EURASIP Journal on Information Security
a SpringerOpen Journal

RESEARCH Open Access

Cross-country analysis of spambots

Vaibhav Garg[1*], Thomas Koster[2] and Linda Jean Camp[2]

Abstract

Spam is a vector for cybercrime and commonly legally prohibited. Why do certain national jurisdictions produce a higher percentage of spam than others despite its prohibition? Why do some countries have a higher percentage of systems acting as spambots compared to other countries? We begin to answer there questions by conducting a cross-country empirical analysis of economic factors that correlate with the prevalence of spam and associated botnets. The economic factors under consideration are grounded in traditional theories of crime offline, as well as prior research in security economics. We found that more than 50% of spam can be attributed to having originated from merely seven countries, indicating that deterrence through policy is both feasible and economically rational. As expected, higher Internet adoption is correlated with higher percentage of spam from a country. Counterintuitively, Internet adoption is also positively correlated with the percentage of infected machines.

Keywords: Economics; Spam; Botnets; Security; Cybercrime

Introduction

The problem of junk email or spam was recognized as early as 1975 [1,2]. In 2010, Symantec reported that 89% of email messages were spam [3], while 88% of spam activity was attributed to spambots [4]. The existence of spambots is attributed to the existence of insecure software on the production side and lack of patching or adoption of security software, e.g., anti-virus software, by end-users (or consumers). This, however, does not (completely) explain why the percentage of spambots is different across countries. Some countries, such as India, always have a high proportion of spambots, and others, e.g., Sweden, do not. Unlike other bot activities, spambots need to be online for a shorter period of time. Thus, all bots are equally valued for sending spam [3]; bots with poor Internet connectivity, such as those in India, are as valuable as those in Sweden, which are likely to have more persistent connections.

The success of cybercrime is contingent on availability of such (spam) botnets. However, individuals whose systems are being exploited as bots do not typically agree to facilitate criminal enterprise. For example, Anonymous (a hacktivist group) tweaked its ‘voluntary’ botnet software, Low Orbit Ion Canon (LOIC), to trick unsuspecting bystanders into launching a distributed denial-of-service (DDoS) attack on the US Justice Department [5]. Similarly, end-users, who are not technically savvy, may unsuspectingly participate as spambots. Deterrence-based approaches when applied to such unsuspecting naive end-users would be undesirable [6]. It would be unrealistic to expect such users to be responsible for their systems or hold them accountable for illicit activities, i.e., sending spam.

Spam, however, is not merely an annoyance but in fact has a significant financial impact. Spam campaigns are integral to the success of online scams, e.g., phishing, pharmaceutical spam [7]. The annual loss due to phishing, and possible gain to phishers, has been claimed to be as much as $178.1 million dollars a year [8]. Pharmaceutical spam revenues have been approximated to $3.5 million dollars [9], with transactions worth $170 million conducted over several years [10].

To alleviate the incentives for cyber-criminals to engage in illegal enterprise online, defenders endeavor to make attacks more expensive [11]. Simultaneously, criminals can be deterred through legal and regulatory solutions, i.e., prosecution [12] or takedowns by law enforcement [13]. While legal deterrence is promising [14], its impact may be limited [15]. For example, even when cyber-criminals are prosecuted, in the long term, they can move to another jurisdiction that is more forgiving of undesirable behavior. Note that for countries with non-extant legitimate information communication technology

*Correspondence: me@vaibhavgarg.net
[1] Department of Computer Science, Drexel University, Philadelphia, PA 19104, USA
Full list of author information is available at the end of the article

(ICT) market, it may be economically rational to allow cybercrime to persist [16]. For example, in a country like Nigeria, 419 scams result in an increased inflow of capital which corresponds with improved (local) social welfare [17].

Current regulatory solutions have been prosecution based. Laws such as the graduated response or the three strikes law [18] have tried to hold individuals accountable for their systems security [19]. Simultaneously, regulators have targeted the prosecution of cybercriminals and other punitive approaches such as takedowns. These punitive regulations are limited in their scope [16], can lead to collateral damage [20], and may be more expensive than the damages due to criminal activity [21].

A third ancillary approach examines the economic indicators of macro behaviors (e.g., participation in botnets) rather than address the micro incentives to protect/attack individual systems [22]. (In fact, there are limited if any incentives for end-users from protect individual systems. From a rational economics perspective, it is better for an individual to free ride since the security of their system depends on the security investment of others [23]. Simultaneously, investment in security is a certain loss, while the loss associated with a security breach is uncertain. Behavioral economics argues that individuals will choose probable rather than certain losses, even when the expected value of the loss is equal [24].) Thus, we need to examine why the percentage of spambots and associated spam differs on a macro level, i.e., across national jurisdictions.

Eeten et al. [25], for example, found that the number of infected machines is driven primarily by the size of an internet service provider's (ISP's) user base. Our research is distinct is three ways. First, the unit of analysis is country rather than ISPs. Second, Eeten et al. exclusively examined countries that were members of Organization for Economic Co-operation and Development (OECD). Our focus is global rather than strictly European. Third, the data for independent variable used here is compiled from the World Bank database, which is free and publicly available. Simultaneously, it is not a one-time measure. Thus, it allows for future research to use the same factors, where the findings would not be an artifact of the measurement strategies for a specific variable. To the extent that measurement impinges the findings, their impact would be consistent across different studies.

We offer preliminary answers to two research questions. First, we investigated why certain countries send a higher total volume of spam than others. Second, we examined why some countries have a higher percentage of infected machines, i.e., spambots, to support spam. Answering these questions required that this work be theoretically grounded in traditional criminology as well as emergent literature on cybercrime economics. With this grounded, we were able to explore the underlying country-level factors that appear to encourage spam and associated botnet infrastructure.

The practical applications for this research are in the guidance of infrastructure investment, and policy formulation. The cross-country analysis of economic variables that correlate with spam/spambots informs not only previously unexplored avenues for policy but also can illuminate the risks of some policies under consideration. (For example, a promising anti-cybercrime effort has been the German Anti-Botnet initiative [26], which provides end-user technical support. However, it is possible that economic constraints prevent the end-user from implementing the recommendations provided, for example, the purchase of anti-virus software. Then, the success of German Anti-Botnet efforts would be contingent on software subsidies for low income markets in some jurisdictions.) This research can serve as a foundation to determine the potential for more widespread patch availability, i.e., a reduction in the number of systems acting as spambots and consequently overall reduction in spam.

Background

Spam requires users to spend a significant amount of time identifying legitimate emails from unsolicited ones. If the web is an attention span economy, then spam is mass theft [27,28]. In addition to costs for the individual end-user, spam also impinges costs on the society as it is often a vector for cybercrime.

Spam has no easy fixes [29]. On the technical side, the effort has been to automate the process of separating spam from legitimate email. This has essentially become an escalating arms race between spammers and security professionals. Anti-spam technologies range from IP address-based techniques [30] to machine learning approaches [31,32]. However, anti-spam efforts are often overcome by strategic innovations by spammers. For example, botnets have emerged as a response to IP blacklisting. Moore et al. [33] investigated temporal correlation between phishing websites and spam campaigns. They found that fast flux attacks pose the greatest threat. While fast flux-based websites comprise only 3% of the hosts, they account 68% of the spam sent.

