Macroeconomic Analysis of Malware

Vaibhav Garg, L. Jean Camp
School of Informatics and Computing
Indiana University
{*gargv, ljcamp*}@*indiana.edu*

***Abstract*—Malware harms infected individuals by stealing computational resources and possibly authenticating information. Malware also creates negative externalities for all users by enabling the creation of botnets for criminal enterprise, e.g. DDoS or phishing. Why is there such a variance in the percentage of malware infections across nations? While there may be idiosyncratic or ad-hoc explanations for concentrations of malicious actors or innocent victims, there has been an absence of systematic cybercrime science. Computer scientists have focused primarily on empirical work without concurrent focus on potentially applicable theories in macroeconomics and criminology. This extended abstract tersely provides the results of why botnets vary by jurisdiction and region, but the focus is on an explanation of and argument for our method: macroeconomic analysis informed by crime science.**[1]

I. Introduction

Technical efforts, such as, anti-virus, aim to thwart cybercriminals, through what is effectively an arms race. Yet without incentive alignment and clear market signals, security solutions often suffer from underconsumption [1]. Regulatory initiatives to defeat cybercrime have focused on deterrence by prosecuting individuals that participate in cybercrime [11]. Deterrence-based approaches have potential [10], but the impact may be limited in duration [17]. Arguably, even successful deterrence has reached the point of diminishing returns [2].

Deterrence-based approaches when applied to unsuspecting end-users participating in cybercrime, e.g. zombies, may be ineffective, even unjust, and thus undesirable [16]. Alternative solutions can instead leverage public-private partnerships; for example, the German anti-botnet initiative where Internet Service Providers (ISPs) inform their customers if they have unknowingly become part of a botnet, while also providing appropriate technical support to mitigate risk.

The systematic development of policy and technical solutions to crime online, not grounded in deterrence, requires a science of cybercrime that is theoretically grounded. Simultaneously, while microeconomic approaches to investigating crime online have been the norm, complementary macroeconomic insights are rare. However, cybercrime (and resulting victimization) much like crime offline is concentrated in specific nations and thus motivates cross-country analysis.

Why are some countries less likely to have a high percentage of systems being infected by malware? In this paper, we analyze the macroeconomic factors that relate to the cross country variance in malware infected systems. In the process we provide a roadmap for appropriate methodology, that utilizes publicly available data from the World Bank thereby providing an analysis that is quantitative, repeatable, and verifiable.

[1]Vaibhav Garg & L Jean Camp, Macroeconomic Analysis of Malware, NDSS (San Diego, CA) 24-27 February 2013, extended abstract.

II. Background and Related Work

Models in previous empirical crime research and much previous cybercrime work presumes voluntary participation in illicit activity. However, this does not hold in cybercrime, most obviously as victims, when hijacked by malware [14]. While cybercrime activities such as spam and phishing may be short lived, the same is not true for the infrastructure, constituting infected machines, that supports these activities. What factors explain the cross-country variance in the number of infected machines? This paper addresses this question by examining the underlying macroeconomic variables.

The lack of individual incentive to invest in a public or private good is often overcome through regulatory or policy solutions [3]. Incentives are also widely used, for example, graduated response or 'three strikes and you are out' [16] prescribes punitive measures to rationally discourage copyright infringement. Such solutions are grounded in a *deterrence theory* of crime [10]. While such solutions are potentially successful in the short run, their effectiveness in the long term may be limited by displacement of criminal enterprise to a different jurisdiction [17].

Microeconomic investigations have been useful in demonstrating the existence of organized underground cybercrime markets, the different stakeholders involved, and the process of transactions[6], [13], [20]. However, complementary macroeconomic insights that explain how these structures emerge and why they persist are rare. Simultaneously, theoretically grounding empirical research is much needed to develop systematic policy insights that are testable hypotheses.

For example, Osorio conducted an empirical examination of massive software copyright infringement [15]. He concluded that copyright violations are a function of access and affordability, being predicted by GDP per capita and availability of post-sales software support in local markets. The immediate policy implication argues for price cuts [5] over SOPA [19].

Garg et al. applied extant models of smuggling offline [4] to crime online [9]. They argue that existing illegal markets can act as a prohibitive tariff suppressing the development of legal services. Simultaneously, cybercrime can be welfare increasing in local jurisdictions skewing the incentives for local law enforcement to crack down on such activities.

Thus, while illegal markets might emerge as a consequence blocked legitimate opportunity [15], they persist as at least in the short term they are locally social welfare increasing [9]. This paper then presents an arguments for a similar examination of other cybercrime activities with an empirical macroeconomic analysis of the variance in the number of malware infected machines in different nations.

III. Methodology

Our core research question is determining which factors explain the variance across countries in the percentage of machines which are zombies or otherwise infected with malware. We examine a publicly available dataset, published by Microsoft, that reports the percentage of malware infected machines in different countries. The data corresponds to systems running different flavors of Windows operating system and has been taken from Microsoft's Security Intelligence Report 2011 [12]. The data is generated using data from 600 million computers across the globe. The total number of countries reported is 119, where the location of the specific machine is determined using geoip. The dependent variable is the percentage of malware infected machines, calculated as the number of computers cleaned for every thousand executions of the Malicious Software Removal Tool.

