malaysia, and Taiwan. In addition to US industry codes, we use the Hong Kong Standard Industrial Classification (HSIC) and the Korean Standard Industrial Classification (KSIC) for Asian organizations. We use the first two digits of industry codes to group organizations, and then rank them according to their malactivity volume metrics—namely, the security metrics.

We are also implementing the treatment websites in two different cloud platforms. In Google Cloud, we provide information with three different languages: Chinese, English, and Korean. In addition, a separate Chinese website was created in Microsoft Azure, since Google service is not accessible in China. The websites are supposed to be accessed by a large number of visitors. In the long run, multiple experiments may be conducted in other countries. Cloud platforms can efficiently be scaled to serve a large number of website visitors with efficient content-caching mechanisms.

Concluding remarks

Cyber-insecurity is a serious problem that calls for efforts from both researchers and governments. The root causes of the issue can be organizations' insufficient security investment and the lack of relevant policies. We argue that the current practice with passive and reactive security information disclosure does not provide sufficient motivations for organizations to resolve the problem. Thus, we propose to set up a government sponsored, third-party institution that proactively monitors organizations' security levels and periodically publishes the evaluation reports for transparency. To evaluate the effectiveness of such an institution, a large-scale RFE on 7919 US organizations was conducted to provide spam reports to the subjects in either private or public ways. The results show that the combinations of information sharing and publicity treatment can significantly decrease large spammers' outbound spam volumes, whereas information awareness treatment by itself is not effective. The significance of peer effect indicates that one of the spam reduction motivations is—peer pressure—from close industry competitors.

We believe that the empirical results of the present paper will provide direct policy implications for governments as well as other institutes devoted to cybersecurity issues—namely, policies that reduce information asymmetry and promote peer pressure and the establishment of a security evaluation measure. More broadly, the results of our paper will benefit the members of cyber community, including various private and public organizations and individuals, by bringing their overall attention to cybersecurity issues and providing them with cybersecurity knowledge.

Cybersecurity research is a burgeoning area and there is still plenty of work to be done. The approaches and results of the present paper suggest some such future directions; our empirical work is just a starting point for Internet security policy evaluation. The experiments

described in the paper can be further extended to settings with more comprehensive security evaluation metrics and in other economic environments. In addition, empirical strategies can be developed to address the issue of the endogenous response of bot-herders. Our ongoing project pursues these directions.

Acknowledgements

This work was supported by the National Science Foundation [award number 1228990]. We thank Yun-sik Choi, Ying-Yu Chen, Mark Varga, Zeyuan Zhu, and Niyati Parameswaran from the University of Texas at Austin and Markus Iivonen from the Helsinki Metropolia University of Applied Science for technical support. We also appreciate all the helpful comments from the participants on the 14th Annual Workshop on the Economics of Information Security and Conference on Information Systems and Technology 2014. We are very grateful for the comments on the experiment web design shared by Sarah R. Benoist, Meredith Bethune from the University of Texas at Austin, and Ping Zhang from the Syracuse University, as well as the comments on statistical analysis shared by Dylan Walker from Boston University, Jason Abrevaya, and Brendan Kline from University of Texas at Austin, and Susan Athey from Stanford University. We thank Yuan Zhang from University of Texas at Arlington for the help in converting the manuscript into Word format. We are responsible for all the possible problems in the paper.

Appendix 1: Randomization details

To get reliable treatment effect estimation from a RFE, we conducted a stratified, pair-wise matching randomization on 7919 organizations [20]. Due to heterogeneity of legal regimes and economic environment across countries, we only included US organizations in the present experiment.

Stratification

One of the standard approaches to avoiding imbalance is stratification on a few key characteristics [35]. In stratification, organizations will be randomly assigned to different treatment groups within each subgroup, defined by key characteristics. In our experiment, we defined 195 subgroups by SIC codes (39 industry sectors) and number of IP addresses (5 segments). The detailed industry and number of IP addresses groups are listed in Tables A1 and A2. Despite the correlation between industry activities, we managed to divide firms into mostly equal sized groups in order to get precise estimation.