Spam, like other cybercrime activities, is profit driven [3,34]. Thus, it is only (economically) rational to send spam when the cost of spamming is lower than the respective profits [35]. Moore et al. [34] argue that unlike crime offline, cybercrime is committed by well-educated individuals who do not have comparable financial opportunities in local markets. They recommend public/private partnerships to share information, for example, regarding phishing websites, to facilitate faster take down. They also suggest that ISPs should be made liable for certain activities if due diligence is not observed. Kanich et al. note that for stand alone retail spam to be profitable, the cost of sending spam should be 20 times cheaper than it currently is [9]. Alternatively, for spam to be profitable, it is must be vertically integrated with the associated scam architecture, i.e., the same individuals running the scam campaigns are also responsible for the associated spam. Economically efficient strategies for takedown are contingent on the analysis of these larger criminal networks as a whole [36].

Given the low marginal revenue, spam is economically sensitive and susceptible to new defenses [37]. Recognizing this, researchers have investigated economic solutions to spam seeking, for example, to raise the cost of sending bulk unsolicited email [38]. For example, CAPTCHAs are used to make the sender prove that they are human. CAPTCHAs, however, can be overcome by using crowdsourced labor markets such as Amazon's Mechanical Turk [39] and its more notorious counterpart Freelancer [40]. Some other techniques to increase the cost of bulk email include proof of work [41] and greylisting [42].

Current economic solutions to both spam in particular and cybercrime in general imagine the attackers to be *homo economicus*, are grounded in microeconomic investigations of individual stakeholder motivation and are thus informed by deterrence theory of crime. Complementary economic insights on a national level, both theoretical [16] as well as empirical [43], have been limited.

Garg et al. [16] proposed an economic model of cybercrime building on the model of smuggling [44]. They assumed certain legitimate networked services to be the smuggled analogue to botnets, in that they both provide bandwidth and computation cycles. Thus, botnets were modeled as an instantiation of eSmuggling. They found that existing illegal markets can act as a prohibitive tariff suppressing the development of legal services. Surprisingly, they found that cybercrime can be welfare increasing in local jurisdictions skewing the incentives for local law enforcement to crack down on such activities, creating a local maximum that can perversely suppress the development of a legitimate market.

Osorio [43] examined the economic factors that drove software copyright infringement. He concluded that copyright violations are a function of access and affordability, being explained by GDP per capita and availability of post sales software support in local markets. A deterrence-based solution to piracy then, such as Stop Online Piracy Act (SOPA) [45], is potentially less effective than Netflix, which allows individuals to participate legitimately [46].

Osorio considered a three-dimensional model [43]: (1) accessibility, (2) affordability, and (3) legal framework. Accessibility was operationalized as the ability of the software to fit local needs, presence of after sales support and corporate presence. Affordability was operationalized as gross domestic product (GDP) per capita. Legal framework was operationalized using the work of Easterly et al. [47]. GDP refers to the market value of all goods and services produced in a country. Osorio's paper empirically examined the theoretical assertions of prior research [48-50].

Osorio's model presumes voluntary participation in illicit activity. However, while this applies to illegal copies of software, this assumption is unlikely to hold for botnets. Facilitators of cybercrime are frequently naive end-users, whose systems have been hijacked, often to support activities of which they do not approve [5]. While these cybercrime activities such as spam and phishing campaigns are short lived, the same is not true for the infrastructure, such as botnets, that supports these activities. What factors facilitate the presence of spambots and associated spam in a country? In this paper, we begin to answer this question by a cross-country examination of the underlying economic variables. We discuss data and methodology in the section immediately following.

Methodology

In this paper, we conduct a cross-country empirical analysis of the economic factors that correlate with and appear to encourage the percentage of spam and associated bots. We implement an ordinary least squared (OLS)-based linear regression analysis, using independent variables that are grounded in traditional theories of crime offline [51] as well as prior research in economics of cybercrime [43]. We consider two dependent variables. First, we investigate the total amount of spam originating from a country as a percentage of total amount of spam received. Second, we examine the number of infected systems in a country as a percentage of total number of spambots.

Independent variables

The independent variables under consideration are grounded in traditional theories of crime offline, specifically: (1) routine activity theory, (2) economic deprivation theory, and (3) social support/altruism theory. These variables have all been operationalized using the publicly available data from the World Bank. The World Bank database provides a consistent measure of country-level economic variables. To the extent that there are measurement errors, the mistakes should be unbiased, evenly distributed, and thus not effect final results (i.e., noise). This database is widely used in economic research [52]. Table 1 lists all the independent variables along with the corresponding year.

Routine activity theory of crime considers crime to be a function of motivated offenders[a], available targets, and absence of guardianship [53]. For spambots, available targets are the vulnerable systems which could be exploited as spambots. A larger population of Internet users creates more potentially vulnerable machines. Thus,

Table 1 Five-dimensional regression model

Model variables	Description	Year
Availability (AVA)	Fixed broadband subscribers	2010
	Fixed broadband subscribers (per100 people)	2010
	Internet users (per 100 people)	2010
Security of ICT infrastructure (SEC)	Secure internet servers	2010
	Secure internet servers (per one million people)	2010
Economic resources or affordability (ECO)	GDP per capita	2010
	GDP per capita by PPP	2010
Governance or legal framework (LEG)	Government effectiveness	2010
	Regulatory quality	2010
	Rule of law	2010
	Control of corruption	2010
Security skills or education (EDU)	Computer, comm., and other services (% imports)	2010
	Computer, comm., and other services (% exports)	2010

we consider the number of Internet users and the number of fixed broadband subscribers. (We use both Internet users and fixed Internet broadband subscribers to account for the difference in resources. For example, broadband subscribers are likely to have higher bandwidth than those that use dial up modem. Simultaneously, broadband subscribers would likely be online more often.) To account for the differences in the proportion of population online, we also included measures of the number of fixed Internet broadband subscribers and Internet users per 100 people in the analysis.

Further, we consider the security of the existing ICT infrastructure, as a variable that measures guardianship. This is operationalized as the number of secure Internet servers (SIS) and SIS per million people. Secure Internet servers is defined by the World Bank as ‘servers using encryption technology in Internet transactions[b]’.

The resilience of the associated infrastructure has been shown to impinge the volume of spam and the percentage of infected systems [54]. (To account for the difference in population, we also consider the number of secure Internet servers per one million people.) Admittedly, this term is vaguely defined. However, it does allow a uniform, consistently available measure that is likely to be repeated over time by the World Bank. This will allow other researchers to reproduce our work and conduct broader empirical examinations of Internet readiness, cybercrime, etc. A more perfect one-time measurement is almost certainly possible, but it would prevent this work from being subject to replication or later repetition and as such be less of a contribution to the science of cybersecurity.

Economic deprivation theory of crime argues that individual participation in crime may be driven by absolute [55] or relative economic deprivation [56]. The lack of economic resources limits the ability of the individual to participate legally in the market. For example, individuals with limited resources may not be able to pay for licensed copies of software, thereby (often) blocking their access to timely security updates and software patches. Alternatively, limited resources would impinge the end-users’ ability to purchase protection services such as anti-virus, making them more vulnerable. Thus, we consider GDP per capita and GDP per capita by purchasing power parity (PPP). Given that GDP is used as a measure of economic resources available to a nation, GDP per capita corresponds to the absolute deprivation of the individual in the country while GDP per capita by PPP indicates the relative deprivation with respect to other countries. Economic deprivation is similar to affordability under Osorio’s [43] framework.

