

Singapore Management University

## Institutional Knowledge at Singapore Management University

---

Research Collection School Of Information Systems

School of Information Systems

---

2013

### Improving Internet Security Through Information Disclosure: A Field Quasi-Experiment

Qian TANG

*Singapore Management University, QIANTANG@smu.edu.sg*

Leigh L. LINDEN

*University of Texas at Austin*

John S. QUARTERMAN

*Quarterman Creations*

Andrew B. WHINSTON

*University of Texas at Austin*

Follow this and additional works at: [https://ink.library.smu.edu.sg/sis\\_research](https://ink.library.smu.edu.sg/sis_research)



Part of the [Information Security Commons](#)

---

#### Citation

TANG, Qian; LINDEN, Leigh L.; QUARTERMAN, John S.; and WHINSTON, Andrew B.. Improving Internet Security Through Information Disclosure: A Field Quasi-Experiment. (2013). 1-43. Research Collection School Of Information Systems.

Available at: [https://ink.library.smu.edu.sg/sis\\_research/1843](https://ink.library.smu.edu.sg/sis_research/1843)

This Working Paper is brought to you for free and open access by the School of Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email [library@smu.edu.sg](mailto:library@smu.edu.sg).

# **Improving Internet Security Through Social Information and Social Comparison: A Field Quasi-Experiment**

**Qian Tang**

McCombs School of Business, University of Texas at Austin, Austin, Texas 78712,  
[qian.tang@utexas.edu](mailto:qian.tang@utexas.edu)

**Leigh Linden**

Department of Economics, University of Texas at Austin, Austin, Texas 78712  
[leigh.linden@austin.utexas.edu](mailto:leigh.linden@austin.utexas.edu)

**John S. Quarterman**

Quarterman Creations, Valdosta, Georgia 31605  
[jsq@quarterman.com](mailto:jsq@quarterman.com)

**Andrew B. Whinston**

McCombs School of Business, University of Texas at Austin, Austin, Texas 78712,  
[abw@uts.cc.utexas.edu](mailto:abw@uts.cc.utexas.edu)

# Improving Internet Security Through Social Information and Social Comparison: A Field Quasi-Experiment

## Abstract

Cybersecurity is a national priority in this big data era. Because of negative externalities and the resulting lack of economic incentives, companies often underinvest in security controls, despite government and industry recommendations. Although many existing studies on security have explored technical solutions, only a few have looked at the economic motivations. To fill the gap, we propose an approach to increase the incentives of organizations to address security problems. Specifically, we utilize and process existing security vulnerability data, derive explicit security performance information, and disclose the information as feedback to organizations and the public. We regularly release information on the organizations with the worst security behaviors, imposing reputation loss on them. The information is also used by organizations for self-evaluation in comparison to others. Therefore, additional incentives are solicited out of reputation concern and social comparison. To test the effectiveness of our approach, we conducted a field quasi-experiment for outgoing spam for 1,718 autonomous systems in eight countries and published SpamRankings.net, the website we created to release information. We found that the treatment group subject to information disclosure reduced outgoing spam approximately by 16%. We also found that the more observed outgoing spam from the top spammer, the less likely an organization would be to reduce its own outgoing spam, consistent with the prediction by social comparison theory. Our results suggest that social information and social comparison can be effectively leveraged to encourage desirable behavior. Our study contributes to both information architecture design and public policy by suggesting how information can be used as intervention to impose economic incentives. The usual disclaimers apply for NSF grants 1228990 and 0831338.

**Keywords:** Internet Security, externality, social comparison, information disclosure, quasi-experiment, reputation, economic incentive

*“A spammer may be based in Latvia, work for a merchant in Moscow, send spam to the United States from a botnet with zombie computers all over the world, and have the final goods shipped from India.”*

*—Rao and Reiley (2012)*

## **1. Introduction**

2011 was a busy year for cyber attacks on many organizations, with targeted attacks increasing by 400%. Industries such as credit card companies, gaming platforms, banks, retailers, TV networks, and government agencies all fell victim to cybercrime, which is not only increasing in frequency but also in the severity of damage. According to the Ponemon Institute, the median cost caused by cybercrime is \$5.9 million per year per company, with a range from \$1.5 million to \$36.5 million. The costs consist of both direct expenses (recovery, detection, etc.) and indirect costs (information loss, business disruption, revenue loss, equipment damages, etc.). However, the study by Ponemon Institute also shows that nearly all of these attacks were avoidable. Most attacks were carried out using fairly simple methods and could have been stopped easily with basic or intermediate controls. Although most attacks were targeted, the target selection was based more on opportunity than on choice. Most organizations fell victims not because they were pre-identified but because they were found to possess exploitable vulnerabilities. About 50-75% of security incidents originated from within an organization (D’Arcy et al. 2009). Ninety-six percent of victim organizations subject to the Payment Card Industry Data Security Standard (PCI DSS) were not in compliance.

Organizations generally underinvest in Internet security because of the following 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. With the proliferation of mobile devices, the increasing number of locations and devices where information can be stored

further adds to the cost for prevention and protection. Second, although the costs for security are too high, the rewards are unclear. It is difficult to measure the risk and potential costs of security breaches beforehand. The frustrating fact about security is that although insecurity is easy to prove, security can never be conclusive. Third, the absence of legislative enforcement leads to the lack of transparency. Although recent progress in data breach notification laws requires companies to notify those customers whose information has been lost or stolen, companies generally can choose not to reveal publicly any attacks, in order to avoid reputation loss. Without transparency, organizations can claim to be secure even if their systems are, in fact, vulnerable, and customers cannot accurately estimate the risk of doing business with them. Moreover, Internet security is a public good in that an organization's security (insecurity) can benefit (hurt) others. The security vulnerabilities of an organization are often used against other organizations. For example, botnets opportunistically scan the Internet to find and compromise systems with exploitable weaknesses. These compromised computers are then utilized to collectively attack other targeted systems as in a typical denial of service attack.

Although they focus on technical solutions, existing studies often tend to ignore the motivational issue, which is a common problem in private provisions of public goods such as charitable giving (List and Lucking-Reiley 2002, Frey and Meier 2004, Shang and Croson 2009) and contribution to online communities (Bulter 2001, Beenen et al. 2004, Ludford et al. 2004, Chen et al. 2010). Social psychologists have documented the existence of social loafing—that people exert less effort on a collective task than they do on a comparable individual task (Beenen et al. 2004). According to social comparison theory, which was initially proposed by Festinger (1954), people have the desire to gain information on others and evaluation on themselves (Taylor and Lobel 1989). When information on others is available, people tend to evaluate themselves in comparison with others. As a result of the self-evaluation, the existence of discrepancy in a social group would lead to action on the part of group members to reduce the discrepancy. People generally care about their social status, often measured by ordinal ranks within their social groups, especially when status is made public and can influence one's

reputation (Griskevicius et al. 2010). As a result, status competition is often utilized to encourage desirable behaviors. Because of the concern for customer switch, organizations have even stronger incentives than individuals to maintain their status among peers. Reputation in Internet security signals a company's valuation for customer information and ability to take appropriate security controls. In the present article, we propose to solve the underinvestment problem by making such information publicly available, in order to solicit social comparison and status competition. Equivalent to rewarding prosocial behavior with status and prestige, we can penalize proself behavior with shame and reputation loss by making these behaviors notorious.

We incentivize organizations to increase security spending through our reputation system, an online website that regularly aggregates individual organizations' security information and releases explicit comparison results as relative performance ranking to the public. It is worth noting that often it is the information aggregation and feedback rather than the information itself that is missing. In the present study, we make use of the available information through third-party monitoring on outgoing spam as the focus security issue. However, the methodology also applies to other security problems, for which data can be collected through public policies on mandatory reporting, in the absence of available data. It has been recognized that a key factor required to improve Internet security is the gathering, analysis, and sharing of information related to security issues (Gal-Or and Ghose 2005). The Securities and Exchange Commission (SEC) formally asked public companies to disclose cyber attacks against them in October 2011. However, no pre-attack information is currently available for businesses and individuals to take precautionary actions. To solicit social comparison, the social information provided needs to reveal what constitutes the right behavior and who behaves that way and who does not.

To test the impact of the specific information released through our website, we conducted a field quasi-experiment in which the released information was used as experimental treatment. To draw attention to our system, we deliberately chose the United States, Canada, Belgium, and Turkey as four treatment countries and did extensive promotion for our website within these countries. The treatment

countries were then matched with four control countries according to population and the severity of the security problem before our experiment. Countries were used as clusters of organizations so that an organization was compared to other organizations within the same country. For organizations in the treatment group, the information on the organizations with the severest security problem in the country was released monthly on our website, whereas similar information was kept internally only for the control group. Although the treatment assignment is at the cluster level, the measurement is at the individual level. The field setting ensures that organizations and the public behave in a natural manner. A difference-in-difference model is used to test the treatment effect. The results show that the treated organizations improved their security situations more than the control organizations. We also find that the more security observed for other organizations, the more likely an organization will be to improve its own security situation.

Our approach for improving Internet security is complementary to existing technical approaches. The vast technical literature, especially in the computer science area, has focused on the development of technologies to secure computer systems, such as secure networking protocols, intrusion detection techniques, database security methods, and access control technologies (Ransbotham and Mitra 2009). By focusing on organizations' incentives to invest in these technologies, we aim to extend prior work and provide a more comprehensive lens for studying Internet security. Our study sheds light on public policy issues concerning security information disclosure and provides a new perspective for dealing with other environmental issues such as pollution, energy conservation, and global warming, where externality leads to a lack of incentives for taking pro-social behavior.