The rationale of choosing the two characteristics is as follows. First, organizations in different industries have different priorities on security. For example, security should be particularly important for software companies. Spammers may also have different incentives based on the "value" of the data that different companies maintain. In that sense, financial and health sectors may have a higher risk. Second, organization size may affect the approaches on the system protection. Organizations with a larger number of IP addresses, generally with larger size, may face more risks and potential problems. On the other hand, large organizations usually have an independent IT department with security experts. With more resources, large organizations can afford better anti-virus software or firewalls to prevent potential security attacks. Therefore, we divided the whole sample into five groups according to their IP address counts.

Pair-wise matching

Stratification can only control for the balance of industry sectors and IP counts and the two variables cannot explain a large share of the spam volume's variance. Since the baseline spam volumes can be

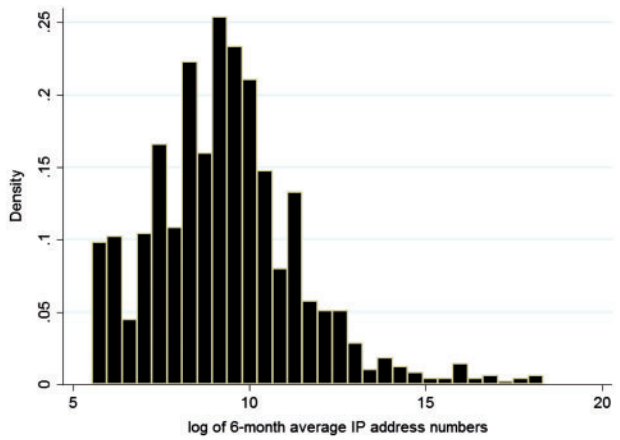

Figure A1. Distribution of communication companies' sizes in our sample (two-digit SIC code equals to 48).

Table A1. Industrial groups' description

Group	Industry sector	Number of organizations
1	Agriculture, mining, and construction	123
2	Electronic devices	103
3	Publishing	133
4	Chemical and measuring manufacturing	156
5	Other manufacturing	245
6	Telephone	836
7	Unclassified communication	164
8	Other communication	163
9	Transportation	253
10	Durable wholesale	215
11	Non-durable wholesale	126
12	Furniture retail	111
13	Non-classified retail	145
14	Other retail	158
15	Depository institutions	186
16	Credit and real estate	133
17	Security	255
18	Insurance	199
19	Holdings and other financial companies	179
20	Health services	337
21	Colleges	423
22	Education service other than colleges	214
23	Management consulting	181
24	Business consulting	150
25	Other management service	116
26	Engineer, accounting, and research	194
27	Non-classified business service	484
28	Computer programming	249
29	Prepackaged software	140
30	Computer integrated systems	157
31	Computer processing	162
32	Information retrieval	102
33	Non-classified computer service	167
34	Other business service	222
35	Legal service	108
36	Membership organization	93
37	Miscellaneous service	115
38	Other service	223
39	Public administration	199

Table A2. Groups based on the number of IP addresses

Number of IP addresses	0–427	428–1024	1024–10^4	10^4–10^5	$>10^5$
Group	1	2	3	4	5

the best proxy of the organizations' security condition, we did the pair-wise matching on organizations' pre-experimental spam volumes. In practice, we found three organizations that minimize the sum of three pairwise differences among them. One problem we faced during this process was the distribution of spam volumes. We found that the distribution for a given organization varies greatly over time and both the distributions of spam volumes and the number of IP addresses for the whole sample was highly skewed. Thus, we used the natural logarithm transformed 6-month average spam volumes as our pre-experimental spam volumes to get higher probability of detecting the treatment effects.

