Analyzing and Modeling Longitudinal Security Data: Promise and Pitfalls

Benjamin Edwards
University of New Mexico
bedwards@cs.unm.edu

Steven Hofmeyr
Lawrence Berkeley National Laboratory
shofmeyr@lbl.gov

Stephanie Forrest
University of New Mexico
Santa Fe Institute
forrest@cs.unm.edu

Michel van Eeten
Delft University of Technology
M.J.G.vanEeten@tudelft.nl

ABSTRACT

Many cybersecurity problems occur on a worldwide scale, but we lack rigorous methods for determining how best to intervene and mitigate damage globally, both short- and long-term. Analysis of longitudinal security data can provide insight into the effectiveness and differential impacts of security interventions on a global level. In this paper we consider the example of spam, studying a large high-resolution data set of messages sent from 260 ISPs in 60 countries over the course of a decade. The statistical analysis is designed to avoid common pitfalls that could lead to erroneous conclusions. We show how factors such as geography, national economics, Internet connectivity and traffic flow impact can affect local spam concentrations. Additionally, we present a statistical model to study temporal transitions in the dataset, and we use a simple extension of the model to investigate the effect of historical botnet takedowns on spam levels. We find that in aggregate most historical takedowns are beneficial in the short-term, but few have long-term impact. Further, even when takedowns are effective globally, they can be detrimental in specific geographic regions or countries. The analysis and modeling described here are based on a single data set. However, the techniques are general and could be adapted to other data sets to help improve decision making about when and how to deploy security interventions.

Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information Systems**]: Security and Protection

Keywords

Spam, takedowns, statistical model

1. INTRODUCTION

Many cybersecurity problems occur at a global scale, involving nations, corporations, or individuals whose actions have impact around the world. Despite these global, persistent problems, there is limited research on the actual effectiveness of the many interventions that have been proposed. Most interventions are based on deep, hands-on experience with specific attacks, and are never evaluated systematically at a large scale. For example, there has been little quantitative analysis of the sustained effect of the many botnet takedowns that occurred over the past decade. Simple qualitative observation of declines in malicious activity following a takedown is not sufficient to determine whether the takedown is effective or causal [19, 31, 33]. Attributing cause is always problematic, but it is especially difficult when empirical datasets have high variance as is often the case in security[13, 15]. Moreover, as the scope of cyber-insecurity has increased, no one security practitioner is able to grasp all of the relevant details associated with global problems [17]. Thus, there is a need for more explicit and rigorous methods to determine which interventions are effective and which are are not.

In this paper, we explore some of the opportunities and impediments to analyzing longitudinal security data, by focusing on the concrete example of spam, developing statistical models to describe a large dataset, and using the model to assess the effect of certain interventions. We ask whether a particular intervention has a temporary or sustained impact and how interventions play out geographically. A potential pitfall in longitudinal datasets, including our dataset, is high variance, and we use careful statistical methods to separate significant effects from noise. A second issue is the retrospective nature of data-driven analyses, which makes predicting the future a challenge. Because intervention methods are often re-used, however, we believe that studying the existing examples, e.g., a historical botnet takedown, can provide insight about the likely effect of similar future interventions.

We illustrate our approach by analyzing a spam dataset, comprising more than 127 billion spam messages sent from over 440 million unique IP addresses, spread across 260 ISPs in 60 countries. Spam is a global problem, and countermeasures have never eliminated it completely. Spam plays a key role in the cyber-crime ecosystem as a vector for various activities such as stealing login credentials through phishing, distributing malware, making fraudulent sales, or selling illegal goods [37]. Spam can be viewed as a proxy for estimating the numbers of infected PCs and the extent of botnets [83, 70, 26].

To compare spam levels across countries, we study a quantity called *wickedness* [25], which can be thought of as the concentration of infected machines sending out spam, either in a single Internet Service Provider (ISP) or in a geographic region. This measure allows us to compare spam levels among different countries or different ISPs, identify how different factors contribute to the concentration of spam sending computers, and assess what

ACSAC '15, December 07 - 11, 2015, Los Angeles, CA, USA

DOI: http://dx.doi.org/10.1145/2818000.2818010

effect interventions have across the globe.