## **2. Literature Review**

### ***2.1 Internet Security***

Existing literature on information security focuses on organizational strategies that can be used for reducing system risk, including deterrence, prevention, detection, and recovery (Forcht 1994, Straub and

Welke 1998). For deterrence and prevention, most previous studies, from the organizational perspective, have examined the impact of security policy and practice on information systems abuse or misuse (Straub 1990, Kankanhalli et al. 2003, D’Acy et al. 2009). For detection and recovery, research has been focused on how to identify attack traffic that could originate from both internal and external sources in a cost-effective way (Toth and Kruegel 2002, Carver et al. 2000, Yue and Cakanyildirim 2007, Mookerjee et al. 2011) Specifically for anti-spam, the filtering techniques consist of machine learning (Sahami et al. 1998, Androutsopoulos et al. 2000, Goodman et al. 2007), crowdsourcing and IP blacklisting (Cook et al. 2006, Ramachandran et al. 2011), screening humans from bots for botnets (Kotadia 2004, Motoyama et al. 2010, Isacenkova and Balzarotti 2011), and Domain Keys Identified Mail (DKIM) (Moyer 2011). However, the problem for any technical solutions is that miscreants can always respond strategically. The interplay is an endless cat-and-mouse game.

There is a growing movement among Internet security professionals towards metric-driven security (Jaquith 2007, Baker et al. 2007), yet much of it remains focused on individual organizations. Some professional organizations survey Internet providers (Manzano 2009) about what they are doing to reduce spam. Some researchers have examined the economic role of ISPs in botnet mitigation (van Eeten et al. 2008) and have used metrics (van Eeten et al. 2010) and country-specific studies (van Eeten et al. 2011) to explore that topic. Most such studies are based on one-time surveys, and do not have ongoing, regular, publication on the net, except at low frequencies such as annually. Internet traffic measurements have always been available, and applications to public policy continue to be studied (Bauer 2012), but comprehensive, frequent, regular, ongoing Internet security measurements are still not common, and there are still fewer examples of such measurements made public in reputational rankings. There is also growing interest by ISPs in stanching outbound spam (CommTouch 2010), but such interest should be compounded by regular reputational rankings. Some researchers have called for cooperation against miscreants and pointed out the consequences of non-cooperation (Moore and Clayton 2008). The various anti-spam blocklists (see Section 3.2) of course represent one form of such cooperation, and some of



them, especially Spamhaus, also focus on specific large spamming organizations and cooperate in takedowns of them. Unfortunately, such takedowns usually have only short-term effects (Quarterman et al. 2011, Quarterman et al. 2012). Some organizations such as Team Cymru do try to deal with organizational cooperation in a quiet way. The present research leverages blocklist and Team Cymru and other data to supply transparent public organizational incentive for such cooperation, and provides theoretical support for such methods and incentives.

Security vulnerability disclosure is an area of public policy that has been subject to considerable debate (Arora et al. 2004b). 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 (Elias 2001 and Farrow 2000). Arora et al. (2004a) found that although vendors are quick to respond to instant disclosure, vulnerability disclosure also increases the frequency of attacks. Arora et al. (2004b) 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. Although there has been no public disclosure on information security vulnerability, industry-based Information Sharing and Analysis Centers (ISACs), where security breach information is 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 (2005) studied the economic incentives for security information sharing and found that information sharing yields greater benefits in more competitive industries. Gordon et al. (2003) 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.

## ***2.2 Regulations on Information Disclosure***

Security information disclosure laws have been focused on data breach notification. Although a national data-breach notification law is still under consideration, 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 (NCSL 2012). The specific requirements of notification laws vary across states. Some laws require notification when the personal information is reasonably assumed to have been acquired by an unauthorized party, whereas others require notification only if it is reasonable to believe the information will cause harm to consumers. The consequences of not complying differ from state to state as well. However, the rationales for these laws are consistent, which is also consistent with our rationale for public disclosure of security vulnerabilities, that notification can provide public information and create an incentive for all firms (even those that have not been breached) (Ponemon Institute 2005, Schwartz and Janger 2007, Romanosky et al. 2011).

However, the impact of data breach disclosure is still in debate. The concerns include the following: (1) Firms must comply with multiple, disparate, and perhaps conflicting state laws (Romanosky et al 2011); and (2) notifications simply shift the burden to consumers if breaches really cause harm (Lenard and Rubin 2005, Cate 2009). Otherwise, they are just unnecessary costs. Romanosky et al. (2011) found that data breach disclosure can reduce identity theft caused by data breaches. Campbell et al. (2003) found a highly significant negative impact of security breaches reported in newspapers on the stock price of the breached company only when the breach involved unauthorized access to confidential data. In contrast, Kannan et al. (2007) found that security breach announcements have no significant negative impact on market return in the long run.

The impact of information disclosure has also been widely studied in areas other than security. Jin and Leslie (2003) studied health information disclosure in the restaurant industry and found that disclosing hygiene quality information increases health inspection scores and lowers certain diseases. Cain et al. (2005) examined the effect of disclosing conflicts of interest and found that the disclosure can have perverse effects because advice receivers do not discount advice sufficiently, and that advice givers exaggerate advice even further. Other information disclosure studies are related to auto safety, public education, and so on (Fung et al. 2007). These studies provide some evidence of how information

disclosure can affect firm behavior. On the basis of these studies, we further add the aggregation and presentation of information, which can leverage reputation and peer influence to enhance disclosure effect.

### ***2.3 The Economics of Internet Security: Externalities and Incentives***

It has long been recognized that Internet security is not a problem that technology alone can solve (Arora et al. 2004a). 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 (Anderson 2001). 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 (Schneier 2002). Security failure is caused as much by bad incentives as by bad design (Anderson and Moore 2006). Meanwhile, vulnerability exploits, botnets, spamming, and other miscreant activities have 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 (Pigou 1920, Coase 1960, Davis and Whinston 1962, Dahlman 1979). 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 miscreants to attack other systems. For example, spam has such an extreme negative externality that the social costs are about 100 times the private benefits (Stone-Gross et al. 2011, Kanich et al. 2008, Caballero et al. 2011, Rao and Reiley 2012). More and more studies have recognized the importance of security externalities and have come up with several 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 (Pigou 1920, Coase 1960, Davis and Whinston 1962, Dahlman 1979). 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 (Kraut et al. 2002, Loder et al. 2004). However, these proposals raise the concerns for privacy and account hijacking by miscreants. 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 (Sipior et al. 2004, Levchenko et al. 2011). However, since most security problems such as spam and phishing may involve parties in different administrative areas, jurisdictional boundaries provide difficulties for such proposals.

#### **2.4 Social Comparison: Status, Shame, and Fame**

In a social community, participants tend to compare themselves to others when social information on other participants' behaviors is available, and such social comparisons in turn affect behaviors (Festinger 1954). Perceptions of relative standing can influence many outcomes. A number of studies have found that self-reported happiness may be more sensitive to relative than to absolute income (Hopkins and Kornienko 2004, Luttmer 2005). The interdependent preferences can be represented either by utility functions that depend not only on the absolute value of consumption but also on the average level of consumption within a population (Duesenberry 1949, Pollack 1976), or by including concern for status, the ordinal rank in the distribution (Frank 1985, Robson 1992, Direr 2001). The reasons for status concern may be intrinsic, a fundamental human characteristic, or instrumental: Status is desirable because it allows better consumption opportunities (Postlewaite 1998, Cole et al. 2001, Hopkins and Kornienko 2004).

The availability of social information is the prerequisite for social comparison. Recent theories on pro-social behavior have focused on "conditional cooperation": People are more willing to contribute when information is provided that many others contribute (Frey and Meier 2004). Satio (2011) suggested that individuals feel ashamed about a choice that does not maximize the payoffs of others only when the choice is made in public. Dillenberger and Sadowski (2010) also proposed that a person's behavior may depend on whether it is observed by someone who is directly affected by it and considered shame as a

moral cost for a person's utility. These concepts can be extended to organizational behavior since organizations are concerned about their social image and reputation (Frei 2010), their relative standing in comparison to other businesses. These social factors such as reputation and social image are valuable assets for a business not only because organizations have the desire for prestige, esteem, popularity, or acceptance (Bernheim 1994), but also because they lead to better business opportunities. It's not all name and shame: it's just as much praise and fame for ranking low, for improving in the rankings, for rapidly ejecting botnets, etc. With the increasing concern for privacy and confidentiality, customers are likely to choose or switch to firms with a more secure information system.

Social comparison and social information are often used to solve the problem of social loafing, the reduction in motivation and effort when individuals work collectively as compared with when they work individually (Beenen et al. 2004). The reasons include reduced individual motivation and coordination loss (Karau and Williams 1993). Both reasons exist in the context of Internet security. The former is due to the externalities, whereas the latter is due to the cost of security efforts. Reputation loss imposed by making relevant social information available can serve as a binding force against social loafing (Akerlof 1980). Social norm formed through social information provision has two effects on pro-social behavior: the focusing influence whereby norms impact behavior only when an individual's attention is drawn to them, and the informational influence whereby norms exerts a stronger impact on an individual the more he observes others behaving consistently with that norm (Krupka and Weber 2009). In the present article, we aim to leverage both effects to motivate pro-social behavior.