The impact of resource deprivation can be alleviated by *social support and altruism* [57]. *Social support* is provided by the government by conditions conducive to adoption of security technologies, for example, through direct and indirect subsidies. An example of a indirect security subsidy is the German Anti-Botnet initiative [26], where offending system owners are informed of malware on their systems as well as advised on how to address the infection. A direct subsidy can be in the form of NSA secure linux in the USA, whereby individuals have a clear signal in the market for secure software as well as free access to it.

Then, better governance should lead to more mature local ICT markets, providing better and cheaper access to ICT technologies and creating an indirect subsidy for the end-user. We operationalize social support using a subset of World Governance Indicators (WGI) [58]: (1) government effectiveness, (2) regulatory quality, (3) rule of law, and (4) control of corruption, i.e., perception of corruption within a country. Government effectiveness measures the perceived quality of public services, quality of civil service, and the degree to which it is independent from political manipulation, the quality of policy formulation and implementation, and the perceived credibility of the government to commit to said policies. Regulatory quality quantifies the perceived ability of the government toward sound policy and the degree to which regulations formulation and implementation encourage private sector development. Rule of law indicates the degree to which the legal framework is implemented. The legal framework can also be thwarted by corruption or perceptions thereof. Control of corruption measures perceptions of corruption, where corruption is defined as misuse of public power for private gain. Governance is similar to legal framework under Osorio's [43] framework.

In addition to social support through public bodies, we also consider *altruism* through private bodies. A developed ICT market should be more invested in protecting its resources. For example, it may be cheaper for the ISPs to proactively protect their networks than provide customer support related to security issues [59]. It would then be rational for ISPs to invest in detection of malware on their networks and actively engage users to clean their machines.

Simultaneously, a bigger ICT market, for example, would result in a larger number of staff personnel who have been trained in basic security practices to address the information security needs of that organization. This information would be relevant when the end-user is using their home machines as well as those at work. Some companies in fact provide access to anti-virus software for home computers as employees often take work home. Thus, we consider the size of the ICT market as a proxy for private altruism. This is operationalized by considering the percentage export and import of computer, communications, and other services. Thus, this variable is used as a proxy to the technical security skills available to the market as a whole[c].

Datasets

Recall we had two research questions. Why do certain national jurisdictions produce a higher percentage of spam than other? Why do some countries have a higher percentage of systems acting as spambots than others? Consequently, we examine two dependent variables and two datasets.

The first dependent variable corresponds to the number of spam emails that appear to originate from a specific geographic location. The data for this variable was procured from an academic source, i.e., computer science servers of Indiana University, Bloomington (in the USA). These servers are primarily used by faculty and staff in computer science. The strategy behind the data collection has been detailed in the Appendix along with the assumptions.

The second dependent variable corresponds to the percentage of infected systems acting as spambots in individual countries. The data corresponding to this variable was obtained from Microsoft's Security Intelligence Report 2011 (Volume 11) [60]. The spambot data is generated by Microsoft's Forefront Online Protection for Exchange (FOPE), which uses a two-stage filter. The first uses a reputation-based filtering at the network edge. The second uses content-based rules and detects issues such as malicious email attachments. The report provides data for the first and the second quarters of 2011. The list is limited to the top 80 countries that host at least 0.1% of the IP addresses used by spambots.

Clearly, we cannot assert if spambot data does or does not represent botnet owners and controllers. However, our research only examines why spambots occur more frequently in certain countries than others. We do not address if command and control centers for such botnets are also endemic to national jurisdictions with specific economic properties. As the focus here is upon the involuntary participation, the issue of botnet control is orthogonal.

Data analysis

We examine two regression Equations 1 and 2, which correspond to the two research questions. For Equation 1, N1 refers to the amount of spam in different countries as a percentage of the total spam volume. The data for dependent variable in this equation is that from Indiana University (Bloomington). For Equation 2, N2 refers to the percentage of infected machines that act as spambots in distinct national jurisdictions. The data for N2 is that from Microsoft. Both Equations 1 and 2 were evaluated using OLS regression, and thus were examined for the underlying assumptions of (the absence) of multicollinearity and heteroskedasticity.

$$\begin{aligned} N1 = {} & \epsilon_1 + \beta_{11} * \mathrm{AVA} + \beta_{12} * \mathrm{SEC} + \beta_{13} * \mathrm{ECO} \\ & + \beta_{14} * \mathrm{LEG} + \beta_{15} * \mathrm{EDU} \end{aligned} \tag{1}$$

$$\begin{aligned} N2 = {} & \epsilon_2 + \beta_{21} * \mathrm{AVA} + \beta_{22} * \mathrm{SEC} + \beta_{23} * \mathrm{ECO} \\ & + \beta_{24} * \mathrm{LEG} + \beta_{25} * \mathrm{EDU} \end{aligned} \tag{2}$$

We began by calculating the variance inflation factor (VIF) to discover and address the presence of multicollinearity in the regression model[d]. The four aspects of WGI were significantly collinear, i.e., VIF>5. Thus, we combined the four governance factors into one by adding all four and called it WGI. Similarly, GDP per capita as well as GDP per capita by PPP were also significantly collinear, i.e., VIF>5. We excluded GDP per capita, while GDP per capita by PPP was retained as an indicator of relative deprivation. Number of fixed broadband Internet subscribers was significantly collinear with the number of Internet users, i.e., VIF>5. Since number of fixed broadband Internet subscribers had less number of missing values, we retained it variable in the model instead of the number of Internet users. For the remaining model VIF values did not indicate strong multicollinearity, i.e., VIF<5.

We also examined the model for heteroskedasticity[e]. We plotted the residuals for the model as a histogram. We also computed the Shapiro test to examine whether the residuals were normally distributed. For Equation 1, the *p* value for the test was much less than 0.001, i.e., the evidence indicates heteroskedasticity. For Equation 2, the p value was 0.715, i.e., the null hypothesis cannot be rejected or there is not enough evidence for heteroskedasticity. For the first model, we then used the White-Huber method to generate heteroskedasticity-corrected covariance matrices.

To facilitate the regression analyses, we transformed both the dependent and the independent variables. Some independent variables, e.g., GDP per capita by PPP, had a wide range and thus the regression will be dominated by the size effect. For such variables, we log transformed the data. Appropriate variables were identified by noticing the presence of outliers in the box plots. Independent variables log transformed were GDP per capita by PPP, number of fixed broadband Internet subscribers, secure Internet servers, and secure Internet servers (per million people).

Given that OLS regression is a parametric test, it makes additional assumptions regarding the dependent variable; specifically, OLS assumes that the dependent variable is continuous and normally distributed. Recall we had two dependent variables: (1) volume of spam and (2) percentage of spambots. The counts for spam volume were converted to percentages, i.e., we divided the amount of spam from a country by the total volume of spam (from all countries in this dataset). The corresponding histogram as well as the Shapiro test did not indicate that the data was normally distributed; p value for the Shapiro test was ≈ 0. Simultaneously, the box plot indicated several outliers. Thus, this dependent variable was also log transformed.

The second dependent variable was the percentage of spambots within a country for different countries. The normality assumption was not satisfied, either by eyeballing the histogram, or by the Shapiro test; p value for the Shapiro test was $<<0.001$. The box plot again indicated several outliers. Thus, this dependent variable was log transformed.

Results

Tables 2 and 3 presents the summary statistics for all the dependent and independent variables. The spam dataset from Indiana University (Bloomington) was highly correlated with the spambot dataset from Microsoft; cor = 0.87, p value $<<0.001$, n = 79 (where n is the number of countries compared). The correlations between the dependent variables and independent variables is given in Table 4[f]. OLS was applied to Equations 1 and 2[g]. The results are given in Tables 5 and 6; the effective sample sizes were 117 and 69, respectively.

