1-2015

Improving Internet Security through Mandatory Information Disclosure

Qian TANG
Singapore Management University, QIANTANG@smu.edu.sg

Andrew B. WHINSTON
University of Texas at Austin

Citation

TANG, Qian and WHINSTON, Andrew B.. Improving Internet Security through Mandatory Information Disclosure. (2015). *2015 48th Hawaii International Conference on System Sciences HICSS: 5-8 January, Kauai, HI: Proceedings*. 4813-4823.

Available at: https://ink.library.smu.edu.sg/sis_research/2637

2015 48th Hawaii International Conference on System Sciences

Improving Internet Security Through Mandatory Information Disclosure

Qian Tang
Singapore Management University
qiantang@smu.edu.sg

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

Abstract

Although disclosure has long been considered as a solution to internalize externalities, mandatory security information disclosure is still in debate. We propose a mandatory disclosure mechanism based on existing data. The information is disclosed as straightforward rankings of organizations for users to understand, interpret, and make comparisons. As a result, the disclosure can influence organizations through reputational effects. We created a public website to disclose information regularly and conducted a quasi-experiment on outgoing spam to test the effectiveness of our mechanism on four matched country groups. For each treated country, we released the ranking list of top 10 most spamming organizations every month, while for the control countries, no information was disclosed. We find that the treatment organizations subject to spam information disclosure reduced significantly more spam than comparison organizations.

1. Introduction

Cyber crime is one of the fastest growing areas of crime. A 2012 cost of cyber crime study conducted by Ponemon Institute shows that the attack frequency has more than doubled over a three-year period and the costs has increased by nearly 40 percent [1]. Although the awareness is growing, the current effort of organizations for cyber security is far from enough, which increases the risks faced by other Internet users [2]. The underinvestment in Internet security is caused by three main reasons. First of all, Internet security is often considered too expensive to achieve. Security products and services are sometimes regarded as useful and desirable, yet not affordable. High-level security practices can be reinforced to prevent security disasters and control the damage. The deployment of such practices, however, is a costly endeavor for organizations without assured significant benefit. Second, the absence of legislative enforcement for disclosure leads to the lack of transparency. Some companies simply do not have the knowledge or internal policies to recognize or deal with cyber threats. Some knowingly choose to cover it up by not reporting to avoid reputational loss. Moreover, because of negative externalities, even if underinvestment is an optimal choice for the company, it is not a social optimum. One system's vulnerabilities may not necessarily harm that system but are often used against others. For example, because malware writers often direct attacks at other targets, they would purposefully minimize the impacts on the infected host.

For lack of transparency and existence of negative externalities, information disclosure could be a good solution and have been encouraged by policy makers. Disclosure makes information transparent. Partners, customers and investors can use the disclosed security information for comparison and evaluation, and make more informative decisions. This would affect companies' financial performance and add onto their incentives for investing in information security. In that sense, disclosure internalizes the negative externalities of insecurity as reputational and financial loss.

Security information disclosure laws have been focused on individual notification. As of August 20, 2012, 46 U.S. states and the District of Columbia, Guam, Puerto Rico and the Virgin Islands have enacted legislation requiring notification of security breaches involving personal information. Regulations on publically disclosing security information so far have mainly been guidelines and suggestions rather than legislation. Therefore, security information disclosure relies mostly on voluntary announcement instead of mandatory reporting. As with voluntary disclosure, the information is less likely to be accurate, comparable, and up to date, making the information not very useful. For example, new healthcare law that became effective early 2012 requires drug companies to disclose the payments they make to doctors for research, consulting, speaking, travel, and entertainment. Major drug companies like Pfizer did disclose the detailed payment information on their websites. Yet the data is called hard to parse, with information scattered here and there. It is difficult for patients to understand the information and compare doctors with each other. What it achieved is only translucency rather than transparency.

1530-1605/15 $31.00 © 2015 IEEE
DOI 10.1109/HICSS.2015.572

What we propose in this paper is using mandatory information disclosure and straightforward information presentation for improving Internet security. Mandatory disclosure can steer management away from any temptation to suppress unfavorable information. It is also more likely to produce standardized data across different companies. With increased standardization, greater comparability also arises, reducing efforts to read and interpret the disclosed information. The goal of public information disclosure is to assist decision making, which is only possible when the reader fully understand the information. We make information straightforward for the reader through explicitly compare companies together in the form of ranking. In this way, one company's security situation is evaluated against other companies and presented in a relative way.

2. Literature review

2.1. Economics of information security

It has long been recognized that Internet security is not a problem that technology alone can solve [3]. Many security questions are at least as much economic as technical. Fundamentally, Internet insecurity is the result of perverse incentives, which are distorted by network externalities, asymmetric information, moral hazard, adverse selection, liability dumping, and the so-called tragedy of the commons [4]. Systems fail often because of misplaced economic incentives: The people who could protect a system are not the ones who suffer the costs of failure [5]. Security failure is caused as much by bad incentives as by bad design [6]. Meanwhile, hacking has evolved over the past a few years to become a well-organized, sophisticated underground market.

The economic incentive problem is caused by negative externality of insecurity. Externality happens because social costs or benefits are not equal to private costs or benefits [7,8,9]. Negative externality happens when social costs are greater than private costs, whereas positive externality happens when social benefits are greater than private benefits. Security vulnerabilities of a system are often exploited by hackers to attack other systems. For example, spam has such an extreme negative externality that the social costs are about 100 times the private benefits [10,11,12]. More and more studies have recognized the importance of security externalities and have come up with various economic and legal policy proposals. The standard economic treatment for negative externality is to impose a Pigouvian tax on the activity that generates negative externality [7,8,9]. For spam, researchers in many studies have proposed to have the spam sender pay the receiver for attention or levy penalties on consumers who purchase goods from spammers [13,14]. However, these proposals raise the concerns for privacy and account hijacking by hackers. The legal treatment is to let government make law or regulation enforcements. For spam, the legal interventions include requiring legal advertisers to offer opt-in or opt-out choices for email receivers and putting legal pressure on banks that process payments from foreign banks known to act on behalf of spam merchants [15]. However, most cyber crime originates from foreign countries and is beyond the reach of legislation.