## **2.5 Previous Publications about This Project**

This paper adds significant new research and analysis beyond our previous publications. The RIPE Labs papers of 2010 (Quarterman et al. 2010c and 2010d), summarized in the RIPE conference presentation (Quarterman et al. 2010d), described what we intended to do with systematic public rankings, elaborating at some length on a pithy presentation at APWG (Quarterman 2010a). After deploying SpamRankings.net, we used it and related drilldown interfaces to illustrate the limited effects

of the Rustock Botnet takedown at the Telecommunications Policy Research Conference (TPRC), and to argue for the use of reputational rankings as a new and more encompassing approach to Internet security (Quarterman et al. 2011) motivated in part by commons theory (Ostrom 1990, Milinski et al. 2002, Dietz et al. 2003). The next year at TPRC, we illustrated the fall of Grum botnet and the rise of Festi botnet to motivate the need for another approach, using an initial successful case of medical organizations, which do seem to have all changed their Internet security behavior due to the launch of SpamRankings.net (Quarterman et al. 2012), and we argued the case for using clustered randomized control trials (Duflo 2010) with countries as the clusters (with supporting statistics on spam distribution per country and ASN) in full-scale worldwide Internet field studies, as well as providing further theoretical background on why reputational rankings should have beneficial effects. The TPRC 2011 (Quarterman et al. 2011) and TPRC 2012 (Quarterman et al. 2012) papers taken together with the even earlier NANOG 48 presentation about FireEye's takedown of the Ozdok or Mega-D botnet (Quarterman 2010a), comprise a side study of how major botnet takedowns did not have much effect on spamming or security beyond temporary short-term effects measured in weeks or a few months, thus mostly irrelevant to our longer-term studies. At ICIS 2012 we presented an overview of the project including some initial analysis of a small set of unrandomized pairs of treated and control countries. The present paper goes much further into that analysis, with further theoretical support from all the areas discussed earlier in this section and also in the next. While the results from this small sample do seem valid, this country-pair approach is limited by the difficulty of finding appropriate pairs of control and treatment countries. Further work, beyond the scope of the present paper, is proceeding on full-scale non-clustered randomized control trials.

### **3. Field Quasi-Experiment**

Field experimentation has been used extensively to provide solid knowledge of causation for policy evaluation (Duflo et al. 2011). It has also been used to study information security and privacy (Hui et al. 2007). Experimental studies randomly assign participants into treatment groups or control groups. Randomization, although more desirable in an ideal environment, is inappropriate given our

circumstance. In the present study, we aimed to evaluate whether public information disclosure can lead to security improvement; thus, the attention to the disclosed information is critical. Rather than randomly choosing some countries for treatment, it is more pragmatic to focus on the countries where the new information is more likely to receive attention. As a result, we used a quasi-experiment with intentional treatment on North American and European countries, to resemble the randomized field experiment, considering the authors' PR connections and promotional activities for our website. Quasi-experiments typically occur in real-world settings that more closely resemble the actual contexts and constraints faced by policymakers and practitioners (Remler and Van Ryzin 2011). Although randomized experiments generally have better internal validity (evidence of causation), quasi-experiments often turn out to have better external validity (generalizability).

### **3.1 Outgoing Spam**

Internet security is a very broad and general concept that has many dimensions. In the present article, we look into outgoing spam as one specific security issue. Referred to as unsolicited bulk emails, most spam messages are sent by botnets, a collection of compromised computers (bots), without the awareness of the legitimate computer owners. Anti-spam blocklists have spam traps scattered across the Internet and can recognize similar messages received at multiple locations. An estimated 88% of daily worldwide email traffic is spam (MAAWG 2011). Inbound spam refers to the spam received, and many organizations are well equipped to filter spam out of incoming emails before these emails reach their employees or users. However, they have very limited techniques to prevent outbound spam originated from computers within the organizations. Outgoing spam is typically generated via zombie computers, compromised user accounts, or spammers who knowingly abuse their accounts (e.g., in snowshoe spam), and it is a common symptom of more damaging security problems (Quarterman et al. 2010e). The same vulnerabilities that enable spam are also openings for other exploits. For example, miscreants can steal existing accounts by tricking end-users (through phishing or by human engineering) into providing their email usernames and passwords. Such stolen accounts can then be used to install botnet spamming malware or other exploits

such as Distributed Denial of Service (DDoS) software or sniffers, causing theft of customer records and intellectual property, fraudulent use of corporate online banking, or even employee blackmail.

It is costly to deal with outbound spam, which often leads to major side effects such as IP blocking by RBL, DNSBL, and IP reputation systems. These side effects cause queue buildup on the affected mail server, delays in message delivery, and may result in lost messages and calls from unhappy end-users. They also lead to compromised user accounts and blocking of legitimate outbound email, which then cause damage to reputation, customer relationship, business operation, and eventually lower profit. Unfortunately, conventional anti-spam measures may not work well for outbound traffic. Spam has an extreme negative externality in the sense that the ratio of external costs to private benefits is as high as 100:1, as compared with about 0.1 for pollution and 7:30 for nonviolent property crime (Stone-Gross et al. 2011, Kanich et al. 2008, Caballero et al. 2011, Rao and Reiley 2012).

Therefore, if ISPs are constantly sending out spam, they not only risk being attacked themselves, but also increase the risk 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, the largest botnet on record (Quarterman et al. 2011). The takedown of this single botnet was followed by an immediate one-third reduction in global email spam (Thonnard and Dacier 2011, Microsoft 2011, Rao and Reiley 2012). Hence, outgoing spam is a typical Internet security problem that lacks transparency, costs a lot to deal with, and generates negative externalities. If our approach proves effective in reducing outgoing spam, it can also be used for improving other security dimensions.

### **3.2 SpamRankings.net**

We launched a website named SpamRankings.net in May 2011 and have since used it to release country-specific outbound spam information. This website serves as our main instrument to study public security information disclosure and presentation. It displays monthly outbound spam volume and rankings



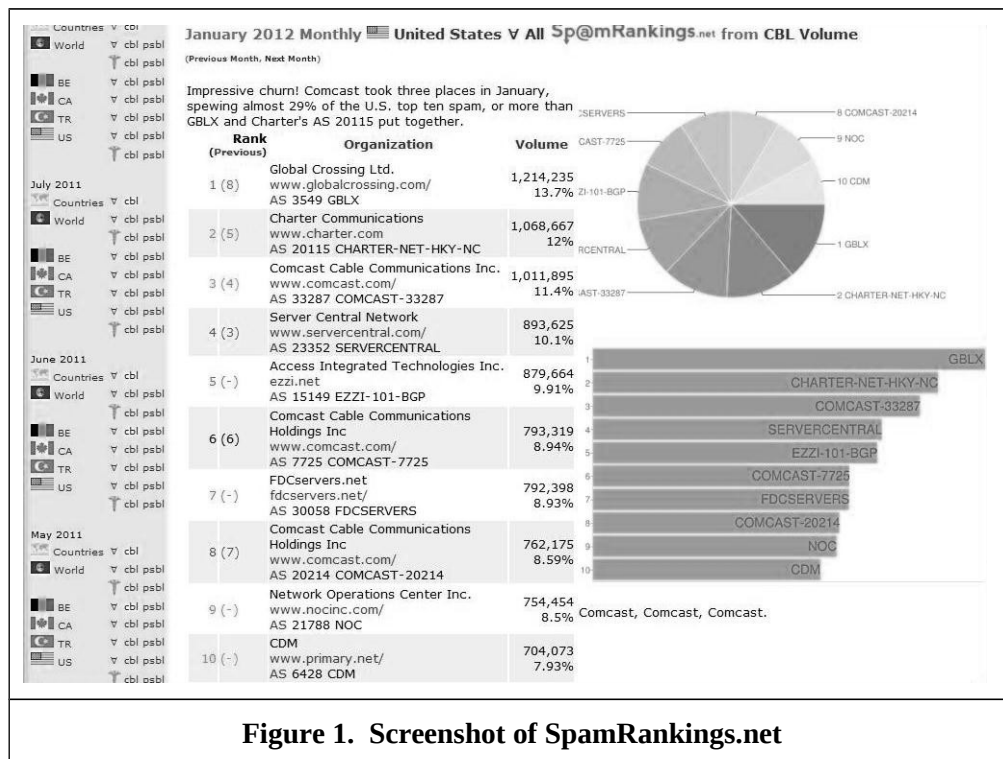
for sample organizations in the treated countries, including the United States, Canada, Belgium, and Turkey. To generate such information as treatment, we gathered and processed a large amount of daily spam data from two blocklists, the Composite Blocking List (CBL), and the Passive Spam Block List (PSBL). The CBL gathers its source data from its own spam traps and very large mail server installations and lists IPs exhibiting characteristics that are specific to open proxies of various sorts and dedicated Spam BOTs that have been abused to send spam, worms/viruses. The PSBL is an easy-on, easy-off blacklist that does not rely on testing and has a lower probability of false positives because any user can remove their ISP's mail server from the list.

We also collect daily blacklist data from Spamhaus (all three lists: SBL, XBL, and PBL), UBL, UCE, URIBL, as well as custom volume data from the University of Texas at Austin Department of Computer Sciences and Quarterman Creations. We occasionally use those other sources of data as cross-checks via our internal drilldown interfaces, and we hold them all in reserve for future studies. Each blacklist has its biases in placement of spam traps, in heuristics used to determine which messages to label as spam, etc. We explicitly addressed this issue before we started publishing rankings (Quarterman et al. 2010f). We use CBL for most of our published rankings and the experiment described in this paper because it provides us a large amount of data each day, with features such as volume counts (number of spam messages) and botnet labelling that are not available from most blocklists. Other blocklists are nonetheless useful for rankings, because while all blocklists are noisy and subject to bias, rankings based on a single blacklist have the same biases and noise. It is not some unattainable Platonic ideal of perfect knowledge about the Internet that matters for our purposes: it is the *relative* rankings that are important.