Analysis of the data shows that spam concentrations are relatively stable for ISPs from one week to the next but are punctuated by spikes that often span several orders of magnitude. These spikes can mask the effect of interventions. Further analysis also reveals that: (1) Gross Domestic Product (GDP) per capita is negatively correlated with wickedness, with less developed countries experiencing higher levels; (2) an ISP's wickedness is correlated with that of surrounding ISPs, suggesting that there are regional influences; and (3) an ISP's network connectivity is correlated with wickedness.

To further understand the impact of ISP connectivity on spam, we construct an *ISP graph* that represents how ISPs are connected to each other. The graph reveals that ISPs with high graph centrality have lower wickedness, while those on the periphery suffer higher rates of infection. Adding a simple model of spam dynamics to the ISP graph shows that spam concentrations at an ISP are influenced by previous levels, suggesting that spam could is one driver in spreading infections across the Internet.

In the last decade, a number of approaches have been suggested and implemented to help fight spam. Of these, the most famous is the botnet takedown. But, email providers have also adopted adaptive IP black lists [21], banks have restricted access to credit card payment processors [29], resources have been devoted to arresting and prosecuting cyber-criminals [51, 36, 1], and users of infected computers have been offered free cleanup tools [4]. Some of these interventions seem to have led to declining spam levels, e.g., real-time filtering and credit card interventions [71, 50, 29, 68].

We show how modeling can help identify when particular interventions likely began affecting spam concentrations. The best model of our dataset identifies three distinct time periods or *eras*, each corresponding to different dynamics. These eras correlate roughly with the introduction of new intervention strategies, and they give some idea of the overall impact of a particular strategy.

When the exact date of an intervention is known (as in the case of botnet takedowns), we can use the model to analyze its impact more precisely, both globally and regionally. Model analysis confirms the hypothesis that most botnet takedowns are effective only in the short-term, with spam levels rebounding in the weeks after a takedown [39]. However, we also find that a few of the takedowns were globally effective in the long term. A closer look at their regional impact, however, shows that effects vary dramatically across different geographic areas and individual countries. Takedowns that are successful globally can be detrimental in specific countries.

Our work uses one particular dataset to illustrate how robust statistical techniques can be applied to study spam trends and the effect of interventions—globally, regionally, and by individual country. Because we studied data taken from a single data source, and focused only on email spam, our conclusions are only as good as the data—a pitfall of any statistical analysis. The methods, however, could readily be applied to other sources of spam and even other security data, as they become avaiable. Additional datasets would certainly improve our confidence in the conclusions of the analysis, and section 2 discusses the idiosyncrasies of our particular dataset.

In summary, statistical analysis of global longitudinal data is a promising approach to understanding the security landscape. This paper makes the following contributions:

1. It presents a robust statistical analysis of longitudinal, global security data, showing how to analyze high variance time series, identify correlations with external factors, and identify the effects of interventions, both when the deployment date is unknown (filters) and when it is known exactly (botnets).
2. It identifies statistically significant correlations between spam concentrations and various risk factors, including GDP, nearby spam concentrations, and ISP connectivity in the ISP graph. Traffic dynamics on this graph influence future wickedness, suggesting that spam is used to spread malware infections.
3. Identification of three statistically distinct eras within the ten-year data set. Although spam levels are highly variable in all eras, the overall concentration of spam declines during the last two eras. These declines may be related to historical events that are outside the scope of our study, and they may have caused discernible shifts in the data.
4. Analysis of the global impact of historical botnet takedowns: only a few of the studied takedowns had lasting impact, while most had only a transient effect, in all eras.
5. Geographic impacts of takedowns. We find that even when a takedown is effective globally, it often results in an increase in wickedness in particular regions or countries.

2. COLLECTING AND MAPPING SPAM DATA TO WICKEDNESS

In this section we describe our dataset, and the *wickedness* metric. We show that wickedness has interesting statistical properties, and identify significant changes in wickedness over time.