Discussion

The relative volume of spam is highly skewed in its distribution. The top seven countries accounted for 51.53% of the total volume of spam received by Indiana University (Bloomington), indicating that most of the offending bots are concentrated jurisdictionally; these seven countries were India, Russian Federation, USA, Vietnam, Indonesia, Brazil, and China, respectively, with India accounting for the largest volume of spam. Previous research noted this

Table 2 Summary statistics: without transformation

	Mean	**Standard deviation**
Dependent variable		
Indiana University*	38,510.0	106,352.2
Microsoft*	2.453	3.83
Independent variable		
GDP per capita by PPP*	14,017.2	15,117.01
Fixed broadband Internet subscribers (FBIS)*	2,859,683	11,877,249
Fixed broadband Internet subscribers (per 100 people)	9.96	12.14
Internet users (per 100 people)	35.47	28.00
Secure Internet servers*	5,560.237	34,338.54
Secure Internet servers (per one million people)*	367.6445	1,151.99
Computer, comm., and other services (% exports)	29.08	19.49
Computer, comm., and other services (% imports)	30.81	16.18
World governance indicators (WGI)	305.7658	158.0306

Asterisk (*) indicates that these variables have *not* been transformed.

Table 3 Summary statistics

	Mean	**Standard deviation**
Dependent Variable		
Indiana University*	-3.29	2.76
Microsoft*	2.45	3.83
Independent Variable		
GDP per capita by PPP*	8.90	1.25
Fixed broadband Internet subscribers (FBIS)*	11.45	3.14
Fixed broadband Internet subscribers (per 100 people)	9.96	12.14
Internet users (per 100 people)	35.47	28.00
Secure Internet servers*	4.62	2.84
Secure Internet servers (per one million people)*	3.09	2.90
Computer, comm., and other services (% exports)	29.08	19.49
Computer, comm., and other services (% imports)	30.81	16.18
World governance indicators (WGI)	205.06	111.38

Asterisk (*) indicates that these variables have been transformed as described in the text.

Table 4 Correlations between spam/spambots and economic factors

Economic variable	Indiana University (*n*)	Microsoft (*n*)
GDP per capita by PPP	0.46*** (145)	0.15 (75)
Fixed broadband Internet subscribers (FBIS)	0.84*** (164)	0.68*** (77)
Fixed broadband Internet subscribers (per 100 people)	0.39*** (164)	0.17 (77)
Internet users (per 100 people)	0.43*** (156)	0.12 (77)
Secure Internet servers	0.74*** (172)	0.46*** (78)
Secure Internet servers (per one million people)	0.23** (172)	0.08 (78)
Computer, comm, and other services (% exports)	0.29*** (141)	0.29* (73)
Computer, comm, and other services (% imports)	0.28*** (141)	0.27* (73)
World governance indicators (WGI)	0.29*** (168)	0.08 (79)

Significant codes: '***' <0.001, '**' <0.01, '*' <0.05.

concentration even for ISPs, i.e., most of the spambots are concentrated to the handful of ISPs. Simultaneously, such ISPs were popular, well established, and thus potentially susceptible to regulatory pressure. Given that most spam is concentrated in a handful of countries, regulatory solutions to spam then appears to be a tangible and tractable option; only a small subset of countries would need to agree to regulate an admittedly larger but still a relatively malleable set of ISPs.

Table 5 OLS regression model: Indiana University [Spam]

Linear regression	Estimate	Standard error	Pr (> |*t*|)
(Intercept)	-11.657	2.622	≈0***
GDP per capita by PPP	0.059	0.403	0.883
Fixed broadband Internet subscribers	0.705	0.172	≈0***
Fixed broadband Internet subscribers (per 100 people)	-0.051	0.033	0.123
Internet users (per 100 people)	0.002	0.016	0.918
Secure Internet servers	0.242	0.212	0.257
Secure Internet servers (per one million people)	0.208	0.161	0.201
Computer, comm., and other services (% exports)	-0.002	0.010	0.835
Computer, comm., and other services (% imports)	-0.006	0.014	0.658
World governance indicators (WGI)	-0.007	0.003	0.0137 *

Significant codes: '***' <0.001, '**' <0.01, '*' <0.05; residual standard error, 1.327 on 108 degrees of freedom; multiple R-squared, 0.7649; adjusted R-squared, 0.7453; F-statistic, 39.03 on 9 and 108 DF; *p* value, <2.2e-16.

Table 6 OLS regression model: Microsoft [Spambots]

Linear regression	Estimate	Standard error	Pr(> |*t*|)
(Intercept)	-4.057	2.592	0.123
GDP per capita by PPP	-0.133	0.392	0.736
Fixed broadband Internet subscribers	0.255	0.125	0.046 *
Fixed broadband Internet subscribers (per 100 people)	-0.019	0.025	0.469
Internet users (per 100 people)	-0.016	0.013	0.210
Secure Internet servers	0.388	0.146	0.010 *
Secure Internet servers (per one million people)	-0.011	0.198	0.954
Computer, comm., and other services (% exports)	0.002	0.008	0.742
Computer, comm., and other services (% imports)	0.018	0.010	0.062
World governance indicators (WGI)	-0.002	0.002	0.295

Significant codes: '***' <0.001, '**'< 0.01, '*' <0.05; residual standard error, 0.8288 on 60 degrees of freedom; multiple R-squared, 0.6057; adjusted R-squared, 0.5466; F-statistic, 10.24 on 9 and 60 DF; *p* value, 2.284e-09.

Overall the evidence in the paper indicates that the prevalence of both spam and spambots is best explained by routine activity theory of crime [53], which considers crime as a function of motivated offenders, available targets, and lack of guardianship. We tested two of these variables, i.e., available targets and guardianship, both of which were important in predicting the volume of spam; the third variable, motivated offenders, is not relevant online as all targets can be considered proximal and thus appropriate for infection[h]. Unlike crime offline, spam and spambots appear to increase with the availability of guardianship. Therefore, previous policy prescriptions grounded in routine activity theory may not be directly applicable online. However, the exploration and translation of such prescriptions from offline to online does suggest a first step in providing potential solutions that can complement regulatory efforts grounded in deterrence.

Availability

We find that all measures of availability, number of users connecting to the Internet, were directly and statistically significantly correlated with the total volume of spam (Table 4). Intuitively, as the number of users increases, so would the number of individual systems and email accounts that can then be used to send out spam. Relative volume of spam is driven by the number of fixed broadband Internet subscribers from a country, as it is the only measure of availability that is statistically significant in the regression analysis (Table 5). In previous research, Eeten et al. [25] similarly noted that the total volume of spam from an ISP was driven by the size of its user base.

The percentage of systems acting as spambots in a country is, however, not statistically correlated with all measures of availability. In fact, only the total number of fixed broadband Internet subscribers (FBIS) was statistically correlated with the percentage of spambots (Table 4). FBIS was also the only statistically significant measure of availability in the regression model (Table 6). This indicates that higher adoption may lead to higher percentage of spambots. This is contrary to previous work by Eeten et al. [25], who found that while bigger ISPs might do worse in total volume of spam, they perform marginally better in terms of percentages. ISPs and national economies may then differ in this respect. Alternatively, the difference might be an artifact of the countries under analysis; Eeten et al. [25] concentrated on OECD countries.