The raw data include observed spamming IP addresses, corresponding outbound spam volume, and botnet tags in the forms of text files from CBL and Network News Transfer Protocol (NNTP) messages from PSBL. One important step in data processing is mapping IP addresses to netblocks and, subsequently, Autonomous Systems (ASes), which are groups of IP addresses owned by an organization. Organizations with very large networks may use multiple ASes as a way to divide their networks.

Therefore, ASes, even within the same organization, are relatively independent of each other. Therefore we use ASes as the measurement level. An AS can be identified by a unique Autonomous System Number (ASN).

Lastly, we aggregate the daily outbound spam data into monthly data and derive rankings for each country. We receive more than eight million records per day from CBL and PSBL, which we summarize into about two million spam messages observed from worldwide IP addresses. On the ASN level, we have seen 27,500 ASNs with spam volume over the lifespan of this project. The ASNs are then grouped and ranked by country. The Top 10 organizations with the most spam are listed on SpamRankings.net (Figure 1). For each Top 10 ASN, we display the following information: 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.



**Figure 1. Screenshot of SpamRankings.net**

To map IP addresses to netblocks and to ASNs initially we used (version 1 or v1) a static copy of mappings from Team Cymru. Recently (published on SpamRankings.net May 2013 for April 2013 rankings) we have revised (version 2 or v2) our methods to use daily mappings from CBL crosschecked

by daily mappings from Team Cymru. After extensive internal consistency checking of v2 and comparisons with v1, it appears clear that while rankings derived from v2 increasingly diverge from those for v1 starting about October 2012 (well *after* the period of the experiment of this paper), the direction of movement of rankings or spam volume is the same for the great majority of ASNs and months; v2 is mostly more precise, including some netblocks and addresses that v1 did not. Precision here means in assignment of IP addresses to netblocks and netblocks to ASNs. Any spam blocklist data will always be imprecise as to the volumes of spam messages seen and the number of IP addresses that send them. As previously noted, the most important feature of the blocklist data for our purposes is relative rankings, and v2 makes those somewhat more precise while demonstrating that v1 rankings were already useful. The data used in the present paper are all from v1.

### **3.3 Quasi-Experimental Design**

To evaluate the impact of spam information released on SpamRankings.net, we used a between-subjects quasi-experimental design with two conditions: the treatment with information released on SpamRankings.net every month, and the control with information kept internally. To strengthen social comparison and reputation concern, information released in the treatment condition is relative ranking with respect to the outgoing spam, which also means that the intervention is at the cluster, the country level. Individual ASN level assignment would have resulted in less meaningful ranking information and weaker social comparison. Therefore, we nested ASNs within countries and assigned countries as clusters of ASNs to the two conditions. Considering the publicity of SpamRankings.net, we specifically chose the United States, Canada, Belgium, and Turkey as four treatment countries. We then matched the treatment countries with four control countries accordingly, based on population and total outgoing spam volume before our experiment, as shown in Table 1. Therefore, the control group consists of ASNs in Indonesia, Malaysia, Netherlands, and Iran.

For the treatment group, we made the monthly spam rankings available through SpamRankings.net from May 2011 to January 2012. We treated different countries with differing starting points in time, with

the United States in May 2011, Canada in June 2011, and Belgium and Turkey in July 2011. This sequential release was designed to accumulate publicity for our ranking site before getting into the full-scale experiment. For the control group, we did not publish any information on outbound spam, 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 outcome of interest is the outgoing spam volume. The sample ASes were included in either the treatment or control condition because they were observed to send out spam during May 2011 to January 2012. Therefore, we have a selection bias toward ASes with more severe outgoing spam problems. In the specific context of the present study, this was not a problem since these ASes were the ones we intend to impact. All of the ASes remained through the entire experiment.

**Table 1 Country Pairs**

Pair	Country	Population	Spam*	Group
1	United States (US)	310,232,863	57,176,031	treated
	Indonesia (ID)	242,968,342	94,435,116	control
2	Canada (CA)	33,679,000	4,387,388	treated
	Malaysia (MY)	28,274,729	6,695,830	control
3	Belgium (BE)	10,403,000	3,781,796	treated
	Netherlands (NL)	16,645,000	6,283,101	control
4	Turkey (TR)	77,804,122	14,759,146	treated
	Iran (IR)	76,923,300	13,291,908	control

\* All the data on spam in this paper refer to the data on outbound spam  
Spam data are taken from data for April 2011.

Since we wanted to engage both organizations and consumers and observe their natural reactions, it was critical for the success of the experiment to accumulate sufficient visibility and attention of SpamRankings.net. We promoted the website through different channels, including social media such as YouTube, Twitter, and blogs, conferences, and press releases, to increase its impact. We also received much feedback and collaboration requests from industries and observed that some organizations had already tried to take their names off the list on SpamRankings.net. For example, we received the

following comment from a Chief Security Officer of a medical center, which also confirms that some outgoing spam could be reduced using basic controls:

*“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 next month, we found another user who had just given up her credentials and got her to change her password as well. I spoke with a colleague at one of the other medical centers ranked on your site and he mentioned they have the same problem...The listing on your site added additional impetus to make sure we ‘stay clean’ so in that regard, you are successful.”*

#### **4. Data**

We collected outbound spam data on the top 250 most spamming ASNs each month from March 2011 to January 2012 for the eight selected countries. Table 2 shows the summary statistics of observed sample ASNs by country. Only the United States had over 250 spammers for some months, but the top 250 ASNs accounted for over 95% of the total outbound spam. The total unique sample size was 1,718 ASNs, with 1,177 ASNs in the treated group and 541 ASNs in the control group. However, if we look at the average number of ASNs with observed outgoing spam per month, we have a more balanced treatment group and control group. The unbalance is the result of the observation that spamming ASNs for the treatment group varied significantly from month to month, indicating that ASNs in the treatment group reduced their outbound spam more quickly than those in the control group. Results for average maximum of spam percentage by ASN show that a few ASNs were responsible for the most outbound spam volume in Indonesia, Turkey, Belgium, and Malaysia. Especially for Indonesia, the most spamming ASN sent out 83.46% of total spam on average. However, for the United States, the most spamming ASN accounted for only 6.89% of total spam on average.

**Table 2 Observed Sample ASNs by Country**

Number of ASNs	Average number of ASNs with positive spam volume	Average max of spam volume by	Average max of spam percentage by
----------------	--	-------------------------------	-----------------------------------

	per month		ASN	ASN
<b>Treated</b>				
US	699	250	3,414,080	6.89%
CA	316	175	1,261,576	20.94%
BE	56	31	1,116,462	44.70%
TR	106	63	5,051,160	48.08%
<b>Sum</b>	<b>1177</b>	<b>519</b>		
<b>Control</b>				
ID	229	190	45,903,492	83.46%
MY	57	44	1,958,958	43.11%
NL	170	101	1,067,763	22.91%
IR	85	77	2,575,711	27.89%
<b>Sum</b>	<b>541</b>	<b>413</b>		

Table 3 summarizes the outbound spam volume by country for both the periods before (Pretest Period) and during (Test Period) the experiment. This comparison presents the average outbound spam volume per month and the difference. On average, the outbound spam volume of the four countries in the treated group dropped by 54.93%, whereas the number for the four countries in the control group was 45.84%.

**Table 3 Outgoing Spam Volume Observed Per Month Per Country**

	Pretest Period*	Test Period*	Difference	Percentage
<b>Treated group</b>				
US	105,347,424	33,007,389	72,340,035	68.67%
CA	7,786,736	3,949,362	3,837,374	49.28%
BE	3,812,537	1,663,925	2,148,612	56.36%
TR	14,758,174	8,052,961	6,705,213	45.43%
<b>Average</b>	<b>32,926,218</b>	<b>11,668,409</b>	<b>21,257,809</b>	<b>54.93%</b>
<b>Control group</b>				
ID	93,416,115	46,320,078	47,096,037	50.42%
MY	6,361,998	3,684,334	2,677,663	42.09%
NL	7,261,624	2,086,952	5,174,672	71.26%
IR	10,590,092	8,515,070	2,075,021	19.59%
<b>Average</b>	<b>29,407,457</b>	<b>15,151,609</b>	<b>14,255,848</b>	<b>45.84%</b>

\*For US, the pretest period is 03/2011-04/2011; the test period is 05/2011-01/2012.

For CA, the pretest period is 03/2011-05/2011; the test period is 06/2011-01/2012.

For other countries, the pretest period is 03/2011-06/2011; the test period is 07/2011-01/2012.

## 5. Statistical Models

We estimate a linear model to test the effect of security information disclosure. First, we employ a simple difference specification to directly compare the treatment and control groups:

$$Y_{ict} = \theta_0 + \theta_1 D_c + \varepsilon_{ict} , \quad (1)$$

where the dependent variable  $Y_{ict}$  is the outcome of interest for AS  $i$  in country  $c$  at time  $t$ , and  $D_c$  is a treatment indicator variable for whether country  $c$  received security information disclosure. Hence, the estimate of the coefficient  $\theta_1$  indicates the difference between treatment and control countries. We utilize this model to compare baseline differences in pre-treatment conditions and to test the effect of spam information disclosure on firms' outbound spam.