2.1 Spam Data

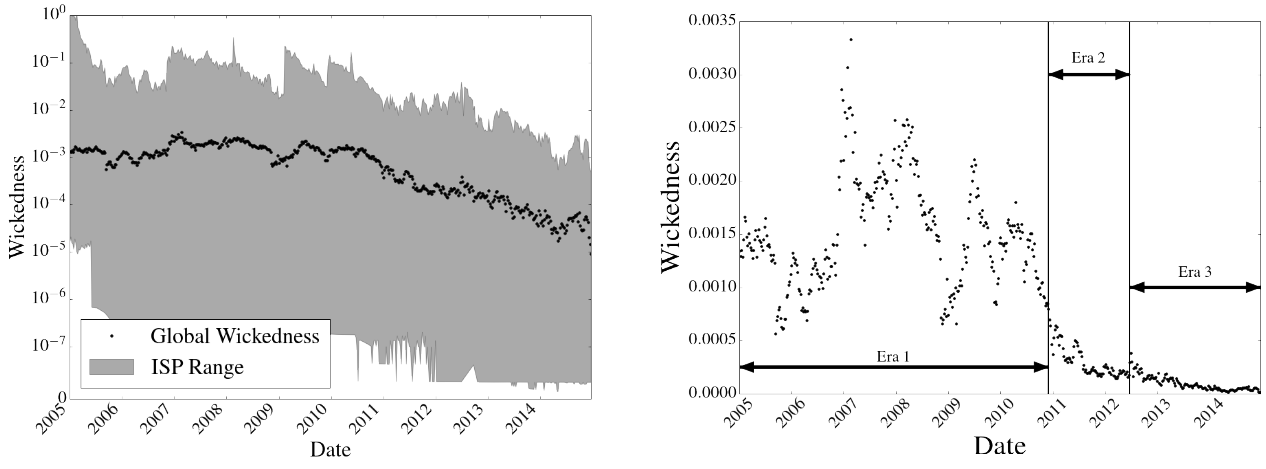
Our spam dataset is based on that used by Van Eeten et al. [73] but greatly expanded. We collected additional data, doubling the timespan covered, and studied the data on a weekly basis. The original study examined spam trends only on a quarterly basis. This dataset was collected from a *spam trap*—an Internet domain designed specifically to capture spam with e-mail addresses that have never been published or used to send or receive legitimate email. Spam traps have been used successfully to identify malware infected hosts, and to measure the extent of botnets, because botnets often send spam [83, 70, 26]. Over the past decade, our spam trap received more than 127 billion spam messages, sent from 440 million unique IP addresses.

In order to make comparisons among different ISPs and geographic regions, the ISP which owns each IP address and the country in which that ISP operates must be identified. To do this we used the following procedure:

1. Each IP address was linked to an ASN (Autonomous System number) using historical BGP data.
2. Each ASN was then manually linked an administrating entity using historical WHOIS records.
3. Industry reports and news media were consulted to connect the administrating entities to the main ISPs in 60 countries, as identified in Telegeography's GlobalComms database. The database also provides us with accurate subscriber numbers for each ISP.
4. Each (part of an) ASN was mapped onto a country using MaxMind's GeoIP database [47].

The manual mapping of ASNs to ISPs prevented us from identifying all possible ISPs which sent spam to our trap. However, we were able to map 659 ASNs to 260 ISPs in 60 countries. These ISPs account for over 80% of the major broadband markets in those countries. These countries also compose the entirety of the Organisation for Economic Co-operate and Development(OECD) and European Union, along with several other major spam sending nations.

This procedure produced two time series for each ISP: a count spam messages and the number of unique IP addresses that sent spam per day. Some ISPs provide dynamic IP addresses with short lease times to their customers. This could potentially cause a single spam-emitting host to be associated with multiple IP addresses over short periods. To help correct for this potential source of overcounting, we use average daily counts of IP addresses over the course of a week to obtain an estimate of the number of



(a)

(b)

Figure 1: Two views of global spam. (a) uses a logged vertical axis to show the high variance in the spam data. The black points indicate global wickedness, and the shaded area shows the range of values for individual ISPs. (b) uses a linear vertical axis to show the qualitative changes in wickedness between different eras.

infected machines associated with an ISP in a given week. This produces slightly coarser granularity data but removes some of the churn caused by dynamic addresses.