2.2. Corporate information disclosure

Corporate information disclosure is a common mechanism used to monitor corporate activities, build trust, dispel erroneous beliefs of the public, and impose pressure for change either on companies themselves or on the government to make legislative change. A wide variety of methods of disclosure and forms of presentation are employed depending on the subjects upon which information is given [16,17]. Many mandatory disclosure provisions concern financial information and information that is relevant to monitoring the management's stewardship. Voluntary disclosures may be more diverse covering issues such as environmental impact, community relations with the company, employee promotions [18].

The benefits of mandatory disclosure include fraud prevention, investor protection, corporate governance and accountability of manager to the shareholders, corporate democracy, efficiency through reduction of monitoring and information search costs, reduction of competitive injury, standardization of information making comparison easier, alternative to regulatory intervention and political and social benefits arising from disclosure [19,20]. Meanwhile, mandatory disclosure may come with the problems such as complexity, overload, cost of providing and interpreting information, potential threat to confidentiality, lack of relevance, lack of interest on the part of the shareholders, misleading and incomplete information.

In addition to or in absence of mandatory disclosure, companies may make voluntary information disclosure beyond their legal and regulatory disclosure obligations. Voluntary disclosure can save a lot of reporting costs [21]. However, reliance on voluntary disclosure would allow companies to avoid disclosing negative information. The disclosed information is difficult to verify even if it is misleading. Different information presentation forms would make comparison between companies difficult or even impossible.

2.3. Security information disclosure

Security vulnerability disclosure is an area of public policy that has been subject to considerable debate. Studies on software vulnerability disclosure have shown that although disclosing vulnerability information provides an impetus to the vendor to release patches early, instant disclosure leaves users defenseless against attackers who can exploit the disclosed vulnerability [22]. Arora et al. [3] found that although vendors can quickly respond to instant disclosure, vulnerability disclosure also increases the frequency of attacks. Arora et al. [23] suggested that the optimal vulnerability disclosure depends on underlying factors such as how quickly vendors respond to disclosure by releasing patches and how likely attackers are to find and exploit undisclosed or unpatched vulnerabilities.

Because of the complexity, it is difficult to implement mandatory security information disclosure on all companies. Yet industry-based Information Sharing and Analysis Centers (ISACs), where security breach information is voluntarily revealed to information-sharing alliance, has been established to facilitate the sharing of security information to enhance and protect critical cyber infrastructure. Gal-Or and Ghose [24] studied the economic incentives for security information sharing and found that information sharing yields greater benefits in more competitive industries. Gordon et al. [25] examined how information sharing affects the overall level of information security when firms face the trade-off between improved information security and the potential for free riding.

Mandatory security information disclosure has focused on data breach notification. As of August 20, 2012, 46 U.S. states and the District of Columbia, Guam, Puerto Rico and the Virgin Islands have enacted legislation requiring notification of security breaches involving personal information [26]. The concern has arisen that notifications may simply shift the burden to consumers if breaches really cause harm [27]. Its effectiveness is also non-conclusive. While Romanosky et al. [28] and Campbell et al. [29] provide evidence that disclosure has positive impacts on reducing cyber crime and would be incorporated as market information, Kannan et al. [30] found that the information does not have significant market value in the long run.

3. Field quasi-experiment

Field experimentation has been used extensively for policy evaluation [31] in information security and privacy [32]. It has the advantage of observing the participants' reactions in a naturally occurring setting over a laboratory experiment [33]. A field quasi-experiment preserves the same benefit but differs from a field experiment in that the treatment assignment is not randomized but rather a purposeful choice by the researcher. In this study, we aim to evaluate whether security information disclosure can lead to security improvement. We would not be able to achieve this unless we generate enough public awareness of the disclosed information. Therefore, it is in our best interest to select subjects more likely to be informed of the disclosed information into the treatment group. As a result, we identified four North American and European countries to be treated with security information disclosure, to make sure that the reach of the disclosure can be achieved in a cost-efficient way. This design also enables us to focus on the four specific treated countries. However, the generalizability of the results to other countries needs to be further tested. We conduct this study following the same experimental setting as described in our earlier work [35] with an improved quasi-experimental design, enriched data sets from multiple sources, extended pre- and post- experimental periods, and additional analyses, to increase the validity of the results and provide more implications.