Re-randomization

After the random assignment with stratification and pair-wise matching, we checked the distances between the control group and two treated groups with respect to companies' various characteristics. We followed the procedures in Morgan *et al.* (2012) to set the pre-specified criteria. With 10,000 simple random draws from our sample followed the previous two steps, we created a simulated distribution of distance between any two groups and set the 5% quantile as the criteria for randomization. Finally, with the 10,000 randomization assignments satisfying the re-randomization criteria for power calculation, we randomly chose one of them as our executed one.

Appendix 2: Additional figures and tables

Table A3. Treatment effects estimation

	Avg. treatment effects		Heterogeneous treatment effects			
	Overall		Public versus control		Private versus control	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Private treatment	−0.0243	−0.0114			0.258***	0.139
	(0.1480)	(0.0655)			(0.0998)	(0.1070)
Public treatment	−0.074	−0.0675	−0.332***	−0.244***		
	(0.1480)	(0.0706)	(0.0923)	(0.0698)		
Indicator			−6.811***	−0.342	−8.276***	0.028
			(0.6290)	(0.3480)	(0.5310)	(0.1420)
Private treatment × indicator					−0.263*	
					(0.1340)	
Public treatment × indicator			0.333**	0.256**		−0.174
			(0.1340)	(0.1050)		(0.1110)
Pre-experimental spam volume		0.702***		0.671***		0.711***
		(0.0269)		(0.0605)		(0.0372)
Number of IP addresses		1.693***		1.618***		1.726***
		(0.1610)		(0.1840)		(0.1410)
(Number of IP addresses)^2		−0.0800***		−0.0756***		−0.0823***
		(0.0076)		(0.00931)		(0.0065)
Number of botnets		0.280***		0.278***		0.276***
		(0.0254)		(0.0229)		(0.0295)
Stock		0.328***		0.23		0.474***
		(0.1060)		(0.1490)		(0.1020)
Intercept	0.886***	−9.099***	6.226***	−9.037***	8.696***	−8.695***
	(0.1080)	(0.7590)	(0.5230)	(0.7500)	(0.2770)	(0.6460)
Industry	No	Two-digit SIC	No	Two-digit SIC	No	Two-digit SIC
P-value for H_0: $a_1 = a_2$	0.7363	0.3220				
Observations	7919	7919	5280	5280	5280	5280
R-squared	0	0.744	0.434	0.751	0.501	0.745

Notes: This table displays the estimated private and public treatment effects with Tobit model. Columns 1 and 2 report the estimates of the differences between the spam volume of treatment groups and control controlling for pre-experimental spam volume, number of pre-experimental IP addresses, number of pre-experimental IP addresses squared, number of pre-experimental infesting botnets, whether or not publicly traded, and industry fixed effects. Columns 3 and 4 report the estimates of the heterogeneous treatment effects with organizations in control and public treatment group. The indicator equals to 1 if organization's log pre-experimental spam volume is less than 3.6. Columns 5 and 6 report the estimates of the heterogeneous treatment effects with organizations in control and private treatment group. The indicator equals to 1 if organization's log pre-experimental botnet is less than 3.4. Standard errors are clustered by industry codes and shown in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A4. Placebo test