Our data was collected from a single spam trap and is only a sample of all the spam sent globally, and it is possible that our data is only capturing the activity of a few unsophisticated spam gangs. It is difficult to exactly compare our data to other publicly available spam reports because most reports rely on relative measures such as fraction of total email that was classified as spam or percentages relative to a peak. However we were able to make some qualitative comparisons to other sources. A subset of the data from 2006 and 2009 was previously found to correspond with industry reports, both in terms of spam volume trends and geographical distribution of sources [73].

Comparing post 2010 trends to longitudinal data available from Spamhaus [66], our data on global wickedness qualitatively matches theirs until mid-2012. After that, however, Spamhaus shows a brief rise in spam, though not to previous levels, while our data show a continued downward trend (Figure 1). Symantec reports a small overall decline in annual average spam in 2013 [69], and Kaspersky also reported a small decline the percentage of spam email compared to legitimate email in 2013 [23]. Our data also shows declines in these two years. The discrepancies between our data and Spamhaus likely reflect changes in tactics of spammers over time that are not captured by our spam trap. However, in this paper we emphasize the procedure used to analyze the data over the exact conclusions drawn from the analysis, which in future work could be verified by analyzing other datasets.

2.2 Estimating Wickedness From Spam Data

We calculate wickedness in terms of IP addresses sending spam. The two time series establish the total number of spam sending hosts within an ISP, but they do not account for the total number of IP addresses actively used by each ISP (i.e. the number of customers). We focus on wickedness rather than the absolute number of spam sending hosts so that we can make valid comparisons between ISPs, countries, and regions in the world. We use data from TeleGeography's Globalcomm database to establish the number of subscribers for each ISP. These data, available quarterly, allowed us to compute the concentration of malicious hosts per customer (the wickedness) and the number of spam messages sent per customer.[1] Using linear interpolation, we inferred the number of customers each week to match the time granularity of the data for malicious hosts. We calculate the wickedness of an ISP i at time t as:

$$W_i(t) = \frac{A_i(t)}{C_i(t)}. \qquad (1)$$

where $A_i(t)$ refers to the number of spam-sending IP addresses and $C_i(t)$ refers to the number of customers for ISP i at time t. The *global* wickedness is defined over all ISPs, i.e. $W(t) = \sum_i A_i(t) / \sum_i C_i(t)$. Figure 1 shows the global wickedness over time calculated from our dataset.

In these data, which capture a sample of the total population of spam-emitting hosts worldwide, between 0.00091% and 0.33% of hosts are sending spam at any given time. However, individual ISP infection rates vary widely as shown by the shaded area in Figure 1a, with some ISPs as high as 80% and others with 0%. Moreover, a single ISP's infection rate often varies by several orders of magnitude from one week to the next. For example, in April, 2011 an ISP in Pakistan experienced a more than 800-fold increase in wickedness in a single week. Previous work has also observed highly dynamic infection levels in IP space [6, 65].

In spite of this large variation, our analysis shows that wickedness at the individual ISP level is highly *autocorrelated*, i.e. the correlation between wickedness in any given week and the previous week is high (Kendall's $\tau = 0.93$). Kendall's τ is a non-parametric measure of statistical dependence. Unlike the more widely used Pearson's r, Kendall's τ does not assume a linear relationship between the data, and is therefore better able to identify non-linear relationships, which abound in our data [32].[2] This counterintuitive result is explained by the fact that in the vast majority of cases week-to-week variation is small, even though a minority of cases break this pattern by varying over several orders of magnitude. Such high variance can often lead to erroneous

[1] Alternatively, wickedness could be defined using messages per customer. We have analyzed the data both ways (data not shown), with essentially identical results.

[2] Measures of linear correlation between the $\ln W_i(t)$ and $\ln W_i(t-1)$ are exceptionally high (Pearson's $r = 0.990$), suggesting a nonlinear relationship similar to a power law. This informs the construction of our model in section 4.