There are several potential explanations for why Internet adoption may increase the percentage of spambots. The foremost and simplest explanation is that increased Internet adoption may simply mean more number of people clicking on links and navigating to different websites, increasing their risk exposure. Then, higher Internet adoption would not just indicate more spambots and spam, but also more malware infections and bots in general. This explanation can be tested by examining the correlations between measures of availability and malware-infected machines and zombies in future studies using other datasets. In concurrent work, we find evidence that suggests that this hypothesis may not be accurate, as number of malware infected machines was negatively correlated with FBIS [61].

A second explanation is that early adopters are typically those who are interested in new technologies, are more technically literate, and have higher education and income. They would be more aware of security risks and more capable of risk mitigation. However, as Internet adoption progresses, systems would be available less informed and economically constrained individuals who would then be limited in their incentives and ability to protect their systems. This explanation is contingent on two variables: education and income (which is highly correlated with education [62]). Education can be examined by using country-level measures of individual literacy rates as well as community-based measures, such as public spending on education. Correspondingly, there would be distinct policy implications[i].

A third potential but tenuous hypothesis may be that when Internet adoption is low, those that adopt technologies are more homogenous. To the extent that security awareness and corresponding mitigation strategies are contingent on stories exchanged in communities [63], the exposure of a homogenous community would be lower. While it is difficult to test this relationship on a country level, it may be possible by looking at indicators of income inequality, e.g., GINI index.

A fourth possibility is the higher penetration allows faster spread of infections as epidemiological models are driven by concentrations [54]. However, there is limited evidence for this. None of the measures for concentration, e.g., Internet users per 100 people, were correlated with the percentage of spambots in a country. However, it is difficult to access the epidemiological impact of malware spread as this is a static model with static independent variables. It may be better to consider the rate of change in the number of Internet users for individual countries. The optimal solution would be to conduct a time series analysis on historical spam data. Such data is, however, not currently available to the researchers.

Guardianship

Security (or guardianship), i.e., secure Internet server (SIS), was directly correlated with both the relative volume of spam from a country as well as the percentage of spambots in individual countries (Table 4). These correlations were statistically significant. However, SIS was not significant in the regression model for the relative volume for spam (Table 5). It may be that the relationship between SIS and relative volumes of spam is not linear as Spearman's coefficient indicates both linear and nonlinear correlations. SIS was, however, significant in explaining the percentage of spambots in individual countries (Table 6).

A first explanation for this is that as the number of users in the local market increases, the number of SIS would consequently have to increase. Thus, more SIS is simply an indicator of Internet adoption. As noted earlier in this section, Internet adoption may lead to more infections. Thus, as the Internet grows not only does the volume of spam increases so does the percentage of spambots. It could be argued that SIS do not need to be jurisdictionally co-located with their target market. However, given that SIS is highly correlated with all measures of availability (p value $<<0.001$), this hypothesis has limited support.

A second explanation then could be that the security of the SIS itself is broken. Personnel in charge may simply be following 'best practices' to secure such servers, which is security often means 'common practices' rather than indicate a measure of quality. It may then be that due to inadequate security, these servers may themselves have become vector for spambot-related malware infections. From a rational choice perspective, trusted systems would attract more attacks as the return on investment would be higher. Problems with SSL implementations are well documented [64]; this could easily extend to other encryption implementations on such servers. However, concurrent research indicates that more SIS do not lead to higher rate of malware infections in general [61]. Thus, this hypothesis currently does not have support.

Economic deprivation

The sole indicator of economic resources, i.e., GDP per capita by PPP, was directly and significantly correlated with the relative volume of spam. However, the correlation with percentage of spambots was not significant. GDP per capita by PPP was also not significant in either of the regression models. Overall, there seems to be limited evidence for the influence of economic resources on prevalence or spam or associated botnets. The evidence that is available indicates a positive relationship; both the correlations are positive and so are the signs for the estimates in both regression models. This is counterintuitive based on previous research. To the extent that economic resources constrain individual ability to purchase legal copies of software [43], higher rates of software piracy should be positively correlated with spam [25]. Similarly, if limited expendable income impinges, the individual ability to purchase security technologies, e.g., anti-virus software, higher GDP per capita by PPP should also lower spam and spambots.

There are two possible explanations for this counterintuitive observation. The simplest explanation is that to the extent that economic resources impact Internet adoption, GDP per capita by PPP also indicates the individual ability to participate in the market as an Internet user. As noted earlier in this section, adoption may lead to higher volume of spam and percentage of spambots. A second, more tenuous, behavioral explanation is that as GDP per capita by PPP increases, adoption of legal software as well as anti-malware technologies is proportionally impinged. However, individuals compensate for risk mitigating technologies by demonstrating higher risk behavior. There is evidence for such static risk budgets offline. For example, the introduction of ABS did not reduce overall risk as drivers compensated by driving closer to other vehicles [65]. Both these hypotheses can be tested by replacing GDP per capita by PPP with more direct measures of economic resources such rates of software copyright infringement and market penetration of anti-virus software. While statistics for copyright infringement are readily available from Business Software Alliance, similar data for market penetration of security technologies is admittedly harder to acquire.

Social support/altruism

Indicator of legal framework and overall governance, i.e., World Governance Indicator (WGI), was significantly and positively correlated with the volume of spam (though not with the percentage of spambots) (Table 4). On the ISP level, better specific governance initiatives have been found to be correlated with lower levels of botnet activity [25]; however, the specific independent variables under consideration were different. WGI was also significant in the regression model for the relative percentage of spam from individual countries (Table 5). To the extent that better governance creates a subsidy for individual end-user consumption of ICT technologies by facilitating the evolution of local ICT markets, this subsidy may not extend to individual investment in security. This is not unexpected, as from a macro-behavioral perspective the security market suffers from clear signals and is thus a market of lemons [66].

Both proxies for technical skills were positively and significantly correlated with both the relative volume of spam as well as the percentage of spambots. These measures do not directly measure either individual literacy rates or technical education in general; instead, they indicate the amount of technical skills that are available to the economy as a whole. Technical skills then indicate a weak relationship with lowering the incidence of spam or spambot, as they were not significant in either of the regression models. The statistical significance of their correlations with the volume of spam and percentage of spambots can be explained as a function of Internet adoption. The relationship between (export and imports of) computer, communications, and other services and Internet adoption is obvious; more exports and imports of ICT services should indicate higher Internet adoption in the local market.

Conclusions

Spam is concentrated, both jurisdictionally and on the network. Merely seven countries appear to generate more than 50% of spam. Similarly, most spam is concentrated to a handful of ISPs. We argue that this allows the possibility of deterrence through regulation. Appropriate legal incentives in less than ten countries would address more than half the spam in the world. Simultaneously, the enforcement of such legislation would be relatively cheap, as only a handful of actors (roughly 50 ISPs) would need to be monitored. Absent the influence of economic deprivation on the prevalence of spam/spambots, there is no argument for software subsidies in low-income markets as a solution to botnets. In fact, the problem of spam seems to be a macro level lack of governance; simultaneously, spambots are impinged by broader infrastructure management issues as those with secure Internet servers. Governance efforts such as the German Botnet Initiative have been promising. Simultaneously, minimum security can be mandated for network providers such as secure Internet servers. While specific policy prescriptions are beyond the scope of this paper, in concurrent research, we argue for ex-ante regulations for addressing spambots [20]. Future research should address the relative costs of policy prescriptions as public or private interventions vs. a common-pool regime.