The macroeconomic factors, which engender the independent variables, are determined by considering previous research in microeconomics of cybercrime and criminology. The data corresponding to these variables have been obtained from the World Bank, whose definitions of specific variables is often vague. However, World Bank data provide uniform, consistently available measures that are likely to be repeated over time by the World Bank. This will allow other researchers to reproduce our work and with data arguably used for broader empirical examinations of Internet readiness, cybercrime etc. Most importantly, given that we are engaged in multiple analyses, if the results are a function of the data then the biases would be consistent. So, for example, should we find significant support across studies for the need for different types of investment in deterrence, prevention, or harm mitigation in cybercrime, these would hold. World Bank data are used across domains (including crime and jurisprudence) to inform major policy debates. The data for all independent variables corresponds to the year 2010.

The *routine activity* theory of crime posits that the probability of crime is a function of motivated offenders, proximity to targets, and lack of effective guardianship [18]. The first independent variable then is the **availability** of machines. Rational malicious actors are drawn to large or rich pools of potential victims; when there are no available computational resources there is no cybercrime. Not surprisingly, Eeten et al. note that the size of the Internet Service Provider (ISP) is the largest driver of the spam generated by the user-base [21]. Size can be operationalized by the size of the market, connectivity measures, or the size of the user base. Corresponding macroeconomic variables are export of computer communications and other services (CCS), fixed Internet broadband subscribers per 100 people (FBIS), and number of Internet users per 100 people (IU).

A second independent variable is network protection or **guardianship**, conceived as private investments in security. For example, individual private investments can be operationalized by the market penetration of security software, e.g., MacAfee software sales. Such data is unfortunately rarely available to general body of researchers. A proxy variable then is the number of secure Internet servers (SIS), which serves as a measure of aggregate market investments in security by private stakeholders. Secure Internet servers is defined by the World Bank as "servers using encryption technology in Internet transactions[2]".

Crime may be a function of blocked legitimate **economic** opportunities or resource deprivation [18]. Certainly economic availability, affordability, and access to resources are then simultaneously implied [15]. As such, lack of economic resources would impinge the individual ability to purchase legitimate software and invest in recovery. While software is often discounted the discount is rarely enough to reflect the difference in purchasing power. One variable that measures individual purchasing power based on local conditions is Gross Domestic Product (GDP) per capita by purchasing power parity (PPP). Simultaneously, GDP per capita by PPP measures not only wealth within a nation but embeds a measurement of income inequality across nations.

Individual lack of economic resources can be alleviated by appropriate **governance** through public investments (i.e., social support theory) [18]; consistent with *structural theory* of crime. Government support can be direct, in the form of subsidies or incentives for adoption of ICT technologies. These may include provision of free products, education, advantageous regulatory regimes, as well as more traditional financial forms. We operationalize social support with a subset of World Governance Indicators (WGI), also provided by the World Bank. The subset we include consists of 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 services, policy formulation and and credibility of the government. Regulatory quality quantifies the government competence and alignment with private sector development. Rule of law is the degree to which the legal framework is followed in practice. Control of corruption measures perceptions of misuse of public power for private gain. (These were combined into one variable titled WGI to account for multicollinearity.)

IV. Results

We detail our analyses in the extended version. In general, normality, multicollinearity, and heteroskedasticity should be addressed to identify the appropriate estimation methods. For example, while Ordinary Least Squares regression not always appropriate [8], exceptions can be made in the case of large data sets [7].

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

Table I provides a list of the macroeconomic factors and respective correlations with number of infected machines.

TABLE I
CORRELATIONS

Variable	Correlation (n)
GDP by PPP	-0.53(108)***
FBIS	-0.33 (116)***
FBIS per 100 people	-0.58(116)***
IU per 100 people	-0.49(116)***
SIS	-0.46(118)***
SIS per 1 million people	-0.51(118)***
CCS (%exports)	-0.34(110)***
CCS(%imports)	-0.34(110)***
WGI	-0.45 (119)***

Signif. codes: '***' < 0.001 '**' <0.01 '*' < 0.05

Table II presents the OLS regression estimates. The model estimates and standard errors have been presented by taking heteroskedasticity into account[3].