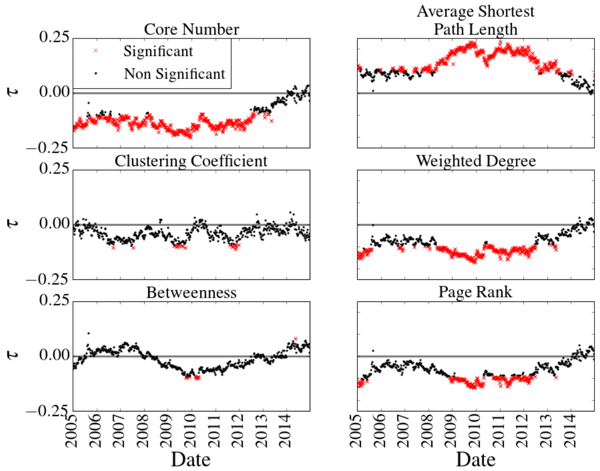


Figure 2: Correlation between wickedness and ISP graph topology. The vertical axis in all plots shows the Kendall's τ between wickedness and the topological measure for the corresponding week on the horizontal axis. Red indicates significant correlations at the $p < .05$ level.

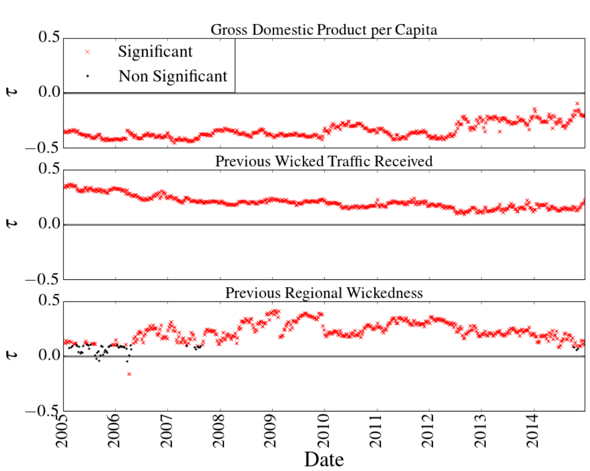


Figure 3: Correlation between wickedness and GDP (top panel), wickedness and traffic (middle panel), and wickedness and average regional wickedness (bottom panel). The vertical axis in all plots is Kendall's τ between wickedness and traffic during the week shown on the horizontal axis. Red indicates significant correlations at the $p < .05$ level.

conclusions about data. Many statistical methods require that data have limited variance, and using such methods might indicate significant changes when none exist [15].

Figure 1b shows several possible qualitative changes in spam volume, and in subsection 4.2 we find that spam exhibits statistically significantly different behavior during these periods.

Era 1: Beginning in 2005, spam increased dramatically until the first botnet takedowns occurred in 2008. During this time period spam levels were volatile and punctuated sharp increases and decreases both globally and at the ISP level.

Era 2: In mid 2010, spam levels began to drop dramatically and seemingly permanently. We find a statistically significant effect in late 2010.

Era 3: In mid 2012, a spike is observed in the data, followed by a further decline in wickedness. During this time period the variance in global wickedness also decreases.

These three eras are highlighted in Figure 1b. In subsection 4.2 we use maximum likelihood techniques to pinpoint when statistically significant transitions occurred and discuss possible causes of these transitions.

3. RISK FACTORS

The previous section defined wickedness and examined its properties in our dataset. Next we ask if certain external "risk factors" are related to an ISP's level of wickedness. In this section, we consider demographic factors, the effect of geography, network effects, and traffic dynamics.

3.1 Demographic Factors

Previous work identified correlations between spam concentrations and measures of development, such as Internet use per capita or education [73, 82]. We find similar results using gross domestic product per capita (GDP). GDP data were obtained from the World Bank, which produces annual data on a per-country for multiple demographic factors [3]. We use GDP per Capita because recent data is readily available, but other measures of development such as unemployment or corruption within institutions may provide different insights. We used linear interpolation to infer weekly values from the annual data.

For each week of data, we compute τ between ISP wickedness and the GDP of the country in which each ISP was operating. The top panel of Figure 3 shows these correlations over the course of our 520 weeks, and indicates that GDP is consistently negatively correlated with wickedness, in agreement with results from previous studies [73, 82]. In subsection 4.1 we calculate the size of this effect. The decline in magnitude in correlation seen in the later portions of the data may indicate that infection rates are becoming less tethered to development, as technology levels rise across the globe.