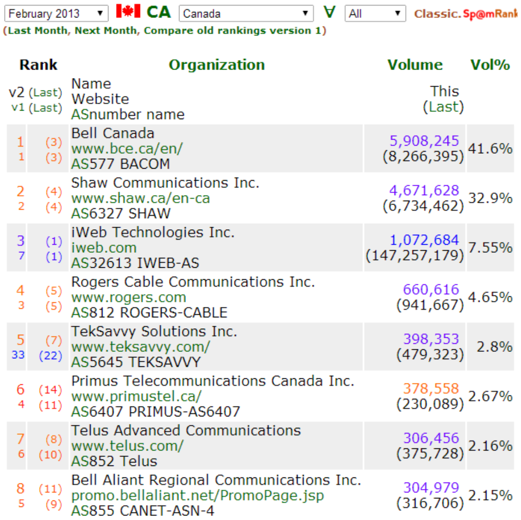
3.1. Experimental setting

We look at outgoing spam as the specific security issue faced by companies. We chose outgoing spam for several reasons. First, the primary data on outgoing spam is already being collected by various anti-spam blocklists. So we can experiment without self-reporting from companies. Second, focusing on specific security issue makes it easier to quantify the problem, standardize the data, and make comparisons. Third, spam is a severe worldwide security issue since an estimated 88% of daily worldwide email traffic is spam [34]. Spam is often sent out through compromised computer accounts or botnets, which are networks of "zombie" computers, and thus is a symptom of more damaging security problems. The same vulnerabilities that enable spam are also openings for other exploits. Lastly, spam demonstrates extreme negative externalities that the ratio of external costs to private benefits is as high as 100:1, compared to 1:10 for pollution and 7:30 for nonviolent property crime [11,12]. Therefore, if an account is constantly sending out spam, it not only risks being attacked itself, but also increases the risks faced by other Internet users. In other words, the efforts of reducing outgoing spam can produce a remarkably large positive externality on other users. For instance, in 2011, Microsoft, Pfizer, FireEye network security, and security experts at the University of Washington collaborated to take down Rustock, one of the largest botnets. The takedown of this single botnet was followed by an immediate one-third reduction in global email spam [12]. Although we look at outgoing spam specifically in this paper, the same disclosure mechanism applies to other security issues.

To publicly disclose outgoing spam information, we launched a website (SpamRankings.net) in May 2011 and have since used it to release outbound spam information of organizations, as described in [35]. The website presents monthly outbound spam volumes and rankings for organizations in the treated countries, including the United States, Canada, Belgium, and Turkey. The disclosed information is processed from the raw data we received from two anti-spam blocklists, the Composite Blocking List (CBL), and the Passive Spam Block List (PSBL). Blocklists detect and list IPs that send out spam emails, and are usually used by email service providers to filter incoming emails. We received the IP level data from blocklists every day in text files from CBL and in Network News Transfer Protocol (NNTP) messages from PSBL. The daily file consists of on average eight million lines like this:

"1.0.17.248, AS2519, 1.0.16.0/23, JP, vectant.ne.jp, , , 1349617960, spamsalot, 39".

Each line contains the spamming IP, the Autonomous System Number (ASN), netblock, country code, domain name, the timestamp, the botnet that spam through the IP, and the total spam volume.



February 2013 ▾ CA Canada ▾ ∀ All ▾ Classic.Sp@mRank

(Last Month, Next Month, Compare old rankings version 1)

Rank		Organization	Volume	Vol%
v2 v1	(Last) (Last)	Name Website ASnumber name	This (Last)	
1 1	(3) (3)	Bell Canada www.bce.ca/en/ AS577 BACOM	5,908,245 (8,266,395)	41.6%
2 2	(4) (4)	Shaw Communications Inc. www.shaw.ca/en-ca AS6327 SHAW	4,671,628 (6,734,462)	32.9%
3 7	(1) (1)	iWeb Technologies Inc. iweb.com AS32613 IWEB-AS	1,072,684 (147,257,179)	7.55%
4 3	(5) (5)	Rogers Cable Communications Inc. www.rogers.com AS812 ROGERS-CABLE	660,616 (941,667)	4.65%
5 33	(7) (22)	TekSavvy Solutions Inc. www.teksavvy.com/ AS5645 TEKSAVVY	398,353 (479,323)	2.8%
6 4	(14) (11)	Primus Telecommunications Canada Inc. www.primustel.ca/ AS6407 PRIMUS-AS6407	378,558 (230,089)	2.67%
7 6	(8) (10)	Telus Advanced Communications www.telus.com/ AS852 Telus	306,456 (375,728)	2.16%
8 5	(11) (9)	Bell Aliant Regional Communications Inc. promo.bellaliant.net/PromoPage.jsp AS855 CANET-ASN-4	304,979 (316,706)	2.15%

Figure 1. Screenshot of SpamRankings.net

The data processing is mainly aggregating and mapping IP level data to netblocks, then to Autonomous Systems (ASes), which are groups of IPs under the administration of an organization, and eventually to the organization. In addition to the mapping data included in the data files, we also used BGP routing data from Team Cymru to cross check as the mapping is dynamically changing. In our dataset, only less than 5% companies have multiple ASes. So the main analysis is performed at the AS level. We then aggregate daily outbound spam volumes into monthly volumes for each AS. Lastly for each country, we derive monthly rankings by outbound spam volume for all companies/ASes within the country. Top 10 organizations with the highest spam volumes for each treated country were disclosed on SpamRankings.net (Figure 1). For each Top 10 AS, the following information is disclosed: rank, rank in the previous month if it was listed in the previous month ("-" if not), name and website of the organization, ASN, and outgoing spam volume.

3.2. Quasi-experimental design

We used a between-subjects quasi-experimental design with two conditions: the treatment of imposing organizations subject to information disclosure on SpamRankings.net, and the control without any potential disclosure threat. Ranking for organizations within a country can strengthen the reputational effect but also makes the selection of treatment organizations clustered by country. We chose to start with treating the organizations in United States, Canada, Belgium, and Turkey, considering the potential publicity of SpamRankings.net in different countries. Then we aim to find comparison countries that can mimic the properties of the four treatment countries. Matching is used to select sufficient observable factors that countries with the same values of these factors will display no systematic differences in their reactions to the treatment [36].

Our matching is based on the combination of two types of observables: economy and IT infrastructure situation, and the spam trend before the treatment. For each country, we collected the country level economy and IT infrastructure statistics for the year of 2011 from the World Bank. The economy indicators include GDP, GDP per capita, GDP growth, unemployment, population, and population growth. The IT infrastructure indicators include Internet users, fixed broadband subscribers, mobile cellular subscribers, and secure Internet servers. To measure the similarity between countries, we standardized these statistics and calculated the Euclidean distance between each treatment country and the other countries. To measure the similarity in spam trend before the treatment, we used spam volume data for each country from January 2011 to April 2011 and calculated the variations in the log volume differences between each treatment country and the other countries over time. Systematic differences in levels were not the main concern, since they can be controlled for using diff-in-diffs methodologies. Instead, the variability in the difference between the two curves was minimized to make the difference as constant over time as possible [36]. Log volume (+1) is used instead of volume because of the extreme skewness in the data, and +1 is used to maintain zero volume observations. Combining the two types of factors, the comparison countries for each treatment country is composed of seven countries (Table 1) with the most similar spam trend (Figure 2) and fairly similar economy and infrastructure statistics (Table 2).