Since the assignment of countries to treatment and control groups is not random in the present study, the outbound spam is likely to be affected by pre-treatment conditions. It is thus necessary to include observable AS characteristics and baseline spam volumes as control variables in equation (1) to improve the precision of the estimated treatment effect. Therefore, we also run the following specification:

$$Y_{ict} = \theta_0 + \theta_1 D_c + \theta_2 X_{ic} + \omega_p + \varepsilon_{ict} , \quad (2)$$

where  $Y_{ict}$  and  $D_c$  are defined as in equation (1), and  $X_{ic}$  is a vector of pre-treatment AS characteristics including baseline spam volume and number of IP addresses. Since the assignment to treatment and control groups was stratified within country pairs (Table 1), we also include country pair fixed effects,  $\omega_p$ , in equation (2).

We also examine whether the treatment effect interacts with baseline AS characteristics by running the following difference in differences model:

$$Y_{ict} = \theta_0 + \theta_1 D_c + \theta_2 X_{ic} + \theta_3 D_c * X_{ic} + \omega_p + \varepsilon_{ict} , \quad (3)$$

where the interactive term  $D_c * X_{ic}$  is added based on equation (2). The estimate of  $\theta_3$  captures the part of treatment effect moderated by baseline AS characteristics.

Outbound spam volumes from organizations within a country may be correlated because of common

policies and regulations for a country. These country-clustered missing variables would result in highly positively correlated error terms. Failure to correct for the correlation could result in underestimated standard errors and thus overestimated treatment effects (Bertrand et al. 2004). We therefore use cluster-robust standard errors at the country level (the level of treatment assignment) in estimation of all of the above models to allow for both error heteroskedasticity and flexible within-cluster error correlation. However, the asymptotic justification based on cluster-robust standard errors assumes that the number of clusters goes to infinity. With few (five to 30) clusters, cluster-robust standard errors can still be understated (Cameron et al., 2008). Since we have only eight clusters (eight countries), this problem is likely to exist. So we further use the wild cluster bootstrap-t procedure suggested by Cameron et al. (2008) to adjust the estimated standard errors for  $\theta_1$  in our main models (equation (2)).

## **6. Baseline Comparison**

Since assigning firms into treatment and control groups is not random, it is necessary to test the difference in pre-treatment conditions that may be correlated with the outbound spam. If the difference is not statistically significant, then any differences in post-intervention outcomes between the two groups can be causally attributed to the intervention. Otherwise, the pre-treatment difference needs to be controlled in order to make a precise estimation on treatment effect. To check whether firm characteristics were similar or not between the two groups, we run regressions of the number of IP addresses and pre-treatment baseline spam volume (average spam volume for March and April 2012) on treatment status using equation (1).

We present the comparison of firms at baseline in Table 4. Column 1 contains the average characteristics for the control group. Columns 2 and 3 present the estimated differences between the treatment and control groups. The results in column 2 do not include any controls, whereas those in column 3 control for country pair fixed effects. The differences in average baseline spam and IP number are statistically significant and large in magnitude. Specifically, the organizations in the treatment group generated about 50% less spam than those in the control group before the treatment. On average, the



organizations in the treatment group also have about four times more IP addresses than those in the control group. Both two variables are likely to be correlated with post-treatment outbound spam.

**Table 4 Baseline Comparison**

	Control Mean	Treatment Difference No Controls	Treatment Difference Country Pair FE
	(1)	(2)	(3)
Baseline spam	218439	-106540 (111622)	-152449** (57470)
IP number	140495	647773* (327720)	625859* (277616)
Observations	540	1717	1717

*Notes.* Column 1 contains the average characteristic of the organizations in the control countries. Columns 2 and 3 contain estimates of the average difference in characteristics between the control and treatment organizations, without controls and with controls for country pair fixed effects, respectively. Standard errors are clustered by country and shown in parentheses. \* indicates statistical significance at the 10% level, \*\* at the 5% level, and \*\*\* at the 1% level.

## 7. Information Disclosure Effect

The estimation of the effect of information disclosure is based on comparing the outgoing spam volumes of the treatment group and the control group, according to equations (1) and (2). The results are presented in Table 5. Column (1) displays the results from the basic model in equation (1), where only treatment indicator is included. It shows that although the treatment organizations sent out less spam than the control organizations, the difference is statistically insignificant. However, once we control for country pair fixed effect, the difference becomes significant and also increases in magnitude (Column (2)).

As suggested by Table 1, the treatment organizations significantly differ from the control organizations in terms of baseline spam and number of IP addresses. Therefore, in addition to country pair fixed effect, Column 3 also controls for baseline spam volume and number of IP addresses and contains the main results. It is not surprising that both baseline spam volume and the number of IP addresses significantly affect post-treatment outbound spam. Baseline spam volume is positively correlated with spam volume during the treatment period, indicating the persistence of certain security

vulnerabilities. Unless the subject organization takes efforts to deal with these vulnerabilities, it will be continuously exploited by malicious attackers. The number of IP addresses is found to be negatively correlated with outgoing spam volume, suggesting that large systems tend to have less vulnerability. On the one hand, large systems provide attackers with more opportunities. On the other hand, large systems are likely to invest more in Internet security. Our finding suggests that the later force dominates the former one.

**Table 5 Effect of Information Disclosure**

	(1) Basic model	(2) Basic model +Country pair FE	(3) Basic model + Country pair FE +Controls	(4) Basic model + Country pair FE +Controls
	Spam	Spam	Spam	Ln(Spam)
Constant	121248* (52992)	163213*** (31879)	-1274 (4250)	6.4448*** (0.3441)
Treatment	-81393 (53820)	-103197** (30849)	-17757** (4076)	-2.8197*** (0.3669)
Baseline spam			0.4922*** (0.0160)	0.0000003 (0.0000002)
IP number			-0.0058*** (0.0005)	0.0000002*** (0.0000002)
Significance level using wild bootstrap-t			0.002	0.002
Observations	14255	14255	14248	14248

*Notes.* Column 1 displays the estimate of treatment effect on outgoing spam using the basic model without controls (equation (1)). Column 2 reports the result controlling for country pair fixed effects. Column 3 controls for country pair fixed effects, baseline spam volume, and the number of IP addresses. Column 4 reports the estimate of treatment effect on log transformed outgoing spam controlling for country pair fixed effects, baseline spam volume, and the number of IP addresses. All standard errors are clustered by country and shown in parentheses. Row significance level using bootstrap-t reports the significance level for the estimate of treatment effect in Column 3. \* indicates statistical significance at the 10% level, \*\* at the 5% level, and \*\*\* at the 1% level.

Controlling for baseline spam and number of IP addresses drops the estimate of the treatment effect by approximately 80% from 103,197 to 17,757, suggesting that the baseline spam and number of IP addresses explain a substantial part of variation in post-treatment outbound spam. Nevertheless, the treatment effect remains significant and sizable even with controls for the two characteristics. Given that the average outgoing spam volume for the treatment group is 111,899, the size of this effect is estimated at approximately 15.9%. Romanosky et al. (2011) found that the adoption of data breach disclosure laws

can reduce identity theft caused by data breaches, on average, by 6.1%. Although the two findings are consistent, the comparison suggests that public information disclosure can generate more effective results than notifying only those who have been affected.

Considering the small number of clusters, we use the wild cluster bootstrap-t procedure suggested by Cameron et al. (2008) to further test the treatment effect estimate. The bootstrap result shows that the estimate is robust to such asymptotic refinement. In addition, to test whether the estimate is subject to the functional specification of the statistical model, we take log transformation of spam volume, which smooths out the skewness in the distribution of spam, and run the same estimation again. According to the results presented in column (4), the treatment effect becomes even more significant statistically. Therefore, we can safely arrive at the conclusion that public disclosure of outbound spam does help reduce outbound spam.

To examine how the treatment effect dynamically changes as the treatment proceeds, we run the estimation for each month of the treatment period separately, controlling for country pair fixed effects, baseline spam, and the number of IP addresses. The results for all seven months of the treatment period are presented in Table 6. Throughout the entire period, the estimates of treatment effect are consistent and increase in magnitude, which provides additional support for our conclusion.

**Table 6 Effect of Information Disclosure by Time**

	(1) 1 month	(2) 2 months	(3) 3 months	(4) 4 months	(5) 5 months	(6) 6 months	(7) 7 months
Constant	7429 (6124)	-10282* (4505)	-9821 (5828)	-7308 (5405)	-2276 (8665)	-6402 (7918)	9327* (4323)
Treatment	-15476* (7296)	-10843* (5229)	-14734** (5925)	-26353*** (4257)	-36988** (9620)	-18685 (12003)	-13304** (4541)
Baseline spam	0.3049*** (0.0043)	0.5125*** (0.0140)	0.4725*** (0.0141)	0.5256*** (0.0160)	0.7260*** (0.0332)	0.7520*** (0.0342)	0.3799*** (0.0103)
IP number	0.0012 (0.0012)	-0.0036** (0.0008)	-0.0035** (0.0014)	-0.0045** (0.0019)	-0.0131** (0.0029)	-0.0149** (0.0032)	-0.0051*** (0.0006)
Observations	1717	1717	1717	1717	1717	1717	1717

*Notes.* Columns 1 to 7 display the estimates for the first to seventh month after the treatment. Country pair fixed effects, baseline spam, and number of IP addresses are included in all estimations. Standard errors are clustered by country and shown in parentheses. \* indicates statistical significance at the 10% level, \*\* at the 5% level, and \*\*\* at the 1% level.