3.2 Geographic Clustering

Qualitatively, we observe that wickedness levels cluster in certain geographic regions during specific periods. For example, during January, 2011 high levels of wickedness are observed in Eastern European countries. Roughly a year later, Eastern Europe experienced lower levels of wickedness while there were higher concentrations in Southeast Asia[3].

To study this geographic clustering, we divide the world into 14 regions, as defined by the United Nations [56], and measured the correlation between the wickedness of an ISP and the average wickedness of all other ISPs in the same region (excluding the original ISP) in the previous week.

We find significant positive correlations between this value and wickedness throughout most of the data (see Figure 3). We study this result more in depth in section 4.

3.3 Autonomous System Topology

Another possible risk factor is an ISP's position in the topological structure of the Internet at the Autonomous System (AS) routing level. To investigate the strength of this effect, we measured the correlation between wickedness and several popular topological metrics [14]. This is not straightforward, however, because our data were collected at the ISP level, and connectivity between ISPs is not identical to Autonomous System connectivity. We address this problem by constructing a hybrid network that reflects both topologies.

We constructed this new network by beginning with the AS level, retrieving AS network data from the *Internet Research*

[3] Map not shown due to space constraints

Lab's Internet AS-Level Topology Archive.[4] The archive collects daily and monthly snapshots of AS-level topology from a number of different sources and, at the time of download on February 11, 2015, was one of the most complete publicly available sources of the AS-level Internet topology [58]. We construct the ISP graph using the following steps:

1. *Aggregate nodes*: Combine all ASNs owned by a single ISP into a single node. This produces a graph that contains both ISP nodes and ASN nodes.
2. *Aggregate edges*: If there are multiple edges between two nodes, combine them into a single weighted edge, with weight equal to the number of connections between the nodes.
3. *Remove stubs*: Remove ASN nodes that are not directly connected to an ISP and have degree equal to one.
4. Combine the daily version of the graph into a weekly snapshot by taking the graph union.

We remove stub ASes because they likely have little real-world influence on traffic flow in the ISP graph [48].

Using this hybrid graph, we investigated the correlation between ISP wickedness and a number of popular measures of graph topology [14]. In total we tested eight different measures.

Figure 2 shows the correlation between wickedness and the six of the eight topological features we tested. Three features are significantly correlated with wickedness throughout the study period (top two panels, and middle right panel): an ISP's location within the Internet hierarchy(Core Number and Average Shortest Path Length) and centrality (weighted degree). Weighted degree is correlated for the majority of time steps (middle right panel), excluding the early part of the time series, a few weeks in 2010 and 2011, and late in the data. By contrast, betweenness centrality and clustering coefficient do not show significant correlation throughout the time series, while page rank is only correlated roughly one third of the time. The correlations that do exist show that in general ISPs with high centrality (degree), tend to have low wickedness, while ISPs on the periphery of the network (low core number, high average shortest path length), have higher wickedness values. It is not clear why this is the case; one possibility is that ISPs on the edge of the network tend to be smaller and thus have fewer resources to counter infections.

ASNs are often categorized based on the type of services they provide [11], and this could influence the amount of wickedness present in a given ISP. We did not include this factor in our analysis because each ISP could be an aggregation of multiple ASNs, and consequently, clear categorizations of the services provided by an ISP are difficult to ascertain. Moreover, since our data on subscriber numbers is at the ISP level, we cannot easily allocate it to different ASNs.