Of the different theories of criminology tested in this paper, routine activity theory was the most significant driver of spam and associated botnets. As expected as

Internet adoption increases, the total volume of spam goes up. Counterintuitively, Internet adoption also increases the percentage of offending machines. The likely explanation is that investments in Internet adoption are not proportionally matched by those in education. Lack of education impinge individuals in two ways. First, they are limited in their awareness of security risks. Second, even when they are aware, they may not have enough information to adequately assess the risk and construct an appropriate response.

Unlike in traditional routine activity theory, where crime is driven by lack of guardianship, spam and associated bots are instead correlated with the presence of secure Internet servers. This difference does not allow us to directly transfer policy insights offline to countering spam. However, it does provide an alternative perspective to the problem and allow us to develop a framework for systematic inquiry toward engendering long-term public policy and technical solutions.

In this paper, we addressed two research questions. First, what economic factors explain the variance in the relative volume of spam seen from different countries. Second, what economic factors explain the variance the percentage of systems acting as spambots across different nations. We find that large volumes of spam and/or spambots are correlated with higher rates of Internet adoption and presence of secure Internet servers. Our research is theoretically grounded in previous investigations of security economics online as well as criminology offline. While some of our results reify previous microeconomic investigations, others are contradictory. Thus, research investigating individual markets may not be generalizable national economies.

Future work includes repeating this analysis over time to evaluate if changes in the population of spambots are correlated with economic changes; this would be helpful in establishing a causal link rather than just a correlational one. Further, our hope is to collaborate with additional institutions who might share their data. In particular, institutional data sets from corporations and non-US institutions are ideal tests of this model.

Endnotes

[a] In our analysis we do not consider the presence of motivated offenders. Offline motivated offenders is given by proximity to the target. Online, however, all attackers are proximal. It may be that attackers prefer to attack systems that are fewer hops away; however, we do not know of any evidence that indicates support for this hypothesis online. Arguably, it is possible to compute a distance metric for the average distance between two systems in two different countries. However, given that we do not have information on the command and control center for the spambots in out data set, even if such distance metric was available, it could not possibly be applied to this data.

[b] See http://data.worldbank.org/indicator/IT.NET.SECR.P6.

[c] Eeten et al. [25] used education as a proxy variable for technical skills. Unfortunately, the World Bank data for education was sparsely populated. Then, using education as a proxy variable would have left out several countries from our analyses. Thus, we decided on a different proxy variable. Note that the size of the ICT market does not encapsulate individual security education, but instead provides a measurement for security skills available to the market as a whole.

[d] Multicollinearity is observed when two or more independent variables are highly correlated, i.e., at least one of the independent variables can be computed as a linear combination of the rest to a statistically significant degree. If the independent variables are multicollinear, then the results for the ordinary least squared regression may be computed incorrectly.

[e] Heteroskedasticity is observed when the residuals for a model are not normally distributed. This happens when the variance of a variable does not increase or decrease proportionally with respect to a second variable.

[f] Since we use Spearman's rank correlation coefficient (which is sensitive to outliers), the variables were transformed as described above before computing the specific values. Note that we do not use Pearson's coefficient as many of the independent variables are not normally distributed. Furthermore, since log transformation is order preserving, it does not impact Spearman. However, log transformation allows us to have a data set without outliers.

[g] Note that while the independent variables are not normally distributed, OLS is still applicable as the sample size is large [67].

[h] For example, previous research has noted that most botnets include bots that are spread over 30 to a 100 countries [68]. Simultaneously, for spam, the associated bots do not require extra bandwidth or longer uptime, it is unlikely that bots in one country would be preferred over another [3].

[i] Eeten et al. [25] note that higher education levels did in fact indicate lower rates of spam. However, as noted by other findings in this paper, results from ISPs may not directly apply to national economies or those from OECD to countries globally.

Appendix

Indiana University (Bloomington) data collection strategy

The spam data spans between 25 July 2010 and 27 March 2011 ($\approx$ 8 months). There were 9.7 million messages. Of these, 6.0 million were classified as spam and 3.7 million were deemed legitimate. From this data set, we have ≈ 62% spam which is below the upper bound given by Symantec.

All email messages are subjected to two spam filters. The primary filter is an IP address based blacklist. This filter provides a binary output as either an 'OK,' which results in the message being exposed to the second filter or a 'REJECT' at which time it is discarded and deemed spam. Additionally, the output of the first filter also provides a time and date stamp as well as an IP address. The second filter subjects the email to several spam classifiers. The email is accepted if the message is internal, has less than a 60% probability of being spam, and is on a whitelist or department supported mailing list, or a virus is successfully removed. Spam messages are accepted but considered to be spam if there is a greater than 60%, but less than 99%, probability of being spam. Messages may be accepted with a warning notification regarding suspicious attachments. Messages are rejected if they have a virus, an illegal attachment type, a greater than 99% probability of being spam or other instantiation of a virus. For our analysis, we combined messages that were accepted spam, rejected spam, as well as messages regarding viruses in one spam category.

We make three assumptions. First, we assume that spam and only spam are caught by the two filter systems, i.e., no spam gets let through accidentally and legitimate messages do get tossed away accidentally. Second, we assume that IP addresses and hostnames correspond to spambots and are not spoofed. Third, we assume that all the spam was sent by spambots.

Additionally, there were messages aborted during reception due to a dropped connection or invalid recipient addresses. It is unknown if these were blanket spam attacks that put in incorrect addresses, people accidentally mistyping emails, or people/organizations that had not updated their email lists and were emailing people no longer at Indiana University (Bloomington) and whose email addresses were no longer valid. Thus, these messages have not been included in the current analysis.

There is also a 'none' status in the second filter, which was used twice in the 8 months, but it is unknown what this status meant. These messages were thrown out, as 2 messages out of 9.7 million are insignificant.

Each line of the spam data from the second filter contained a unique identifier, stating whether a message is local, (from) Indiana (University), or external, a time and date stamp, anonymized 'from' and 'to' tags, the result of the analysis of the email, filters, the percent chance that it is spam (if it does not hit one of the automatic exceptions, e.g., the whitelist), and the relay hostname or IP address. We used an IP lookup database with a 95% accuracy rate to determine the spam 'messages' country of origin. For the secondary messages without IP addresses, we looked at country-level domains (.br, .ua, .gov, etc.) to determine the originating country of spam messages. For all top-level domains (.com, .net, .org), we had to throw these spam messages into an unknown origins category for a lack of accuracy with respect to the originating location. These methods are reflected in other spam studies [9,69].

Microsoft data

There were a total of 80 countries in the data set. India had the higher percentage of spambots with 10.9% of the machines acting as spambots in the first quarter of 2011 and 11.0 % in the second. Jordan had the lowest percentage with 0.06% of machines acting as spambots in the first quarter of 2011 and 0.10 % in the second. Table 7 provides summary statistics for the data.

Table 7 Summary statistics for Microsoft spambot data

Statistics	1Q11	2Q11
Minimum	0.06	0.10
First quarter	0.21	0.19
Median	0.45	0.45
Mean	1.14	1.31
Third quarter	1.30	1.00
Maximum	10.90	11.00

Competing interests

The authors declare that they do not have any competing interests.

Acknowledgements

We would like to thank Prof. Angela Sasse and Nathaniel Husted for their insights and comments. We are also thankful to the Stat/Math Center at Indiana University, specifically Thomas Jackson, for their help with the statistical analysis. Last but not the least, we are indebted to the anonymous reviewers whose feedback was critical to improving this research. Any mistakes in the paper are the authors' own.

Author details

[1] Department of Computer Science, Drexel University, Philadelphia, PA 19104, USA. [2] School of Informatics and Computing, Indiana University, Bloomington, IN 47408, USA.