**Table 7 Interaction Effect of Information Disclosure**

	(1) No Interaction	(2) Interact with baseline spam	(3) Interact with IP number	(4) Interact with baseline rank	(5) Interact with baseline top10
Constant	-1274 (4250)	-1478 (2363)	-1746 (2945)	-49205* (23181)	-9727 (10995)
D <sub>c</sub>	-17757** (4076)	10656 (9605)	-17091*** (2791)	-86068 (47539)	-4212 (13915)
Baseline spam	0.4922*** (0.0160)	0.5027*** (0.0006)	0.4919*** (0.0166)	0.4942*** (0.0133)	0.4944*** (0.0128)
IP number	-0.0058*** (0.0005)	-0.0002 (0.0007)	-0.0002 (0.0193)	-0.0036** (0.0014)	-0.0049*** (0.0007)
D <sub>c</sub> * Baseline spam		-0.2524** (0.0776)			
D <sub>c</sub> * IP number			-0.0057 (0.0191)		
Baseline rank				244.2** (89.05)	
D <sub>c</sub> * Baseline rank				-336.6 (217.7)	
Baseline top10					7820 (90440)
D <sub>c</sub> * Baseline top10					-294164 (294193)
Observations	14248	14248	14248	14248	14248

*Notes.* Column 1 displays the main results without any interaction effect from column 3 of Table 5 for comparison. Columns 2 and 3 present the results allowing the treatment effect to interact with baseline spam volume and number of IP addresses respectively. Column 4 and 5 present the results allowing the treatment effect to interact with the baseline rank and baseline top10 respectively. Country pair fixed effects, baseline spam, and number of IP addresses are included in all estimations. Standard errors are clustered by country and shown in parentheses. \* indicates statistical significance at the 10% level, \*\* at the 5% level, and \*\*\* at the 1% level.

Then, we allow the treatment effect to interact with pre-treatment characteristics to see whether the effect will differ on different organizations. Table 7 presents the results. Column (1) simply displays the main results from column 3 of Table 5 without any interaction as a benchmark for comparison. Columns (2) and (3) present the results allowing the treatment effect to interact with baseline spam volume and the

number of IP addresses, respectively. Column (4) presents the results allowing the treatment effect to interact with the baseline rank, which is the spam ranking among all organizations in the same country at the time one month before treatment. Column (5) presents the results allowing the treatment effect to interact with baseline top10, which is a binary indicator for whether the organization ranked top 10 among all organizations in the same country at one month before treatment. Country pair fixed effects, baseline spam, and number of IP addresses are included in all columns.

According to these results, the treatment effect does not interact with pre-treatment characteristics except for baseline spam volume. The significant negative coefficient (-0.2524) for  $D_c^*$  Baseline spam shows that information disclosure is more effective on organizations with more baseline spam. Although we listed only the top 20 spamming organizations, organizations currently off the list were also encouraged to take effort to remain that way. However, when public disclosure was imposed, organizations with severe a outgoing spam problem had stronger incentives to deal with the problem to reduce reputation loss. This could have been partially because of the specific presentation we used in the present study, that we disclosed the worst behavior instead of advocating the best practice.

Since the data we used were collected repeatedly for the same sample for many months, the outcome may be serially correlated, and the resulting standard errors may be inconsistent. Besides bootstrap, Bertrand et al. (2004) proposed that the correction that collapses the time series information into a “pre”- and “post”-period can explicitly take into account the effective sample size. Using this method as an additional robustness check, we collapse our data into a pre-treatment and a post-treatment period by taking the average of spam volume for the months before the treatment and the months with treatment for each sample AS. Then we run the statistical model using the collapsed data. The results (in Table 8) are consistent with the results using original data.

**Table 8 Information Disclosure Effect Using Average Spam Volume for Pre- and Post-Treatment Periods**

	(1) Basic model	(2) Basic model +Country pair FE	(3) Basic model + Country pair FE +Controls
	Post-spam	Post-spam	Post-spam
Constant	112267* (53094)	157823*** (34448)	-1694 (4138)
Treatment	-72612 (53993)	-96041** (32421)	-17333*** (4261)
Pre-spam			0.4928*** (0.0158)
IP number			-0.0058*** (0.0006)
Observations	1718	1718	1718

*Notes.* Standard errors are clustered by country and shown in parentheses. \* indicates statistical significance at the 10% level, \*\* at the 5% level, and \*\*\* at the 1% level.

## 8. Social Comparison Effect

As we showed in the previous section, information disclosure does positively encourage organizations to reduce outgoing spam. Although we release information only on the top-most spamming organizations, the disclosure has a positive spillover effect on organizations that are currently off the list. The disclosure effect mainly comes from the concern for reputation loss; that is, the more spam an organization sends out, the more likely it will be listed on SpamRankings.net. Another force that would potentially affect organizations' behavior is the social comparison effect caused by the information of other organizations' behavior. The important information on others' behavior released on our website is the maximum spam volume observed from the most spamming AS (referred to as Max spam) and the minimum listed spam volume observed from the AS that ranked 10th in the treatment country (referred to as *Min spam*). According to social comparison theory (Festinger 1954), organizations will react to the specific information disclosed as well as to the disclosure mechanism in that organizations have the tendency to behave in consistency with others. Therefore, in addition to the treatment variable, we add two variables, Treatment\*Max spam and Treatment\*Min spam, to test whether treatment organizations will react to the

specific information disclosed on other organizations.

Table 9 shows the results after including these two additional variables. The coefficients indicate how the specific information would modify the treatment effect. We find that organizations react only to the specific *Max spam*, not the *Min spam*, meaning that they pay attention only to the worst spam sending behavior but not to the mediocre behavior. The coefficient estimate of Treatment\*Max spam is positive, suggesting that the more outgoing spam observed from the worst behavior, the less likely organizations will be to take effort to improve their own behavior. In other words, it shows that if even the worst behavior is not so bad, organizations will have more pressure or desire to improve themselves. This finding is consistent with the prediction by social comparison process, during which individuals evaluate themselves against others and the existence of a discrepancy leads to actions toward reducing it.

**Table 9 Impact of Specific Information Disclosed**

	(1) Basic model + Country pair FE +Controls	(2) Basic model + Country pair FE +Controls	(3) Basic model + Country pair FE +Controls
	Spam	Spam	Spam
Constant	353.3 (5172)	-44.00 (5041)	375.0 (5271)
Treatment	-24601*** (6320)	-17514** (7266)	-24922** (7375)
Baseline spam	0.4899*** (0.0177)	0.4899*** (0.0177)	0.4899*** (0.0177)
IP number	-0.0066*** (0.0014)	-0.0066*** (0.0014)	-0.0066*** (0.0014)
Treatment*Max spam	0.0002* (0.0001)		0.0002** (0.00008)
Treatment*Min spam		-0.0051 (0.0114)	0.0007 (0.0096)
Observations	14248	14248	14248

Notes. Standard errors are clustered by country and shown in parentheses. \* indicates statistical significance at the 10% level, \*\* at the 5% level, and \*\*\* at the 1% level.

## 9. Conclusion

Governments, businesses, and consumers are constantly exposed to the risk of cybercrime. Our society has recognized the need for additional laws and co-operation 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 (Anderson and Moore 2006). An organization's security vulnerabilities are also shared by other organizations but are often kept private. The negative externality gives Internet security the feature of partial public good. The private provision of public goods 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 social comparison theory, we propose a social information provision to encourage organizations to improve their Internet security. The information disclosure mechanism can incur concern for reputation loss, whereas the specific information disclosed can incur a social comparison. Through making information on other organizations' behavior publicly available, we aimed to solicit the comparison among them and impose reputation loss on those who do not behave pro-socially. To dose up the comparison, we disclosed the relative rankings for all of the organizations in a country in addition to absolute performance and listed the top 10 worst organizations by their standings. Such "shame" lists make bad behaviors notorious. Using a field quasi-experiment on outgoing spam for 1,718 ASes in eight countries, we show that providing social information on outgoing spam encouraged the treatment organizations to reduce it by approximately 15.9%. As compared with an existing study (Romanosky et al. 2011), which documented a 6.1% effect of adopting data breach disclosure laws on identity theft, this result shows that making social information publicly available is more effective than notifying only affected consumers in motivating desirable pro-social behavior.



We find a positive spillover effect in that even though only the top 10 worst organizations are listed, such listing incentivizes both listed and unlisted organizations. However, the impact is stronger on organizations with more spam. The number of IP addresses is found to be negatively correlated with outgoing spam volume. The number of IP addresses measures the size of the AS. On the one hand, large ASes have more incentives to invest in Internet security because of economies of scale. On the other hand, large ASes have more exploitable opportunities for miscreants. The result indicates the former dominates the latter. A closer look at the social comparison process reveals that the more outgoing spam observed from the worst spammer, the less likely an organization will be to improve its own behavior. This finding again reflects that improving Internet security requires collective work of all organizations and that individual behavior can generate strong positive externalities on others. We can utilize the desire for information on others' behavior and use of the information for self-evaluation to intervene in individual organizations' security decisions and solicit the desirable behavior.

Our present study has implications for information architecture design and public policy making. With the ubiquity of Internet, the things that people do online can be tracked, which provides us with an abundance of data. The question we currently face is not the lack of data but how to make use of the data available. Online users care about their popularity, reputation, and social status within the community. If we can capture users' actions, aggregate and display the relevant information, and provide the right feedback as the right intervention at the right time, we can lead their behaviors in our intended direction. With respect to public policy, our present work is among the few empirical studies on Internet security using security vulnerabilities data. Policy makers have hesitated to use security information disclosure for a long time. Although a fierce argument has been observed surrounding disclosure, little attention has been paid to information display or presentation. We believe what is more important than disclosure is whether the information is easy for users to interpret and compare. In the present study, we used relative rankings to enhance the disclosure effect. For policy evaluation, more information presentation methods can be considered and compared before carrying out 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 such as energy conservation and pollution. In the case where data is not available, the legislation that requires mandatory reporting can be employed to collect data.