TABLE II
REGRESSION ESTIMATES

Variable	Estimate	Std. Error	Pr(> $\|t\|$)
(Intercept)	3.633	0.983	0.00 ***
GDP by PPP	-0.062	0.139	0.66
FBIS	0.051	0.058	0.38
FBIS per 100 people	-0.026	0.014	0.07
IU per 100 people	-0.007	0.004	0.13
SIS	-0.049	0.053	0.36
SIS per 1 million people	0.079	0.066	0.24
CCS (%exports)	-0.006	0.003	0.07
CCS(%imports)	0.001	0.003	0.74
WGI	0.0003	0.001	0.76

Signif. codes: '***' < 0.001 '**' <0.01 '*' < 0.05

Residual standard error: 0.4461 on 90 degrees of freedom
Multiple R-squared: 0.4818, Adjusted R-squared: 0.43
F-statistic: 9.299 on 9 and 90 DF, p-value: 7.443e-10

V. CONCLUSIONS AND FUTURE WORK

Given that FBIS per 100 people is negatively correlated with number of infected machines implies that while higher adoption may lead to an overall increase in the total bots, the percentage of offending systems is reduced. Previously, Eeten et al. [21] too noted that while ISPs with bigger user base had higher number of infected machines, bigger ISPs did better in terms of percentages. Arguably then as Internet grows, it becomes more secure on average.

Table II indicates that our macroeconomic model explains a significant amount of variance in the percentage of infected machines in different countries (≈43%). All the macroeconomic variables indicate a negative relationship with the percentage of infected systems; thus, suggesting a host of potential candidates to be addressed by public policy and private enterprise. For example, if higher SIS lowers malware, regulation can mandate encryption. Alternatively, if GDP by PPP is the primary driver, price cuts may improve the adoption of secure technologies. While a causal relationship is not obviously implied, it does argue for further analysis by back ward time series analysis using historical botnet data.

This paper we concludes with a restatement of the core argument, that is, for a systematic macroeconomic investigation of crime online and victimization, grounded in criminology to engender a science of cybercrime.

[3]Specifically, we used the White-Huber method to generate heteroskedasticity corrected covariance matrices

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grants 1259172 and 1250367. 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.

REFERENCES

[1] R. Anderson. Why information security is hard-an economic perspective. In *Computer Security Applications Conference, 2001. ACSAC 2001. Proceedings 17th Annual*, pages 358–365, 2001.

[2] R. Anderson, C. Barton, R. Bhme, R. Clayton, M. van Eeten, M. Levi, T. Moore, and S. Savage. Measuring the cost of cybercrime. In *Workshop on the Economics of Information Security*. WEIS, 2012.

[3] R. Anderson and T. Moore. The economics of information security. *Science*, 314(5799):610–613, 2006.

[4] J. Bhagwati and B. Hansen. A theoretical analysis of smuggling. *The Quarterly Journal of Economics*, 87(2):172–187, 1973.

[5] Y. Chen and I. Png. Software pricing and copyright enforcement: private profit vis-a-vis social welfare. In *Proceedings of the 20th international conference on Information Systems*, ICIS '99, pages 119–123, Atlanta, GA, USA, 1999. Association for Information Systems.

[6] K. Choo and R. Smith. Criminal exploitation of online systems by organised crime groups. *Asian journal of criminology*, 3(1):37–59, 2008.

[7] P. Diehr and T. Lumley. The importance of the normality assumption in large public health data sets. *Annual Review Public Health*, 23:151–169, 2002.

[8] T. Dietz, R. Frey, and L. Kalof. Estimation with cross-national data: robust and nonparametric methods. *American Sociological Review*, 52(3):380–390, 1987.

[9] V. Garg, N. Husted, and J. Camp. Smuggling theory approach to organized digital crime. In *eCrime Researcher's Summit*. IEEE, 2011.

[10] G. Higgins, A. Wilson, and B. Fell. An application of deterrence theory to software piracy. *Journal of Criminal Justice and Popular Culture*, 12(3):166–184, 2005.

[11] T. Holz, M. Engelberth, and F. Freiling. Learning more about the underground economy: a case-study of keyloggers and dropzones. In *Proceedings of the 14th European conference on Research in computer security*, ESORICS'09, pages 1–18, Berlin, Heidelberg, 2009. Springer-Verlag.

[12] Microsoft. Microsoft security intelligence report. Technical report, Microsoft, 2011.

[13] T. Moore and R. Clayton. An empirical analysis of the current state of phishing attack and defence. In *Workshop on the Economics of Information Security*. WEIS, 2007.

[14] Q. Norton. Anonymous tricks bystanders into attacking justice department. Technical report, Wired, 2012.

[15] C. Osorio. A contribution to the understanding of illegal copying of software: Empirical and analytical evidence against conventional wisdom. In *Program on Internet and Telecoms Convergence*. MIT, 2002.

[16] K. Peter. The graduated response. *Fla. L. Rev.*, 62:1373, 2010.

[17] I. Png, C.-Y. Wang, and Q.-H. Wang. The deterrent and displacement effects of information security enforcement: International evidence. *J. Manage. Inf. Syst.*, 25(2):125–144, Sept. 2008.

[18] T. Pratt and F. Cullen. Assessing macro-level predictors and theories of crime: A meta-analysis. *Crime and Justice*, pages 373–450, 2005.

[19] J. Sanchez. Sopa, internet regulation and the economics of piracy. Technical report, CATO Institute, 2012.

[20] R. Thomas and J. Martin. The underground economy: priceless. *USENIX; login*, 31(6):7–16, 2006.

[21] M. van Eeten, J. M. Bauer, H. Asghari, and 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, May 2010.