Received: 14 August 2013 Accepted: 20 October 2013
Published: 30 October 2013

References

1. J Postel, On the junk mail problem. Request for comments. Netw Working Group (1975)
2. P Denning, Electronic Junk. Commun. ACM **25**(3), 163–165 (1982)
3. B Stone-Gross, T Holz, G Stringhini, G Vigna, in *Proceedings of the 4th USENIX conference on Large-scale exploits and emergent threats*, LEET'11. The underground economy of spam: a botmaster's perspective of coordinating large-scale spam campaigns (USENIX Association Berkeley, CA, 2011), p. 4. http://dl.acm.org/citation.cfm?id=1972441.1972447. Accessed 29 October 2013
4. MessageLabs, MessageLabs Intelligence; 2010 Annual Security Report. Tech. rep., MessageLabs (2010)

5. Q Norton, Anonymous Tricks Bystanders Into Attacking Justice Department. Tech. rep., Wired (2012), http://www.wired.com/threatlevel/2012/01/anons-rickroll-botnet/. Accessed 29 October 2013
6. P Yu, The graduated response. Fla. Law Rev. **62**, 1373–1430 (2010)
7. C Grier, K Thomas, V Paxson, M Zhang, in *Proceedings of the 17th ACM conference on Computer and communications security,* CCS '10. @spam: the underground on 140 characters or less (ACM, New York, NY, 2010), pp. 27–37. http://doi.acm.org/10.1145/1866307.1866311. Accessed 29 October 2013
8. T Moore, R Clayton, in *Workshop on the Economics of Information Security*. An empirical analysis of the current state of phishing attack and defence. http://weis07.infosecon.net/papers/51.pdf. Accessed 29 October 2013
9. C Kanich, C Kreibich, K Levchenko, B Enright, GM Voelker, V Paxson, S Savage, in *Proceedings of the 15th ACM Conference on Computer and Communications security,* CCS '08. Spamalytics: an empirical analysis of spam marketing conversion (ACM, New York, 2008), pp. 3–14. http://doi.acm.org/10.1145/1455770.1455774. Accessed 29 October 2013
10. D McCoy, A Pitsillidis, G Jordan, N Weaver, C Kreibich, B Krebs, GM Voelker, S Savage, K Levchenko, in *Proceedings of the 21st USENIX conference on Security symposium,* Security'12. PharmaLeaks: understanding the business of online pharmaceutical affiliate programs (USENIX Association, Berkeley, 2012), p. 1. http://dl.acm.org/citation.cfm?id=2362793.2362794. Accessed 29 October 2013
11. Z Li, Q Liao, in *Managing Information Risk and the Economics of Security*. A Striegel, Botnet Economics: Uncertainty Matters (Springer US, 2009), pp. 245–267. http://dx.doi.org/10.1007/978-0-387-09762-6_12. Accessed 29 October 2013
12. T Holz, M Engelberth, F Freiling, in *Proceedings of the 14th European conference on Research in computer security,* ESORICS'09. Learning more about the underground economy: a case-study of keyloggers and dropzones (Springer-Verlag, Berlin, Heidelberg, 2009), pp. 1–18. http://dl.acm.org/citation.cfm?id=1813084.1813086. Accessed 29 October 2013
13. D Bleaken, Botwars: the fight against criminal cyber networks. Comput. Fraud Secur. **2010**(5), 17–19 (2010)
14. G Higgins, A Wilson, B Fell, An application of deterrence theory to software piracy. J. Crim. Justice Pop. Cult. **12**(3), 166–184 (2005)
15. I Png, C Wang, Q Wang, The deterrent and displacement effects of information security enforcement: international evidence. J. Manag. Inf. Syst. **25**(2), 125–144 (2008)
16. V Garg, N Husted, N Camp, in *eCrime Researcher's Summit*. Smuggling Theory Approach to Organized Digital Crime (IEEE, San Diego, 2011), pp. 1–7. http://ieeexplore.ieee.org/xpl/abstractAuthors.jsp?arnumber=6151980. Accessed 29 October 2013
17. J Buchanan, AJ Grant, Investigating and prosecuting Nigerian fraud. U. S. Attorneys' Bull. **49**(6), 39–47 (2001)
18. PK Yu, The graduated response. Fla. Law Rev. **62**, 1373 (2010)
19. P Sayer, French court levies first fine under three-strikes law on illegal downloads under Hadopi. Tech. rep., Computer World (2012)
20. V Garg, J Camp, in *The Research Conference on Communication, Information, and Internet Policy*. Ex Ante vs. Ex Post: Economically Efficient Sanctioning Regimes for Online Risks (Arlington, VA, 27–29, September 2013)
21. R Anderson, C Barton, R Böhme, R Clayton, M van Eeten, M Levi, T Moore, S Savage, in *Workshop on the Economics of Information Security*. Measuring the Cost of Cybercrime (Berlin, 25–26 June 2012)
22. TC Schelling, *Micromotives and macrobehavior.* (WW Norton & Company, New York, 2006)
23. H Varian, *System Reliability and Free Riding* (Kluwer Academic Publishers, Norwell, 2004), pp. 1–15
24. B Schneier, in *Proceedings of the Cryptology in Africa 1st International Conference on Progress in Cryptology,* AFRICACRYPTŠ08. The psychology of security (Springer-Verlag Berlin, Heidelberg, 2008), pp. 50–79. http://dl.acm.org/citation.cfm?id=1788634.1788642. Accessed 29 October 2013
25. M van Eeten, JM Bauer, H Asghari, S Tabatabaie, The Role of Internet Service Providers in Botnet Mitigation: An Empirical Analysis Based on Spam Data. OECD Science, Technology and Industry Working Papers 2010/5, OECD Publishing (2010). http://dx.doi.org/10.1787/5km4k7m9n3vj-en. Accessed 29 October 2013
26. S Karge, The role of Internet intermediaries in advancing public policy objectives. Tech. rep., OECD (2009)
27. T Lewis, Something for nothing [electronic commerce]. Comput. **32**(5), 120–118 (1999)
28. R Kannan, An empirical study of long-run impact of Internet advertising on consumer response behavior. *PhD thesis*, Massachusetts Institute of Technology, 2006
29. L Cranor, B LaMacchia, Spam! Commun. ACM **41**(8), 74–83 (1998)
30. H Esquivel, A Akella, T Mori, in *Proceedings of the 2nd International Conference on COMmunication Systems and NETworks,* COMSNETS'10. On the effectiveness of IP reputation for spam filtering (IEEE Press, Piscataway, 2010), pp. 40–49. http://dl.acm.org/citation.cfm?id=1831448. Accessed 29 October 2013
31. I Androutsopoulos, J Koutsias, KV Chandrinos, CD Spyropoulos, in *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. An experimental comparison of naive Bayesian and keyword-based anti-spam filtering with personal e-mail messages (ACM, New York, 2000). pp. 160–167
32. A Bratko, B Filipič, G Cormack, T Lynam, B Zupan, Spam filtering using statistical data compression models. J. Mach. Learn. Res. **7**, 2673–2698 (2006)
33. T Moore, R Clayton, H Stern, in *Proceedings of the 2nd USENIX Conference on Large-scale Exploits and Emergent Threats: Botnets, Spyware, Worms, and more,* LEET'09. Temporal correlations between spam and phishing websites (USENIX Association, Berkeley, 2009), p. 5. http://dl.acm.org/citation.cfm?id=1855676.1855681. Accessed 29 October 2013
34. T Moore, R Clayton, R Anderson, The economics of online crime. J. Econ. Perspect. **23**(3), 3–20 (2009)
35. E Allman, The economics of spam. Queue **1**(9), 80 (2003)
36. Y Nadji, M Antonakakis, R Perdisci, W Lee, in *International Symposium on Research in Attacks, Intrusions, and Defenses*. Connected Colors: Unveiling the Structure of Criminal Networks (Rodney Bay, 23–25 October 2013)
37. H Liu, K Levchenko, M Félegyházi, C Kreibich, G Maier, GM Voelker, S Savage, in *Proceedings of the 4th USENIX conference on Large-scale Exploits and Emergent Threats,* LEET'11. On the effects of registrar-level intervention (USENIX Association, Berkeley, 2011), p. 5. http://dl.acm.org/citation.cfm?id=1972441.1972448. Accessed 29 October 2013
38. J Goodman, G Cormack, D Heckerman, Spam and the ongoing battle for the inbox. Commun. ACM **50**(2), 24–33 (2007)
39. E Bursztein, S Bethard, C Fabry, JC Mitchell, D Jurafsky, in *Proceedings of the 2010 IEEE Symposium on Security and Privacy,* SP '10. How Good Are Humans at Solving CAPTCHAs? A Large Scale Evaluation (IEEE Computer Society, Washington, 2010), pp. 399–413. http://dx.doi.org/10.1109/SP.2010.31. Accessed 29 October 2013
40. M Motoyama, K Levchenko, C Kanich, McD Coy, GM Voelker, S Savage, in *Proceedings of the 19th USENIX conference on Security,* USENIX Security'10. Re: CAPTCHAs: understanding CAPTCHA-solving services in an economic context (USENIX Association, Berkeley, 2010), p. 28. http://dl.acm.org/citation.cfm?id=1929820.1929858. Accessed 29 October 2013
41. D Liu, LJ Camp, in *Workshop on the Economics of Information Security*. Proof of Work can Work. Cambridge, 26–28 June 2006
42. D Twining, MM Williamson, M Mowbray, M Rahmouni, in *USENIX Annual Technical Conference, General Track*. Email Prioritization: Reducing Delays on Legitimate Mail Caused by Junk Mail (Boston, MA, 27 June–02 July 2004), pp. 45–58
43. C Osorio, in *Program on Internet and Telecoms Convergence*. A contribution to the understanding of illegal copying of software: empirical and analytical evidence against conventional wisdom (MIT, 2002). http://hdl.handle.net/1721.1/1479. Accessed 29 October 2013
44. J Bhagwati, B Hansen, A theoretical analysis of smuggling. Q. J. Econ. **87**(2), 172–187 (1973)
45. B Barrett, What is SOPA? Tech. rep., Gizmodo (2012)
46. J Sanchez, SOPA Internet Regulation And the Economics of Piracy. Tech. rep., CATO Institute (2012)
47. W Easterly, M Sewadeh, *Global Development Network Growth Database*. (World Bank Group, Washington, 2001). http://go.worldbank.org/ZSQKYFU6J0. Accessed 29 October 2013
48. I Al-Jabri, A Abdul-Gader, Software copyright infringements: an exploratory study of the effects of individual and peer beliefs. Omega **25**(3), 335–344 (1997)
49. M Katz, C Shapiro, Technology adoption in the presence of network externalities. J. Pol. Econ. **94**(4), 822–841 (1986)
50. H Varian, *Economics of Information Technology*. (University of California, Berkeley, 2001)