Our present article is only our first step in studying Internet security and relevant public policy issues from social psychology and economics perspectives. We are planning to further extend the present study in several dimensions. First, we experimented only with ranking information in our study to focus on relative standing. To identify the exact effect of using ranking information versus absolute volume information, a new treatment group can be added by which organizations receive information only on absolute outbound spam volume with organizations listed alphabetically. With the established visibility of SpamRankings.net, we can experiment with more countries, more industries, and more treatment conditions. 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 infosec, affecting both vulnerability to spam and other threats such as phishing. This would mean that broad improvements in infosec can be achieved by presenting public 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 will still ignore other security problems. If this happens, it means that companies instead need to be individually incentivized to make improvements on individual dimensions of security. As a result, we can encourage companies to make anti-spam improvements by releasing social information on spam. However, to encourage companies to prevent phishing, we need to also release phishing information. With phishing data in addition to spam data, we can distinguish these two possibilities by exploring their correlations. In addition, we can drill down outgoing spam to botnets or snowshoe spammers to consider attackers' reactions to information disclosure.

## 10. Acknowledgements and Disclaimer

This material is based upon work supported by the National Science Foundation under Grants No. [1228990](#) and [0831338](#).

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

We also gratefully acknowledge custom data from CBL, PSBL, Team Cymru, the University of Texas Computer Science Department, and Quarterman Creations. None of them are responsible for anything we do, either.

## References

- Akerlof, G. A. 1980. A theory of social custom, of which unemployment may be one consequence. *The Quarterly Journal of Economics* 94(4) 749-775.
- Anderson, R. 2001. Why information security is hard: An economic perspective. *Proceedings of the 17th Annual Computer Security Applications Conference (ACSAC)*, New Orleans, LA.
- Anderson, R., T. Moore. 2006. The economics of information security. *Science* 27(314) 610-613.
- Androutsopoulos, I., J. Koutsias, K. V. Chandrinou, G. Paliouras, C. D. Spyropoulos. 2000. An evaluation of naïve bayesian anti-spam filtering. arXiv preprint cs/0006013. <http://arxiv.org/abs/cs.CL/0006013>.
- Arora, A., R. Krishnan, A. Nandkumar, R. Telang, Y. Yang. 2004a. Impact of vulnerability disclosure and patch availability: An empirical analysis. *Third Workshop on the Economics of Information Security* (24) 1268-1287.
- Arora, A., R. Telang, H. Xu. 2004b. Timing disclosure of software vulnerability for optimal social welfare. *Proceedings of the 3rd Workshop of Economic Information Systems*, Minneapolis, MN 1–47.
- Baker, W.H., Rees, L.R., Tippett, P.S.. 2007. Necessary measures: metric-driven information security risk assessment and decision making. *Communications of the ACM*. 50 (10) 101-106.
- Banerjee, A.V., E. Duflo, R. Glennerster, D. Kothari. 2010. Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives. *BMJ* doi 10.1136/bmj.c2220. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2871989/>
- Bauer, S., D.D. Clark, W. Lehr. 2012. A Data Driven Exploration of Broadband Traffic Issues: Growth, Management, and Policy. *TPRC 2012*. [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2029058](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2029058)

- Beenen, G., K. Ling, X. Wang, K. Chang, D. Frankowski, P. Resnick, R. E. Kraut. 2004. Using social psychology to motivate contributions to online communities. Human-Computer Interaction Institute. Paper 88. <http://repository.cmu.edu/hcii/88>
- Bernheim, B. D. 1994. A theory of conformity. *Journal of Political Economy* 102(5) 841-877.
- 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.
- Butler, B. S. 2001. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research* 12(4) 346-362.
- 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*.
- Cain, D.M., G. Loewenstein, D. A. Moore. 2005. The dirt of coming clean: Perverse effects of disclosing conflicts of interest. *Journal of Legal Studies* (34) 1-25.
- Cameron, A. C., J. B. Gelbach, D. L. Miller. 2008. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics* 90(3) 414-427.
- 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) 431-448.
- Carver, C. A., J. M. D. Hill, J. R. Surdu, U. W. Pooch. 2000. An intrusion response taxonomy and its role in automatic intrusion response. *Proceedings of the first IEEE Workshop on Information Assurance and Security*, Los Alamitos, CA: IEEE Computer Society 129-135.
- Cate, F. 2009. Comparative approaches to security breaches. *Symposium on Security Breach Notification Six Years Later: Lessons Learned about Identity Theft and Directions for the Future*, Berkeley Center for Law and Technology, Berkeley, CA.
- Chen, Y., F. M. Harper, J. Konstan, S. Xin Li. 2010. Social comparisons and contributions to online communities: A field experiment on MovieLens. [\*The American Economic Review\* 100\(4\) 1358-1398.](#)

- Coase, R.H. 1960. The problem of social cost. *Journal of Law and Economics* (3) 1-44.
- Cole, H. L., G. J. Mailath, A. Postlewaite. 2001. Investment and concern for relative position. *Review of Economic Design* 6(2) 241-261.
- Cook, D., J. Hartnett, K. Manderson, J. Scanlan. 2006. Catching spam before it arrives: Domain specific dynamic blacklists. *ACSW Frontiers '06: Proceedings of the 2006 Australasian Workshops on Grid Computing and E-Research* (54) 193–202, Darlinghurst, Australia.
- Commtouch. 2010. *First Ever Outbound Spam Study Finds Service Providers Struggling to Deal with Spam Sent from Their Own Networks*. Commtouch and Osterman. <http://www.commtouch.com/press-releases/first-ever-outbound-spam-study-finds-service-providers-struggling-deal-spam-sent-thei>
- D'Arcy, J., A. Hovav, D. Galletta. 2009. User awareness of security countermeasures and its impact on information systems misuse: A deterrence approach. *Information Systems Research* 20(1) 79-98.
- Dahlman, C.J. 1979. The problem of externality. *Journal of law and economics* 22(1) 141-162.
- Davis, O. A., A. Whinston. 1962. Externalities, welfare, and the theory of games. *The Journal of Political Economy* 70(3) 241-262.
- Dietz, T., E. Ostrom, P.C. Stern. 2003. The Struggle to Govern the Commons. *Science*. 302(5652) 1907-1912.
- Dillenberger D., P. Sadowski. 2012. Ashamed to be selfish. *Theoretical Economics* 7(1) 99-124.
- Direr, A. 2001. Interdependent preferences and aggregate saving. *Annals of Economics and Statistics* (63-64) 297-308.
- Duesenberry, J. 1949. *Income, saving, and the theory of consumer behavior*. Cambridge, MA: Harvard University Press.
- 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.
- Elias, L. 2001. Full disclosure is a necessary evil. SecurityFocus.com, [www.securityfocus.com/news/238](http://www.securityfocus.com/news/238).
- Farrow, R. 2000. The pros and cons of posting vulnerability. *The Network Magazine*,

[www.networkmagazine.com/shared/article](http://www.networkmagazine.com/shared/article).

- Festinger, L. 1954. A theory of social comparison processes. *Human Relations* (7) 117 -140.
- Forcht, K. A. 1994. *Computer Security Management*. Boyd & Fraser, Danvers, MA.
- Frank, R. H. 1985. The demand for unobservable and other nonpositional goods. *American Economic Review* 75(1) 101-16.
- Frei, S., 2010. The security of end-user PCs: An empirical analysis. *DDCSW: Collaborative Data-Driven Security for High Performance Networks*, Internet2 and WUSTL, August.
- Frey, B. S., S. Meier. 2004. Social comparisons and pro-social behavior – testing “conditional cooperation” in a field experiment. *American Economic Review* (94) 1717–1722.
- Fung, A., M. Graham, D. Weil. 2007. *Full Disclosure: The Perils of and Promise of Transparency*. Cambridge, MA: Cambridge University Press.
- Gal-Or, E., A. Ghose. 2005. The economic incentives for sharing security information. *Information Systems Research* 16(2) 186-208.
- Goodman, J., G. V. Cormack, D. Heckerman. 2007. Spam and the ongoing battle for the inbox. *Communications of the ACM* 50(2) 24–33.
- 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.
- Griskevicius, V., J. M. Tybur, B. Van den Bergh. 2010. Going green to be seen: Status, reputation, and conspicuous conservation. *Journal of Personality and Social Psychology* 98(3) 392-404.
- Hopkins, E., T. Kornienko. 2004. Running to keep in the same place: Consumer choice as a game of status. *American Economic Review* 94(4) 1085-1107.
- 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.
- Isacenkova, J., D. Balzarotti. 2011. Measurement and evaluation of a real world deployment of a challenge-response spam filter. *Proceedings of ACM ICM 2011*.
- Jaquith, A. 2007. *Security Metrics: Replacing Fear, Uncertainty, and Doubt*. Addison-Wesley

Professional.