Table 1. Matched country groups

Pair	Group	Country	Pair	Group	Country
1	treated	United States (US)	3	treated	Belgium (BE)
	control	United Kingdom (GB)		control	Austria (AT)
		Japan (JP)			Germany (DE)
		Hong Kong(HK)			Greece (GR)
		Estonia (EE)			Italy (IT)
		Egypt (EG)			Portugal (PT)
		Moldova (MD)			Russia (RU)
		Australia (AU)			Singapore (SG)
2	treated	Canada (CA)	4	treated	Turkey (TR)
	control	Barbados (BB)		control	Iraq (IQ)
		Bahamas (BS)			Costa Rica (CR)
		Peru (PE)			Cyprus (CY)
		Poland (PL)			Dominica (DM)
		Luxembourg (LU)			Ecuador (EC)
		Hungary (HU)			Morocco (MA)
		Finland (FI)			Jamaica (JM)

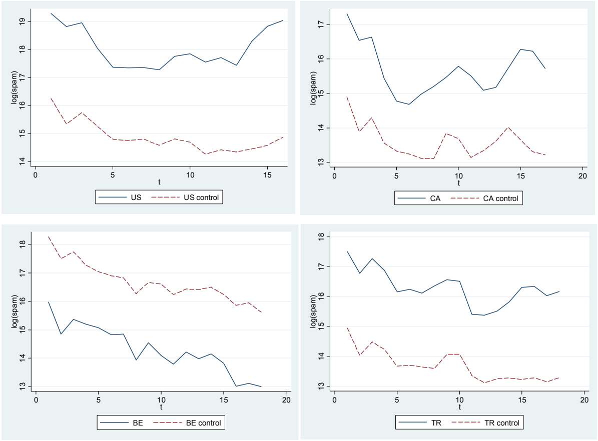


Figure 2. Spam trends of the treatment vs. comparison countries

Table 2. Economy and IT infrastructure statistics for the treated vs. control countries

	US	US control	CA	CA control	BE	BE control	TR	TR control
Internet User (per 100 people)	77.9	67.4	83	69.7	78	64.3	43.1	39.7
Broadband Internet Subscribers (per 100 people)	27.35	21.84	31.83	18.56	32.81	22.89	10.26	8.37
Mobile cellular subscribers (per 100 people)	95	128.6	80	126.4	120	142.1	89	107.9
Secure Internet servers (per 1 million people)	1563	784	1369	664	597	460	144	248
GDP per capita (current US$)	49,854	29,241	51,554	32,960	46,422	34,071	10,605	9,105
Unemployment (% of total labor force)	8.9	7.4	7.4	8.8	7.1	8.3	9.8	8.3
GDP growth (annual %)	1.8	3.7	2.5	2.9	1.8	1.1	8.8	4.0
Population growth (annual %)	0.7	0.67	1	0.94	1.4	0.34	1.3	1.21
GDP (current US$)	1.55 E+13	1.47 E+12	1.78 E+12	1.67 E+11	5.13 E+11	1.27 E+12	7.75 E+11	6.25 E+10
Population (million)	311	43	34	12	11	46	73	13

Figure 2 shows the time trends in spam volume in the pre-treatment period for the treatment countries and in average spam volume for each of their comparison countries. For each treatment country, the comparison countries exhibit very similar systematic changes in spam volumes. Table 2 presents the economy and IT infrastructure statistics for each treatment country and the average for its comparison countries. In spite of significant differences, each treatment country and its comparison countries are generally comparable in these aspects. As a result, our treatment group consists of ASes in four countries, and our control group consists of ASes in 28 countries. This matched pair design helps to control for the treatment variability among different country pairs, and thus can reduce the variance of the estimated treatment effect and lead to greater precision [37].

We started the treatment (information disclosure on SpamRankings.net) in May 2011 for the United States, June 2011 for Canada, July 2011 for Belgium and Turkey. This sequential release was designed to accumulate publicity for SpamRankings.net before getting into the full-scale experiment. For the control group, we did not disclose any information, but the same data were collected and kept internally. We also collected static information on each AS, including number of IP addresses, number of unique IP addresses, number of prefixes, number of regions, network name, website, network type, traffic level, inbound versus outbound traffic ratio, and geographic scope. The primary dependent variable is the outgoing spam volume after treatment. The sample ASes were included in either the treatment or control condition because they were observed to send out spam. Therefore, we have a selection bias toward ASes with severe outgoing spam problems. This is not an issue in this study since these ASes are the ones we aim to target.

Because we need to engage both organizations and consumers and test the treatment effect based on their natural reactions, it was critical for the success of the experiment to accumulate sufficient visibility and attention of SpamRankings.net in the four countries. We promoted the website through different channels, including social media such as YouTube, Twitter, and blogs, traditional media like newspaper and magazines, conferences, and press releases, to increase its visibility. We also received positive feedback and collaboration requests from industries. For example, we received the following comment from a Chief Security Officer of a medical center:

> *"The first time we were rated #1 on your list, we noticed that one of our users had generated thousands of spam messages and asked her to change her password—that stopped the spam immediately…The listing on your site added additional impetus to make sure we 'stay clean'…"*

4. Data analysis

4.1. Data and manipulation check

We collected data on all the spamming ASes each month from January 2011 to April 2013 for the selected 32 countries. We then dropped the ASes without any spam in any period before the treatment. The total unique sample size is 11,333 ASes, with 5,948 s in the treated group and 5,385 in the control group. Table 3 shows the summary statistics of network characteristics for sample ASNs by matched country pair, including prefixes, BGP peers, IP addresses, and AS path length. The treated country and its control countries are generally comparable in these network characteristics.