3.4 Network Traffic Dynamics

Traffic dynamics affect the concentration of malicious hosts [25], but appropriate network traffic datasets are not publicly available. Numerous models of traffic flow have been proposed for the AS network, ranging from simple [64] to elaborate [5], and for this study we adapted Roughan et al.'s gravity model [64] to simulate malicious traffic between nodes in the ISP graph. In the gravity model, the traffic received by node i from j is expressed as:

$$r_{ij} = \frac{C_i C_j}{d_{ij}^2} \tag{2}$$

where C_i is the number of customers for ISP i, and d_{ij} is the shortest path length between the two ISPs in the ISP graph. We assume that malicious traffic is proportional to the total traffic received by an ISP i, and then calculate the expected per customer rate of malicious traffic:

$$R_i = \frac{\sum_{j \neq i} R_{ij} W_j}{C_i} \tag{3}$$

where W_j is the concentration of spam-emitting IP addresses at ISP j and R_{ij} is fraction of j's traffic destined for i (normalized r_{ij}). Normalizing by C_i allows us to interpret R_i as the expected fraction of malicious traffic received by each customer of ISP i.

We test whether this calculated value correlates with wickedness in the same way we did for the topological factors, except we consider time by introducing a one time-step lag between the two series. This allows us to identify possible *causal* relationship between traffic and wickedness[22], as shown in the top panel of Figure 3. The figure shows that there is a statistically significant positive correlation through time. This indicates that the flow of malicious traffic, in particular, the amount of malicious traffic received per customer in the previous week, correlates with increased wickedness in the next week.

4. MODELING

In the previous section we identified external factors that are individually correlated with wickedness. In this section, we develop an autoregressive model that incorporates and combines these factors. We then use the model to explore the relative strengths of these effects and identify the transitions between spam eras.

4.1 Autoregressive Model

An *autoregressive* model is a type of linear regression, which uses previous values in a time series to predict future values. We have already discovered in subsection 2.2 that our dataset is highly autocorrelated, which justifies this model selection, and we include the external risk factors identified in section 2.

Visual inspection of the data reveals an obvious drop-off in wickedness levels somewhere after 2010. We incorporated this observation into the model by hypothesizing up to three distinct temporal eras. In each era y, the wickedness of ISP i at time t is modeled as:

$$\begin{aligned}\ln(W_{i,y}(t)) = {} & \beta_{0,y}\ln(W_i(t-1)) + \beta_{1,y}\ln(R_i(t-1)) + \\ & \beta_{2,y}\ln(G_i(t-1)) + \beta_{3,y}\ln(E_i(t)) + \\ & \beta_{4,y}P_i(t) + \beta_{5,y}\ln(D_i(t)) + \epsilon_y\end{aligned} \tag{4}$$

Each symbol in Equation 4 is described in Table 1. In section subsection 3.3 we found that both average shortest path length and core number are correlated with wickedness. However, these two measures are highly correlated with each other, and including both metrics in the model could cause estimates of $\beta_{x,y}$ to be incorrect [79], so we selected average path length.

All autoregressive models include a distribution of error terms, here represented by ϵ, and they are usually assumed to be normally distributed [79]. In our case, given the high variance of the data (section 2), we assume $\epsilon_y \sim T(\nu,\sigma)$, where $T(\nu,\sigma)$ is the non-standardized Student's T distribution, which is considered to be more appropriate when a dependent variable has high variance [79] which we observed in section 2.

In the model some variables are log transformed because preliminary inspection revealed that their functional relationships were non-linear in particular ways (i.e. roughly linear on log/log plots).[5]

4.2 Identifying Model Transitions

In section 1 we noted that the data experiences several possible qualitative changes, and that they may correspond to changes in spam tactics or the development of new spam fighting tools. However, it is unclear exactly when these changes might have

[4] http://irl.cs.ucla.edu/topology/ipv4/daily/.

[5] We speculate that the log/log relationship between $W_i(t)$ and $W_i(t-1)$, may arise from an underlying growth or decay process in malware infected hosts.

Table 1: Coefficients for the autoregressive model. Range indicates the range of possible values for each variable. **Bold** coefficients are statistically significant at the $p<0.01$ level.