- Jin, G. Z., P. Leslie. 2003. The effect of information on product quality: Evidence from restaurant hygiene grade cards. *Quarterly Journal of Economics* (118) 409–451.
- 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*.
- Kankanhalli, A., H. H. Teo, B. C. Y. Tan, K.-K. Wei. 2003. An integrative study of information systems security effectiveness. *International Journal of Information Management*, 23(2) 139–154.
- 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.
- Karau, S. J., K. D. Williams. 1993. Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, 65(4) 681-706.
- Kotadia, M. 2004. Porn gets spammers past Hotmail, Yahoo barriers. CNET News, May 6. <http://news.cnet.com/2100-1023-5207290.html>.
- 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.
- Krupka, E., R. A. Weber. 2009. The focusing and informational effects of norms on pro-social behavior. *Journal of Economic Psychology* 30(3) 307-320.
- Lenard, T. M., P. H. Rubin, P.H. 2005. Slow down on data security legislation. *Progress Snapshot 1.9*, Washington, DC: Progress & Freedom Foundation.
- Levchenko, K., A. Pitsillidis, N. Chachra, B. Enright, Mark F elegyhazi, C. Grier, T. Halvorson, C. Kanich, C. Kreibich, H. Liu, D. McCoy, N. Weaver, V. Paxson, G. M. Voelker, S. Savage. 2011. Click trajectories: End-to-end analysis of the spam value chain. *IEEE Symposium on Security and Privacy* 2011 431–46.
- List, J. A., D. Lucking-Reiley. 2002. The effects of seed money and refunds on charitable giving:

- Experimental evidence from a university capital campaign. *Journal of Political Economy* 110(1) 215-233.
- 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.
- Ludford, P. J., D. Cosley, D. Frankowski, L. Terveen. 2004, Think different: Increasing online community participation using uniqueness and group dissimilarity. *Proceedings of the SIGCHI conference on Human factors in computing systems* 631-638.
- Luttmer, E. F. P. 2005. Neighbors as negatives: Relative earnings and wellbeing. *The Quarterly Journal of Economics* 120(3) 963–1002'
- Manzano, P. 2009. *ENISA 2009 spam survey: Measures used by providers to reduce spam*. ENISA. <http://www.enisa.europa.eu/act/res/other-areas/anti-spam-measures/studies/spam-survey>
- Messaging Anti-Abuse Working Group (MAAWG). 2011. Email metrics program: The network operator's Perspective. *Report 14*. [http://www.maawg.org/sites/maawg/files/news/MAAWG\\_2010\\_Q3Q4\\_Metrics\\_Report\\_14.pdf](http://www.maawg.org/sites/maawg/files/news/MAAWG_2010_Q3Q4_Metrics_Report_14.pdf).
- Microsoft. 2011. Battling the rustock threat. *Microsoft Security Intelligence Report*, Special Edition.
- Milinski, M., D. Semmann, H-J. Krambeck. 2002. Reputation helps solve the 'tragedy of the commons'. *Nature* 415 424-426.
- Mookerjee, V., R. Mookerjee, A. Bensoussan, W. T. Yu. 2011. When hackers talk: Managing information security under variable attack rates and knowledge dissemination. *Information Systems Research* 22(3) 606-623.
- Moore, T., R. Clayton. 2008. *The Consequence of Non-Cooperation in the Fight Against Phishing*. Third APWG eCrime Researchers Summit.
- Motoyama, M., K. Levchenko, C. Kanich, D. McCoy, G. M. Voelker, S. Savage. 2010. Re: Captchas—understanding captcha-solving services in an economic context. *USENIX Security Symposium* (10).
- Moyer, E. 2011. Breach exposes chase, capital one, TiVo customers. *CNET News*, April 2.



[http://news.cnet.com/8301-1009\\_3-20050068-83/bleach-exposes-chase-capital-one-tivo-customers/](http://news.cnet.com/8301-1009_3-20050068-83/bleach-exposes-chase-capital-one-tivo-customers/).

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

Ostrom, E 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press. ISBN 0521405998.

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

Pollak, R. A. 1976. Interdependent preferences. *American Economic Review* 66(3) 309-320.

Ponemon Institute. 2005. *National Survey on Data Security Breach Notification*. Traverse City, MI.

Postlewaite, A. 1998. The social basis of interdependent preferences. *European Economic Review* 42(3-5), 779-800.

Quarterman, J.S., A.B. Whinston. 2010a. FireEye's Ozdok Botnet Takedown in Spam Blocklists and Volume Observed by IIAR Project, NANOG 48. Austin, TX.

[http://www.nanog.org/meetings/nanog48/presentations/Wednesday/Quarterman\\_light\\_N48.pdf](http://www.nanog.org/meetings/nanog48/presentations/Wednesday/Quarterman_light_N48.pdf)

Quarterman, J.S., A.B. Whinston. 2010b. Economic Incentives for Internet Security through Reputation and Insurance. invitation-only first *APWG and IEEE-SA Roadmapping Session, Toward a Global Public Health Initiative Model for eCrime Response*.

Quarterman, J.S. 2010c. Economic Incentives for Cooperation to Fight Spam. *RIPE Labs*. <http://www.ripelabs.net/Members/jsq/content-cooperation-to-fight-spam>

Quarterman, J. S., A. B. Whinston, S. Sayin, E. V. Kumar, J. Reinikainen, J. Ahlroth. 2010d. Internet cloud layers for economic incentives for Internet security. *RIPE Labs*, <https://labs.ripe.net/Members/jsq/economic-incentives-for-internet-security>.

Quarterman, J.S., A.B. Whinston, S. Sayin, E.V. Kumar, J. Reinikainen, J. Ahlroth. 2010e. Transparency as Incentive for Internet Security: Organizational Layers for Reputation, *RIPE* 61. <http://ripe61.ripe.net/presentations/116-Quarterman-presentation-Rome.pdf>

Quarterman, J.S., A.B. Whinston, S. Sayin, E.V. Kumar, J. Reinikainen, J. Ahlroth. 2010f. ASN Ranking

Correlations Between Spam Blocklists. *RIPE Labs*.

<https://labs.ripe.net/Members/jsq/asn-ranking-correlations-between-spam-blocklist>

Quarterman, J.S., S. Sayin, A.B. Whinston. 2011. Rustock Botnet and ASNs. *Telecommunications Policy Research Conference*. <http://www.spamrankings.net/about/publications/publications/tprc2011/>

Quarterman, J.S., L.L. Linden, Q. Tang, A.B. Whinston. 2012. Reputation as Public Policy for Internet Security. *Telecommunications Policy Research Conference*. <http://www.spamrankings.net/about/publications/publications/tprc2012/>

Ramachandran, A., A. Dasgupta, N. Feamster, K. Weinberger. 2011. Spam or ham? Characterizing and detecting fraudulent 'not spam' reports in web mail systems. *Proceedings of ACM CEAS 2011: 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference*.

Ransbotham, S., S. Mitra. 2009. Choice and chance: A conceptual model of paths to information security compromise. *Information Systems Research* 20(1) 121-139.

Rao, J. M., D. H. Reiley. 2012. The economics of spam. *The Journal of Economic Perspectives* 26(3) 87-110.

Remler, D. K., G. G. Van Ryzin. 2011. *Research methods in practice: Strategies for description and causation*. Sage Publications.

Robson, A. J. 1992. Status, the distribution of wealth, private and social attitudes to risk. *Econometrica*, 60(4) 837-857.

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.

Saito, K. 2011. Role conflict and choice: Shame, temptation, and justifications. Working paper.

Schneier, B. 2002. Computer security: It's the economics, stupid. *Workshop on Economics and Information Security*, University of California, Berkeley, CA.

Schwartz, P., E. Janger. 2007. Notification of data security breaches. *Michigan Law Review* (105) 913-984.

- Shang, J., R. Croson. 2009. A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The Economic Journal* 119(540) 1422-1439.
- Sipior, J. C., B. T. Ward, P. G. Bonner. 2004. Should spam be on the menu? *Communications of the ACM* 47(6) 59-63.
- Stone-Gross, B., T. Holz, G. Stringhini, G. Vigna. 2011. The underground economy of spam: A botmaster's perspective of coordinating large-scale spam campaigns. *USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*.
- Straub, D. W., R. J. Welke. 1998. Coping with systems risk: Security planning models for management decision making. *MIS Quarterly*, 22(4) 441-469.
- Tang, Q., L. Linden, J.S. Quarterman, A.B. Whinston. 2012. Reputation as Public Policy for Internet Security: A Field Quasi-Experiment. 2012. Presented at ICIS 2012.
- Taylor, S. E., M. Lobel. 1989. Social comparison activity under threat: downward evaluation and upward contacts. *Psychological Review* 96(4), 569-575.
- Thonnard, O., M. Dacier. 2011. A strategic analysis of spam botnets operations. *CEAS '11: Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference* 162-171.
- Toth, T., C. Kruegel. 2002. Evaluating the impact of automated intrusion response mechanisms. *Proceedings of the Eighteenth Annual Computer Security Applications Conference*, Los Alamitos, CA.
- van Eeten, M.J.G., J.M. Bauer. 2008. Economics of Malware: Security Decisions, Incentives and Externalities. *DSTI/DOC 1(29)*.
- van Eeten, M., J. M. Bauer, H. Asghari, S. Tabatabaie, D. Rand. 2010. The Role of Internet Service Providers in Botnet Mitigation: An Empirical Analysis Based on Spam Data. *WEIS 2010*.
- van Eeten, M., H. Asghari, J.M. Bauer, S. Tabatabaie. 2011. *Internet Service Providers and Botnet Mitigation: A Fact-Finding Study on the Dutch Market*, Netherlands Ministry of Economic Affairs, Agriculture and Innovation, The Hague, Netherlands.  
<http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2011/01/13/internet-service-provid>

ers-and-botnet-mitigation.html

Yue, W., M. Cakanyildirim. 2007. Intrusion prevention in information systems: Reactive and proactive response. *Journal of Management Information Systems* 24(1) 329–353.