Table 3. Summary of sample ASes by country

	Number of ASNs	Average number of spamming ASNs per month	Average number of prefixes	Average number of BGP peers	Average number of IPs	Average AS path length
Treated						
US	5320	1805	59.6	36.4	218,485	4.1
CA	447	193	77.7	29.7	113,590	4.0
BE	61	23	17.1	20.8	162,576	3.9
TR	120	46	140.6	15.8	162,119	4.7
All treated	**5948**	**517**	**62.1**	**35.1**	**208,910**	**4.1**
Control						
US control	1155	502	74.7	66.5	283,647	4.0
CA control	948	342	26.2	26.3	54,820	4.1
BE control	3180	1300	31.0	72.5	77,650	4.1
TR control	102	39	190.1	9.7	129,347	4.6
All control	**5385**	**546**	**47.7**	**61.6**	**124,095**	**4.1**

The major concern for quasi-experiment is the potential selection bias as a result of deliberate choice. The experimental results are attained by comparing the outcomes of the treated group and the control group, which is based on the assumption that the two groups would have generated similar outcomes without the treatment. Experiment maintains this assumption by assigning subjects into two groups randomly. For quasi-experiment where the assignment is by choice rather than by chance, additional manipulation check is necessary to make sure no significant difference exists. On the country level, Figure 3 and Table 2 have shown that the treated countries and the control countries have similar time trends and economy and IT infrastructure statistics. On the organizational level, we use simple regressions to test the difference in pre-treatment conditions that may be correlated with the outbound spam as in the following equation:

$$Y_{icp} = \theta_0 + \theta_1 D_c + \varepsilon_{icp}\,, \qquad (1)$$

where Y_{icp} is the variable that needs to be balanced between the treated and the control, i is the index for AS, c is the index for country, p is the index for country pair, and D_c is the treatment indicator. θ_1 is the coefficient of interest. Significant θ_1 suggests selection bias. We also include pair fixed effect as a covariate

$$Y_{icp} = \theta_0 + \theta_1 D_c + \omega_p + \varepsilon_{icp}. \qquad (2)$$

ε_{icp} is the random error term. Because companies of the same country share the same regulation, culture, policy, etc., ε_{icp} is likely to be correlated within each country. The correlation would lead to underestimated standard errors of θ_1 [38]. To solve this issue, we use cluster-robust standard errors clustered at the country level. Cluster-robust standard errors allow for both error heteroskedasticity and flexible within-cluster error correlation. If not otherwise indicated, all standard errors reported in the paper are clustered by country.

We then apply Equation (1) and (2) to AS characteristics such as number of prefixes, number of BGP peers, IPs, and AS path length, and AS spam volume. For AS pre-treatment spam volume, time fixed effects are included to control for the systematic variations over time. Table 4 presents the results of the balance check. Column 1 contains the mean of the control organizations. Column 2 presents the estimated differences between the treatment and control organizations without country pair fixed effects (Equation (1)). Column 3 shows the results when country pair fixed effects are included (Equation (2)). According to Table 4, after controlling for matched pair fixed effects, the differences between the treated and the control ASes are insignificant for all AS characteristics and pre-treatment spam volume. In other words, our matching design yields a well-balanced treatment and control samples.

Table 4. Balance check between the treated and the control organizations

	Control mean (1)	Treatment difference without Country Pair FE (2)	Treatment difference with Country Pair FE (3)
Prefixes	47.66	14.39	-2.68
		(10.31)	(14.57)
BGP peers	61.61	-26.53***	-23.50
		(8.50)	(14.29)
IPs	124096	84814	-29087
		(52446)	(69546)
AS path length	4.08	0.03	0.03
		(0.09)	(0.14)
Spam before treatment	218439	-64490***	-62698
		(11463)	(42835)

For spam before treatment, time fixed effects are also included.
Standard errors are clustered by country and shown in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

4.2. Disclosure effect

We start with the standard difference-in-differences (DID) model [38] as follows:

$$Y_{icpt} = \omega_p + \lambda_t + \delta D_{ct} + \varepsilon_{icpt}, \tag{3}$$

where Y_{icpt} is the spam volume of AS *i* in country *c* of country pair *p* at time *t*, D_{ct} is the treatment indicator. According to Figure 3, different treatment-control pairs have substantial different time patterns in spam volume. For this concern, we further allow the time fixed effects to be different for different pairs as in Equation (4).

$$Y_{icpt} = \omega_p + \lambda_{pt} + \delta D_{ct} + \varepsilon_{icpt}. \tag{4}$$

Additionally, outgoing spam volume may also be influenced by AS characteristics such as prefixes, BGP peers, IPs, and AS path length. We use X_{icp} to present the vector of time invariant AS characteristics. Equation (5) shows the full model with all covariates.

$$Y_{icpt} = \omega_p + \lambda_{pt} + \delta D_{ct} + \beta X_{icp} + \varepsilon_{icpt}. \tag{5}$$

Equation (3)-(5) are then used to estimate the effect of our treatment, the information disclosure. The results are reported in Table 5. The estimates of the treatment effect are consistent and significantly negative across different models, supporting that information disclosure can help reduce outgoing spam. Comparing column (1) with (5) or (4) with (6), we can see that estimation with all country sample without controlling for country differences tends to overestimate the treatment effect. Since our dependent variable is log(Spam), the average spam with treatment is estimated to be 57% (=exp(-0.559)) of the average spam without treatment, which means the disclosure reduces spam volume by about 43%. Romanosky et al. [28] found that the adoption of data breach disclosure law can reduce identity theft caused by data breaches, on average, by 6.1%. Besides dependent variables, two possible reasons may explain the difference in the outcome. First, the public disclosure can impose more incentive, than individual data breach notifications, on organizations to deal with security issues. Second, many spam sent out from the same account are due to the same vulnerability. So dealing with the vulnerability can reduce spam volume significantly or even stop it completely. We observe that spam volumes of many sample ASes dropped to zero in subsequent month.