	Post-experimental spam volume					
	June 2013		July 2013		August 2013	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Private treatment	0.0372		−0.00670		−0.00487	
	(0.0455)		(0.0428)		(0.0433)	
Public treatment	−0.00349	0.06500	−0.04360	−0.06450	−0.05090	−0.12600
	(0.0465)	(0.0600)	(0.0431)	(0.0702)	(0.0433)	(0.0988)
Indicator		−3.545***		−3.093***		−2.830***
		(0.1420)		(0.1170)		(0.1640)
Public treatment × indicator		−0.114***		0.0251		0.1000
		(0.0574)		(0.0635)		(0.1060)
Pre-experimental spam volume	0.588***	0.275***	0.561***	0.299***	0.559***	0.296***
	(0.0151)	(0.0364)	(0.0149)	(0.0361)	(0.0151)	(0.0257)
Number of IP addresses	0.259***	0.134**	0.437***	0.251***	0.418***	0.289***
	(0.0413)	(0.0546)	(0.0396)	(0.0536)	(0.0372)	(0.0557)
(Number of IP addresses)^2	−0.00595**	−0.00353	−0.0182***	−0.0107***	−0.0165***	−0.0116***
	(0.0026)	(0.0036)	(0.0025)	(0.0035)	(0.0023)	(0.0034)
Number of botnets	0.187***	0.218***	0.279***	0.274***	0.258***	0.277***
	(0.0176)	(0.0274)	(0.0194)	(0.0343)	(0.0188)	(0.0289)
Stock	−0.111	0.0767	−0.0548	0.0955	−0.0105	0.0984
	(0.0677)	(0.0925)	(0.0659)	(0.0714)	(0.0681)	(0.0785)
Intercept	−1.002***	3.116***	−1.746***	2.199***	−1.688***	1.734***
	(0.2550)	(0.3080)	(0.1990)	(0.2920)	(0.1930)	(0.3410)
Industry	Two-digit SIC	Two-digit SIC	Two-digit SIC	Two-digit SIC	Two-digit SIC	Two-digit SIC
Observations	7919	5280	7919	5280	7919	5280
R-squared	0.740	0.847	0.768	0.846	0.760	0.821

Notes: This table displays the robustness check with placebo test. Columns 1–2, 3–4, and 5–6 use October, November, and December 2013 as our experiment start time, respectively. Standard errors are clustered by industry codes and shown in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A5. Treatment effects with alternative measures of pre-experimental spam volume

	Post-experimental spam volume							
	Two-month average pre-experimental				Four-month average pre-experimental			
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Private treatment	−0.0196				−0.0176			
	(0.0357)				(0.0317)			
Public treatment	−0.07310	−0.237**	−0.226**	−0.233**	−0.06110	−0.208*	−0.196***	−0.205**
	(0.0442)	(0.1120)	(0.0895)	(0.0925)	(0.0407)	(0.1120)	(0.0732)	(0.0836)
Indicator		−0.983***	−0.965***	−0.967***		−0.735***	−0.728***	−0.737***
		(0.1230)	(0.1060)	(0.1330)		(0.1260)	(0.0936)	(0.1210)
Public treatment × indicator		0.220*	0.203**	0.211*		0.194	0.177**	0.187*
		(0.1200)	(0.0976)	(0.1090)		(0.1200)	(0.0852)	(0.1050)
Pre-experimental spam volume	0.585***	0.497***	0.479***	0.477***	0.597***	0.536***	0.517***	0.515***
	(0.0435)	(0.0234)	(0.0386)	(0.0332)	(0.0401)	(0.0236)	(0.0323)	(0.0238)
Number of IP addresses	0.436***	0.414***	0.393***	0.394***	0.389***	0.387***	0.364***	0.366***
	(0.0567)	(0.0508)	(0.0496)	(0.0436)	(0.0522)	(0.0498)	(0.0511)	(0.0434)
(Number of IP addresses)^2	−0.0168***	−0.0149***	−0.0144***	−0.0147***	−0.0145***	−0.0139***	−0.0131***	−0.0134***
	(0.0039)	(0.0031)	(0.0033)	(0.0026)	(0.0037)	(0.0030)	(0.0033)	(0.0025)
Number of botnets	0.306***	0.304***	0.308***	0.311***	0.288***	0.289***	0.291***	0.293***
	(0.0592)	(0.0268)	(0.0572)	(0.0430)	(0.0544)	(0.0261)	(0.0504)	(0.0374)
Stock	0.0232	0.0224	0.0235	0.0446	0.0314	0.0218	0.015	0.0332
	(0.0732)	(0.0861)	(0.1010)	(0.0841)	(0.0650)	(0.0840)	(0.0961)	(0.0777)
Intercept	−1.758***	−0.445*	−0.726***	−0.716***	−1.679***	−0.650***	−0.862***	−0.852***
	(0.1990)	(0.2360)	(0.2350)	(0.2510)	(0.1830)	(0.2360)	(0.2040)	(0.2310)
Industry	No	No	Two-digit SIC	Three-digit NAICS	No	No	Two-digit SIC	Three-digit NAICS
Observations	7919	5280	5280	5280	7919	5280	5280	5280
R-squared	0.741	0.740	0.746	0.748	0.747	0.744	0.750	0.751