51. T Pratt, F Cullen, Assessing macro-level predictors and theories of crime: a meta-analysis. Crime Justice **32**, 373–450 (2005)
52. C Kilby, Supervision and performance: the case of World Bank projects. J. Dev. Econ. **62**, 233–259 (2000)
53. M Felson, L Cohen, Human ecology and crime: a routine activity approach. Hum. Ecol. **8**(4), 389–406 (1980)
54. T Kelley, L Camp, in *11th Annual Workshop on the Economics of Information Security*. Online promiscuity: prophylactic patching and the spread of computer transmitted infections (WEIS, Berlin, 2012). http://weis2012.econinfosec.org/papers/Kelley_WEIS2012.pdf. Accessed 29 October 2013
55. W Bonger, *Race and Crime* (Patterson Smith, Montclair, 1969)
56. J Blau, P Blau, The cost of inequality: Metropolitan structure and violent crime. Am. Sociol. Rev. **47**, 114–129 (1982)
57. F Cullen, Social support as an organizing concept for criminology: presidential address to the academy of criminal justice sciences. Justice Q. **11**(4), 527–559 (1994)
58. D Kaufmann, A Kraay, M Mastruzzi, The worldwide governance indicators: methodology and analytical issues. SSRN eLibrary (2010). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1682130. Accessed 29 October 2013
59. MJ van Eeten, JM Bauer, Economics of Malware: Security Decisions, Incentives and Externalities. OECD Science, Technology and Industry Working Papers 2008/1, OECD Publishing (2008). http://ideas.repec.org/p/oec/stiaaa/2008-1-en.html. Accessed 29 October 2013
60. Microsoft, Microsoft Security Intelligence Report (Volume 11). Tech. rep., Microsoft (2011). http://www.microsoft.com/security/sir/default.aspx. Accessed 29 October 2013
61. V Garg, LJ Camp, in *Network and Distributed System Security Symposium Extended Abstracts*. Macroeconomic Analysis of Malware (San Diego, CA, 24–27 February 2013)
62. S Baum, J Ma, Education pays for individuals and society. High. Educ. **57**(4), 1–48 (2007)
63. E Rader, R Wash, B Brooks, in *Proceedings of the Eighth Symposium on Usable Privacy and Security, SOUPS '12*. Stories as informal lessons about security (ACM, New York, 2012), pp. 6:1–6:17. http://doi.acm.org/10.1145/2335356.2335364. Accessed 29 October 2013
64. M Marlinspike, in *BlackHat DC*. New tricks for defeating SSL in practice (Arlington, VA, 16–19 February 2009)
65. B Jonah, R Thiessen, E Au-Yeung, Sensation seeking, risky driving and behavioral adaptation. Accid. Anal. Prev. **33**(5), 679–684 (2001)
66. R Anderson, in *Proceedings of the 17th Annual Computer Security Applications Conference,* ACSAC '01. Why Information Security is Hard-An Economic Perspective (IEEE Computer Society, Washington, 2001), p. 358. http://dl.acm.org/citation.cfm?id=872016.872155. Accessed 29 October 2013
67. P Diehr, T Lumley, The importance of the normality assumption in large public health data sets. Annu. Rev. Publ. Health **23**, 151–169 (2002)
68. L Zhuang, J Dunagan, DR Simon, HJ Wang, JD Tygar, in *Proceedings of the 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats,* LEET'08. Characterizing botnets from email spam records (USENIX Association, Berkeley, 2008), pp. 2:1–2:9. http://dl.acm.org/citation.cfm?id=1387711. Accessed 29 October 2013
69. A Ramachandran, N Feamster, in *Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications,* SIGCOMM '06. Understanding the network-level behavior of spammers (ACM, New York, 2006), pp. 291–302. http://doi.acm.org/10.1145/1159913.1159947. Accessed 29 October 2013

doi:10.1186/1687-417X-2013-3
Cite this article as: Garg *et al.*: Cross-country analysis of spambots. *EURASIP Journal on Information Security* 2013 **2013**:3.