Comparing column (3) and (4), controlling for AS characteristics improves R square substantially. Among the AS characteristics, number of prefixes and IPs measures the size of the AS. BGP peers are the routing neighbors of the AS on the Internet. Number of peers measures the connectivity of the AS to other networks. AS path describes a sequence of connected domains that form a path from the current point to the originating domain. BGP path selection algorithm chooses AS of shorter path length. Among these measures, we find only prefixes and IPs have significant positively impacts on outgoing spam volume (Column (6)), indicating that the larger the AS, the more spam it generates.

Table 5. Information disclosure effect

	Matched country sample only				All country sample	
	(1)	(2)	(3)	(4)	(5)	(6)
D_{ct}	-0.016	-0.491***	-0.407*	-0.559**	-0.318*	-1.204***
	(0.298)	(0.134)	(0.220)	(0.258)	(0.175)	(0.240)
Prefixes				0.002***		0.002***
				(0.0004)		(0.0005)
BGP peers				0.0004		0.00013
				(0.0003)		(0.0003)
log(IPs)				1.121***		1.197***
				(0.143)		(0.113)
AS path length				-0.305		-0.086
				(0.196)		(0.161)
Constant	2.719***	3.237***	2.732***	-6.254***	3.042***	-6.596***
	(0.363)	(0.402)	(0.171)	(2.032)	(0.251)	(1.042)
Time fixed effects	N	Y	N	N	N	Y
Country pair fixed effects	N	N	Y	Y	/	/
Country pair specific time effects	N	N	Y	Y	/	/
R-squared	0.0%	1.8%	3.5%	36.9%	0.6%	34.1%
Observations	318,164	318,164	318,164	73,612	545,692	125,608

Standard errors are clustered by country and shown in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

To examine how treatment effect varies with AS characteristics, we introduce the interaction term between the treatment and AS characteristics into

Equation (5) to allow the treatment effect to be different for ASes with different number of prefixes, BGP peers, IP addresses, and AS path length. The results are presented in Table 6. Country pair fixed effects and country pair specific time effects have been controlled for in all columns.

Table 6. Interaction effects between information disclosure and AS characteristics

	(1)	(2)	(3)	(4)
D_{ct}	-0.517*	-0.531*	1.532	0.820
	(0.281)	(0.273)	(1.119)	(0.769)
Prefixes	0.00224***	0.00205***	0.00206***	0.00201***
	(0.000426)	(0.000414)	(0.000410)	(0.000453)
BGP peers	0.000393	0.000578	0.000386	0.000381
	(0.000301)	(0.000377)	(0.000294)	(0.000296)
log(IPs)	1.120***	1.120***	1.173***	1.121***
	(0.142)	(0.142)	(0.151)	(0.143)
AS path length	-0.305	-0.303	-0.295	-0.232
	(0.196)	(0.197)	(0.203)	(0.206)
D_{ct}* Prefixes	-0.000639			
	(0.000385)			
D_{ct}* BGP peers		-0.000684		
		(0.000512)		
D_{ct}* log(IPs)			-0.203*	
			(0.110)	
D_{ct}* AS path length				-0.337*
				(0.177)
Constant	-6.252***	-6.261***	-6.836***	-6.558***
	(2.023)	(2.032)	(2.145)	(2.013)
Country pair fixed effects	Y	Y	Y	Y
Country pair specific time	Y	Y	Y	Y
R-squared	36.9%	36.9%	37.0%	36.9%
Observations	73,612	73,612	73,612	73,612

Standard errors are clustered by country and shown in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

We find significant interaction effects between the treatment and two AS characteristics: IP addresses and AS path length. According to Column (3) and (4), the treatment would be more effective on ASes with more IP addresses and longer AS path length. It suggests that the disclosure would improve spam more effectively for large companies. This is consistent with the intuition that large companies care about their reputations more than small companies.

4.3. Robustness checks

To check if the result is driven only by one outlier in the four country pairs, we further allow the treatment effects to be different for different pairs. This also gives us an idea of how treatment effects vary across matched pairs. We find that except for Turkey, the treatment effects are negative and significant for United States, Canada, and Belgium.

Our specific disclosure mechanism only lists top 10 most spamming ASes. It is possible that the estimated treatment effect is purely driven by the spam reduction of the listed top 10 organizations only without any effect on the rest. To test whether there exists significant difference between the reactions of listed and non-listed organizations, we let Top10_{icpt}=1 if AS *i* was listed as top 10 ASes in previous month, and 0 otherwise, add the interaction term between Top10_{icpt} and D_{ct} in equation (5), and run the regression again. We do not find significant difference in treatment effects between listed and non-listed organizations.

If the reputation incentive is indeed driving the treatment effect, we should be able to observe stronger treatment effect on organizations more security sensitive, such as banks and financial service companies, and weaker effect on less sensitive organizations such as public services. We were able to collect industry data on US ASes from LexisNexis Academic database. We find the treatment effect the least for organizations in agriculture, forestry and fishing, mining, construction, and public services, and the most for organizations in manufacturing, finance, insurance, and real estate services. The treatment effect is not significant for communications services companies such as AT&T, T-Mobile, Comcast Cable, and Time Warner Cable. Telecommunication companies are special cases because their customers are partially responsible for the observed spam but our data collection can only trace back to the company.