Notes: This table displays the robustness check with alternative measures of pre-experimental spam volume. Columns 1–4 use monthly average spam volume from November 2013 to December 2013 while Columns 5–8 use monthly average spam volume from September 2013 to December 2013. The results in Columns 2, 3, 4, 6, 7, 8 are from the subsample of organizations that are in control and public treatment groups. The indicator is a dummy variable that indicates whether the company's log pre-experimental spam volume is less than 3.6. Standard errors are clustered by industry codes and shown in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A6. Treatment effects with difference-in-differences model

	Tobit model				OLS model			
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Public treatment	−0.0118	0.00786	0.00895	0.01000	−0.0118	0.00898	0.0107	0.00961
	(0.0756)	(0.0553)	(0.0397)	(0.0562)	(0.0756)	(0.0577)	(0.0385)	(0.0582)
After	−0.735***	−0.652***	−0.653***	−0.653***	−0.776***	−0.686***	−0.687***	−0.687***
	(0.1620)	(0.1100)	(0.1600)	(0.1500)	(0.1650)	(0.1150)	(0.1690)	(0.1590)
Indicator	−6.181***	−3.785***	−3.714***	−3.696***	−7.668***	−4.535***	−4.421***	−4.399***
	(0.1100)	(0.1020)	(0.2260)	(0.1970)	(0.1340)	(0.1240)	(0.3430)	(0.3020)
Public treatment × after	−0.287	−0.263*	−0.264***	−0.264*	−0.305	−0.277*	−0.279***	−0.279**
	(0.2310)	(0.1570)	(0.0660)	(0.1350)	(0.2350)	(0.1640)	(0.0697)	(0.1410)
Public treatment × indicator	1.027***	0.901***	0.902***	0.901***	1.087***	0.958***	0.953***	0.954***
	(0.1680)	(0.1160)	(0.1520)	(0.1440)	(0.1960)	(0.1400)	(0.1590)	(0.1520)
Indicator × after	−0.0153	−0.0543	−0.0581	−0.0613	−0.08	−0.104	−0.12	−0.121
	(0.1550)	(0.1140)	(0.0811)	(0.1130)	(0.1810)	(0.1370)	(0.0875)	(0.1210)
Public treatment × indicator × after	0.324	0.310*	0.310***	0.310**	0.409	0.379*	0.382***	0.380**
	(0.2390)	(0.1660)	(0.0672)	(0.1440)	(0.2800)	(0.2000)	(0.0910)	(0.1620)
Number of IP addresses		0.146***	0.132***	0.129***		0.288***	0.255***	0.246***
		(0.0081)	(0.0213)	(0.0172)		(0.0133)	(0.0380)	(0.0333)
Number of botnets		0.569***	0.563***	0.565***		0.571***	0.568***	0.572***
		(0.0240)	(0.0727)	(0.0530)		(0.0275)	(0.0976)	(0.0716)
Stock		−0.0386	−0.0373	−0.0219		0.0899	0.0847	0.109
		(0.0546)	(0.1190)	(0.0975)		(0.0889)	(0.1850)	(0.1480)
Intercept	−1.758***	−0.445*	−0.726***	−0.716***	−1.679***	−0.650***	−0.862***	−0.852***
	(0.1990)	(0.2360)	(0.2350)	(0.2510)	(0.1830)	(0.2360)	(0.2040)	(0.2310)
Industry	No	No	Two-digit SIC	Three-digit NAICS	No	No	Two-digit SIC	Three-digit NAICS
Observations	10,560	10,560	10,560	10,560	10,560	10,560	10,560	10,560