In addition, we test for the spillover effect among ASes within an organization. We find that among the non-listed ASes from organizations with multiple ASes, those with other ASes within the same organization being listed as top 10 most spamming ASes in previous month has reduced more spam than those without.

5. Conclusions

Governments, businesses, and consumers are constantly exposed to the risk of cybercrime. Our society has recognized the need for disclosure laws to protect consumer privacy, enterprise assets, intellectual property, and critical national infrastructure. In the thriving and fast-moving discipline of Internet security, many are searching for technical solutions such as firewall and antivirus software. We propose that Internet security needs to be improved from the perspective of fundamental motivations. Systems are prone to failure when the person guarding them is not the person who suffers when they fail [6]. An organization's security vulnerabilities are also bared by other organizations but are often kept as private. The negative externality gives Internet security the feature of partial public good. For public goods, the private provision often results in underinvestment because of the lack of incentives. In social psychology, the underinvestment problem is often addressed through making relevant social information available and soliciting social comparison process.

Drawing upon theories on corporate reputation and social comparison, we propose an information disclosure mechanism to encourage organizations to improve their Internet security. Our disclosure mechanism so the disclosed information is easier to be standardized and more comparable. Our disclosure is based on existing Internet data collected by third party organizations so no individual reporting is needed from organizations, which avoids high reporting and auditing costs for security information disclosure. Our disclosed information is rankings for organizations so the information is easy to understand and interpret for the public, and thus shame companies who are not behaving socially responsible.

The information disclosure leverages reputational effect to influence companies' behaviors. Using a field quasi-experiment on outgoing spam for over 10,000 organizations in 32 countries, we provide evidence for information disclosure effect. We show that disclosing social information on outgoing spam encouraged the treatment organizations to reduce spam significantly. Comparing to an existing study [28] which documented a 6.1% effect of adopting data breach disclosure laws on identity theft, our result shows that making social information publicly available is more effective than notifying only affected consumers in motivating organizations to improve information security. Rao and Reiley [12] conservatively estimate that American firms and consumers experience costs on the order of \$20 billion annually due to spam. The estimated magnitude of the treatment effect, reducing spam volume by 33%-43%, shows that it is worthwhile to enable such a mandatory disclosure procedures of spam information, taking into account the total costs of data collection, processing, and disclosure.

The disclosure more strongly affects organizations with more problems. The size of the organization is found to be negatively interacting with the treatment, suggesting large organization tend to be more reactive to information disclosure. We also explore the industry difference in their reactions to the disclosure of spam information, and find that organizations in the industry more sensitive to information security react more strongly to the disclosure. The insignificant effect on communications service companies calls for special attention for enforcing disclosure in communications sector, since these companies often end up being blamed for their customers' conduct.

Our findings have implications for information architecture design and public policy on cyber security. With the rise of big data, the question we currently face is not the lack of data but how to make use of the available data. Both individual and corporate users care about their popularities, reputations, and social status within their communities. We can capture users' actions, aggregate and display the relevant information, and provide the appropriate feedback as the intervention, and eventually influence their behaviors. This gives rises to many inexpensive and efficient solutions for social problems such as pollution, donation, energy conservation, etc.

With respect to public policy, our present work is among the few empirical studies on Internet security using security vulnerabilities data. Our study provides a cost-efficient way to enforce mandatory disclosure. Policy makers have hesitated to use security information disclosure for a long time. The debate has been focused on whether we should have disclosure or not. Yet little attention has been paid to information display or presentation. We believe what is more important than disclosure is whether the information is understandable for users to compare and interpret. We use relative rankings as the specific information presentation form to enhance the disclosure effect. For policy evaluation, more information presentation methods can be considered and compared before implementing the policy extensively. Field experimentation provides an efficient and effective method to evaluate potential policies beforehand. The same approach applies to other security, social, or environmental problems as well. In the case where data is not available, the legislation that requires mandatory reporting can be employed to collect data.

6. Limitations

First, we experimented only with ranking information in our study to focus on relative standing. To identify the relative effect of using ranking information versus absolute volume information, we can add a new treatment group where organizations receive information only on absolute outbound spam volume and they are listed alphabetically. Second, the observation of reduction in outgoing spam may or may not reflect the improvement in overall Internet security. If overall Internet security improves while spam decreases, it indicates that companies take the initiative to improve their overall security, affecting both vulnerability to spam and other threats such as phishing. In this case, overall improvement in information security can be achieved by disclosing information on certain security issues. It is also possible that in response to public information disclosure of outbound spam, organizations may take effort to address only outbound spam issue but ignore other security problems. In this scenario, each security issue needs to be addressed individually. We can distinguish these two scenarios with data on multiple security issues. In addition, we do not consider possible reactions of attackers to information disclosure. These are important directions to extent our study in the future.

7. References

[1] Ponemon Institute. 2012. “2012 Cost of Cyber Crime Study: United States”, October (http://www.ponemon.org/local/upload/file/2012_US_Cost_of_Cyber_Crime_Study_FINAL6%20.pdf)

[2] Verizon, 2012 Data Breach Investigation Report, pp. 3.

[3] Arora, A., R. Krishnan, A. Nandkumar, R. Telang, Y. Yang. 2004. “Impact of Vulnerability Disclosure and Patch Availability: An Empirical Analysis,” *Third Workshop on the Economics of Information Security* (24), pp. 1268-1287.

[4] Anderson, R. 2001. “Why Information Security Is Hard: An Economic Perspective,” *Proceedings of the 17th Annual Computer Security Applications Conference (ACSAC)*, New Orleans, LA.

[5] Schneier, B. 2002. “Computer Security: It's the Economics, Stupid,” *Workshop on Economics and Information Security*, University of California, Berkeley, CA.

[6] Anderson, R., T. Moore. 2006. “The Economics Of Information Security,” *Science* (27:314), pp. 610-613.