Notes: This table displays the robustness check with DID model on subsample of organizations in public and control groups. Columns 1–4 report the estimates from Tobit models and Columns 5–8 report the estimates from OLS models. The "After" is a dummy variable indicates whether the outbound spam volume is collected after our experiment has started. The "Indicator" is a dummy variable represents whether the organization's log pre-experimental spam volume is less than 3.6. Standard errors are clustered by industry codes and shown in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A7. Treatment effects with different security measures

	CBL host				PSBL volume			
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Private treatment	−0.0198	−0.0136			−0.0126	−0.01680		
	(0.0571)	(0.0262)			(0.0431)	(0.0222)		
Public treatment	−0.02650	−0.03080	−0.12300	−0.0979*	−0.05090	−0.0519**	−0.218*	−0.180**
	(0.0570)	(0.0264)	(0.0984)	(0.0517)	(0.0422)	(0.0223)	(0.1270)	(0.0889)
Indicator			−3.375***	−0.330***			−1.945***	−0.318***
			(0.0743)	(0.0630)			(0.0917)	(0.0748)
Public treatment × indicator			0.1100	0.0888			0.217*	0.175*
			(0.1070)	(0.0604)			(0.1280)	(0.0886)
Pre-experimental spam metric		0.724***		0.662***		0.493***		0.467***
		(0.0162)		(0.0242)		(0.0255)		(0.0320)
Number of IP addresses		0.160***		0.150***		0.145***		0.121***
		(0.0187)		(0.0229)		(0.0294)		(0.0312)
(Number of IP addresses)^2		−0.00453***		−0.00356**		−0.00809***		−0.00723***
		(0.0012)		(0.0014)		(0.0018)		(0.0020)
Number of botnets		0.0842***		0.0862***		0.254***		0.244***
		(0.0101)		(0.0120)		(0.0151)		(0.0176)
Stock		0.0592		0.0387		−0.0426		−0.0484
		(0.0432)		(0.0528)		(0.0333)		(0.0383)
Intercept	1.845***	−0.650***	4.322***	−0.273	0.597***	−0.493***	2.024***	−0.168
	(0.0404)	(0.1190)	(0.0683)	(0.1830)	(0.0304)	(0.1230)	(0.0912)	(0.1430)
Industry	No	No	Two-digit SIC	Three-digit NAICS	No	No	Two-digit SIC	Three-dig NAICS
Observations	7919	7919	5280	5280	7919	7919	5280	5280
R-squared	0.000	0.792	0.505	0.792	0.000	0.685	0.284	0.689

Notes: This table displays the robustness check with different security measures. Columns 1–4 report the treatment effects for number of infesting hosts by CBL in each quantile. Columns 5–8 report the treatment effects for spam volume by PSBL in each quantile. The results in columns 3, 4, 7, and 8 are from the subsample of organizations that are in control and public treatment groups. The indicator is a dummy variable that indicates whether the company's log pre-experimental spam volume is less than 3.6. Standard errors are clustered by industry codes and shown in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A8. Treatment effects without ISPs

	Average treatment effects				Heterogeneous treatment effects			
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Private treatment	0.0172	−0.00679	−0.00222	−0.0111				
	(0.0802)	(0.0462)	(0.0385)	(0.0428)				
Public treatment	−0.03430	−0.05830	−0.05720	−0.06210	−0.31300	−0.306**	−0.293***	−0.313**
	(0.0795)	(0.0465)	(0.0472)	(0.0498)	(0.2030)	(0.1440)	(0.0984)	(0.1210)
Indicator					−4.429***	−0.274*	−0.283	−0.287*
					(0.1480)	(0.1650)	(0.1730)	(0.1600)
Public treatment × indicator					0.311	0.312**	0.293***	0.310**
					(0.2100)	(0.1510)	(0.1080)	(0.1400)
Pre-experimental spam volume		0.521***	0.506***	0.506***		0.504***	0.486***	0.487***
		(0.0240)	(0.0251)	(0.0206)		(0.0372)	(0.0239)	(0.0297)
Number of IP addresses		0.331***	0.323***	0.315***		0.323***	0.311***	0.307***
		(0.0437)	(0.0404)	(0.0385)		(0.0537)	(0.0440)	(0.0512)
(Number of IP addresses)^2		−0.00985***	−0.0105***	−0.0102***		−0.00925***	−0.00958***	−0.00963***
		(0.0027)	(0.0029)	(0.0023)		(0.0032)	(0.0029)	(0.0030)
Number of botnets		0.361***	0.377***	0.375***		0.353***	0.369***	0.367***
		(0.0440)	(0.0538)	(0.0307)		(0.0565)	(0.0425)	(0.0322)
Stock		0.0872	0.125**	0.134**		0.0828	0.0867	0.108
		(0.0679)	(0.0491)	(0.0571)		(0.0853)	(0.0749)	(0.0691)
Intercept	1.999***	−1.185***	−1.364***	−1.310***	5.531***	−0.904***	−1.065***	−1.012***
	(0.0566)	(0.1740)	(0.1430)	(0.1510)	(0.1430)	(0.2560)	(0.2680)	(0.2690)
Industry	No	No	Two-digit SIC	Three-digit NAICS	No	No	Two-digit SIC	Three-digit NAICS
Observations	6755	6755	6755	6755	4506	4506	4506	4506
R-squared	0.000	0.663	0.673	0.675	0.420	0.660	0.671	0.674

Notes: This table displays the robustness check without ISPs' observations. Columns 1–4 report the average treatment effects and Columns 5–8 report the heterogeneous treatment effects. The results in Columns 5–8 are from the subsample of organizations that are in control and public treatment groups. The indicator is a dummy variable indicates whether the company's log pre-experimental spam volume is less than 3.6. Standard errors are clustered by industry codes and shown in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