[7] Coase, R.H. 1960. “The Problem of Social Cost,” *Journal of Law and Economics* (3), pp. 1-44.

[8] Dahlman, C.J. 1979. “The Problem of Externality,” *Journal of law and economics* (22:1), pp. 141-162.

[9] Pigou, A. 1920. *The Economics of Welfare*. McMillan&Co., London.

[10] Caballero, J., C. Grier, C. Kreibich, V. Paxson. 2011. “Measuring Pay-Per-Install: The Commoditization of Malware Distribution,” *Proceedings of the 20th USENIX Security Symposium*.

[11] Kanich, C., C. Kreibich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson. S. Savage. 2008. “Spamalytics: An empirical analysis of spam marketing conversion,” *Proceedings of the 15th ACM Conference on Computer and Communications Security*.

[12] Rao, J. M., D. H. Reiley. 2012. “The Economics of Spam,” *The Journal of Economic Perspectives* (26:3), pp. 87-110.

[13] Kraut, R. E., J. Morris, R. Telang, D. Filer, M. Cronin, S. Sunder. 2002. “Markets for attention: Will postage for email help?” *CSCW ‘02 Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work* 206–215, New York, NY.

[14] Loder, T., M. Van Alstyne, R. Wash. 2004. “An economic answer to unsolicited communication,” *Proceedings of the 5th ACM conference on Electronic commerce* 40-50.

[15] Sipior, J. C., T.W. Burke, P. G. Bonner. 2004. “Should Spam Be On The Menu?” *Communications of the ACM* (47:6), pp. 59-63.

[16] Balakrishnan K, A. Ghose, P. Ipeirotis. 2008. “The Impact of Information Disclosure On Stock Market Returns: The Sarbanes-Oxley Act and The Role of Media as An Information Intermediary,” *Proceedings of Seventh Workshop on the Economics of Information Security* (WEIS 2008), Hanover, New Hampshire.

[17] Loughran T., B. McDonald. 2011. “When is a liability not a liability? Textual analysis, dictionaries and 10-Ks,” *Journal of Finance*, 66(1) 35–65.

[18] Villiers, C. 2012. *Corporate Reporting and Company Law*, Cambridge, NY: Cambridge University Press.

[19] Field L., M. Lowry, S. Shu. 2005. “Does disclosure deter or trigger litigation?” *Journal of Accounting and Econom*ics, 39(3) 487–507.

[20] Verrecchia R. E. 1983. “Discretionary disclosure,” *Journal of Accounting and Economics,* 5(3) 179–194.

[21] Jorgensen B. N., M. T. Kirschenheiter. 2003. “Discretionary risk disclosures,” *The Accounting Review*, 78(2) 449–469.

[22] Farrow, R. 2000. “The pros and cons of posting vulnerability,” *The Network Magazine,* www.networkmagazine.com/shared/article.

[23] Arora, A., R. Telang, H. Xu. 2004. “Timing Disclosure of Software Vulnerability for Optimal Social Welfare,” *Proceedings of the 3rd Workshop of Economic Information Systems*, Minneapolis, MN, pp. 1–47.

[24] Gal-Or, E., A. Ghose. 2005. “The economic incentives for sharing security information,” *Information Systems Research* 16(2) 186-208.

[25] Gordon, L. A., M. Loeb, W. Lucyshyn. 2003. “Sharing information on computer systems security: An economic analysis,” *Journal of Accounting Public Policy* 22(6) 461–485.

[26] National Conference of State Legislatures (NCSL), 2012, http://www.ncsl.org/issues-research/telecom/security-breach-notification-laws.aspx.

[27] Lenard, T. M., P. H. Rubin, P.H. 2005. “Slow down on data security legislation,” *Progress Snapshot 1.9*, Washington, DC: Progress & Freedom Foundation.

[28] Romanosky, S., R. Telang, A. Acquisti. 2011. “Do Data Breach Disclosure Laws Reduce Identify Theft?” *Journal of Policy Analysis and Management* 30(2) 256-286.

[29] Campbell, K., L. A. Gordon, M. P. Loeb, L. Zhou. 2003. “The Economic Cost of Publicly Announced Information Security Breaches: Empirical Evidence from the Stock Market,” *Journal of Computer Security* (11), pp. 431–448.

[30] Kannan, K., J. Rees, S. Sridha. 2007. "Market reactions to information security breach announcements: An empirical analysis," *International Journal of Electronic Commerce* 12(1) 69-91.

[31] Duflo, E., P. Dupas, M. Kremer. 2011. "Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya," *American Economic Review* 101(5) 1739-1774.

[32] Hui, K. L., H. H. Teo, S. Y. T. Lee. 2007. "The value of privacy assurance: An exploratory field experiment," *MIS Quarterly* 31(1) 19-33.

[33] Suh, K., I. Benbasat, E. Suh. 2013. "The Impact of listing location on visits, bids, and final prices in online auctions: A field experiment," *International Journal of Electronic Commerce*, 17(3) 87-108.

[34] Messaging Anti-Abuse Working Group (MAAWG). 2011. "Email metrics program: The network operator's Perspective," *Report 14*.

[35] Tang, Q., L.L. Linden, J. S. Quarterman, A. B. Whinston. 2012. "Reputation as public policy for Internet security: A field study," *International Conference on Information Systems (ICIS)*, Orlando, FL.

[36] Blundell, R., Costa Dias, M. 2000. "Evaluation methods for non-experimental data," *Fiscal Studies,* 21 (4) 427-468.

[37] Raudenbush, S. W., X. Liu. 2000. "Statistical Power and Optimal Design for Multisite Randomized Trials," *Psychological Methods*, 5(2) 199-213.

[38] Bertrand, M., E. Duflo, S. Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics*, 119(1) 249-275.