References

1. Krebs B. *Spam Nation: The Inside Story of Organized Cybercrime—From Global Epidemic to your Front Door*. Sourcebooks, Inc., 2014.
2. Anderson R. Why information security is hard? An economic perspective. In: *Proceedings of the 17th Annual Computer Security Applications Conference, December 2001*, pp. 358–365. IEEE 2001.
3. Bauer JM, Van Eeten MJ. Cybersecurity: stakeholder incentives, externalities, and policy options. *Telecommun Policy* 2009;**33**:706–19.
4. Van Eeten M, Asghari H, Bauer J *et al*. Internet service providers and botnet mitigation: a fact-finding study on the Dutch market. Delft University of Technology, 2011.
5. Wood D, Rowe B. Assessing home Internet users' demand for security: Will they pay ISPs? In: *Proceedings of the Workshop on the Economics of Information Security*, 2011.
6. Moore TW, Clayton R. The impact of public information on phishing attack and defense. *Commun Strat* 2011;**81**:45–68.
7. Tang Q, Linden L, Quarterman JS *et al*. Improving internet security through social information and social comparison: a field quasi-experiment. In: *Proceedings of the Workshop on the Economics of Information Security*, 2013.
8. D'arcy J, Hovav A, Galletta D. User awareness of security countermeasures and its impact on information systems misuse: a deterrence approach. *Inform Syst Res* 2009;**20**:79–98.
9. Gal-Or E, Ghose A. The economic incentives for sharing security information. *Inform Syst Res* 2005;**16**:186–208.
10. Duflo E, Glennerster R, Kremer M. Using randomization in development economics research: a toolkit. *Handbook Dev Econ* 2007;**4**:3895–962.
11. Harrison GW, List JA. Field experiments. *J Econ Lit* 2004;**42**:1009–55.
12. List JA. Why economists should conduct field experiments and 14 tips for pulling one off. *J Econ Perspect* 2011;**25**:3–15.
13. Arora A, Krishnan R, Nandkumar A *et al*. Impact of vulnerability disclosure and patch availability—an empirical analysis. In: *Proceedings of the Third Workshop on the Economics of Information Security, vol.* **24**, pp. 1268–87.
14. Durumeric Z, Kasten J, Adrian D *et al*. The matter of heartbleed. In: *Proceedings of the 2014 Conference on Internet Measurement Conference*, pp. 475–488. ACM, 2014.
15. Rossow C. Amplification hell: revisiting network protocols for DDoS abuse. In: *Proceedings of the Symposium on Network and Distributed System Security (NDSS)*, Internet Society, 2014.
16. Stone-Gross B, Kruegel C, Almeroth K *et al*. Fire: finding rogue networks. In: *Proceedings of the Computer Security Applications Conference*, pp. 231–240. IEEE, 2009.
17. Vasek M, Moore T. Do malware reports expedite cleanup? An experimental study. In: *Proceedings of the Workshop on Cyber Security Experimentation and Test*, USENIX, 2012.
18. Athey S, Imbens G. Machine learning methods for estimating heterogeneous causal effects. arXiv Preprint arXiv:1504.01132, 2015.
19. Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. arXiv Preprint arXiv:1510.04342, 2015.
20. Morgan KL, Rubin DB. Rerandomization to improve covariate balance in experiments. *Ann Stat* 2012;**40**:1263–82.
21. Rao JM, Reiley DH. The economics of spam. *J Econ Perspect* 2012;**26**:87–110.
22. Pitsillidis A, Kanich C, Voelker GM *et al*. Taster's choice: a comparative analysis of spam feeds. In: *Proceedings of the 2012 ACM Conference on Internet Measurement Conference*, pp. 427–440. ACM, 2012.
23. Dharmapurikar S, Krishnamurthy P, Taylor DE. Longest prefix matching using bloom filters. In: *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, pp. 201–212. ACM, 2003.
24. Popadak JA. Dividend payments as a response to peer influence. Available at SSRN 2170561, 2012.
25. Festinger L. A theory of social comparison processes. *Hum Relat* 1954;**7**:117–40.
26. Sacerdote B. Peer effects with random assignment: results for Dartmouth roommates. Technical Report, National Bureau of Economic Research, 2000.
27. Aral S, Walker D. Identifying influential and susceptible members of social networks. *Science* 2012;**337**:337–41.
28. Brown JR, Ivković Z, Smith PA *et al*. Neighbors matter: causal community effects and stock market participation. *J Finan* 2008;**63**:1509–31.
29. Shue K. Executive networks and firm policies: evidence from the random assignment of MBA peers. *Rev Finan Stud* 2013;**26**:1401–42.
30. Bruhn M, McKenzie D. In pursuit of balance: randomization in practice in development field experiments. World Bank Policy Research Working Paper Series, 2008.
31. Wager S, Hastie T, Efron B. Confidence intervals for random forests: the jackknife and the infinitesimal jackknife. *J Mach Learn Res* 2014;**15**:1625–51.
32. Breiman L. Random forests. *Mach Learn* 2001;**45**:5–32.
33. Manski CF. Identification of endogenous social effects: the reflection problem. *Rev Econ Stud* 1993;**60**:531–42.
34. Graham BS. Identifying social interactions through conditional variance restrictions. *Econometrica* 2008;**76**:643–60.
35. Fisher RA. The Logic of Inductive Inference. *J R Stat Soc* 1935;**98**:39–82.