Variable	**Symbol**	$\boldsymbol{\beta_{i,y}}$	**Range**	**Era 1** Jan 2005-Dec 2010	**Era 2** Dec 2010-June 2012	**Era 3** June 2012-Dec 2014
Log Prev Wickedness	$\ln(W_i(t\text{-}1))$	$\beta_{0,y}$	[-18.1,0.26]	**0.994**	**0.991**	**0.976**
Log Prev Wicked Traffic	$\ln(R_i(t\text{-}1))$	$\beta_{1,y}$	[-29.4,-10.4]	0.0002	0.0003	**0.0145**
Log Prev Region Wickedness	$\ln(G_i(t\text{-}1))$	$\beta_{2,y}$	[-13.3,-2.1]	**-0.0039**	**-0.0158**	**-0.0188**
Log GDP per capita	$\ln(E_i(t))$	$\beta_{3,y}$	[6.5,11.7]	**-0.0080**	**-0.0255**	**-0.0359**
Shortest Path Length	$P_i(t)$	$\beta_{4,y}$	[0.0,8.2]	**0.0052**	0.0109	**0.0658**
Log Weighted Degree	$\ln(D_i(t))$	$\beta_{5,y}$	[2.7,4.9]	-0.00009	-0.0006	**0.0175**
			$\boldsymbol{R^2}$	0.985	0.975	0.937

occurred. Rather than pre-define transitions between eras based on industry reports or qualitative evaluations of the data, we used the model to determine the most likely dates when significant changes in spam concentrations occurred, testing for zero, one, or two significant transitions.

For each possible combination of two transition dates, we use maximum likelihood estimation (MLE) to estimate the values for all $\beta_{x,y}$ and their standard errors. We then selected transition dates which gave the model the highest likelihood.

To measure whether dividing the data into three eras is justified, we compared the model to one with a single division into two eras, and one with no divisions. We used the Akaike Information Criteria (AIC)[79], which is a measure of goodness of fit based on likelihood that penalizes more complex models. We found a statistically significant improvement between the model with two divisions and models with a single or no divisions. It is also possible that there are more statistically significant transitions in the data than we were able to test for due to computational constraints. We leave this topic for future investigation.

The first change identified by our methodology begins in December 2010, after which we see a steady decline in spam levels. This may be due to the increasing efficacy of adaptive, real-time filtering, although filtering systems were first deployed at companies such as Google as early as 2006 [71]. There is evidence that improved filtering forced spammers to deploy new more costly methods of spamming, such as large-scale account hacking [21]. Filtering even impacted delivery of legitimate bulk email in the first half of 2011 [59]. Microsoft's Security Intelligence Report attributes the decline in 2011 to both more sophisticated filtering techniques, and to the takedown of the Cutwail and Rustock botnets [50].

We identify a second transition beginning in June 2012. In May, 2011 Kanich et al. published a paper which identified a handful of banks that were responsible for processing most of the payments made by spam victims [29]. Shortly after the paper was published, Visa tightened requirements for merchants, and effectively disrupted many spammers' revenue streams [76]. Seven months after the announcement of these requirements, spammers reported difficulty maintaining reliable credit card processing [41] and spam volume dropped significantly, e.g. Symantec's Internet Security Threat Report from 2012 notes a significant drop in pharmaceutical spam [68].

4.3 Model Results

Table 1 gives the MLE values for the $\beta_{x,y}$. Examining Table 1 we see that the autoregressive term has the largest influence on future wickedness. Surprisingly, one of the other terms (regional wickedness during the previous week) in all eras has an opposite effect from what was reported in section 3 (Figure 3). This is an example of *Simpson's Paradox* [63], indicating that in the presence of other variables, high levels of wickedness in neighboring ISPs actually reduce future wickedness. One possible explanation is that spammers initially try to infect as many machines in a region as possible, and then concentrate on vulnerable ISPs as they discover them, reducing attacks on the less vulnerable ISPs. This factor and the other variables identified in section 2 are statistically significant, but at low levels. This simple model accounts for the vast majority of the variance in our data, with a combined coefficient of determination of $R^2=0.980$ for data in all eras[6]. It is possible that more sophisticated models might provide more predictive power than our simple linear, autoregressive model. We tested support vector machines, feed forward neural networks, decision tree regression, and gradient tree boosting, and found that none outperformed our model (measured by R^2) or had similar explanatory power. Moreover, our robust statistical approach can determine statistical significance without computationally expensive procedures, such as cross validation.

5. THE EFFECT OF TAKEDOWNS

Section 4 presented a statistical model that accurately assesses the relative contribution of a variety of factors on spam levels over almost a decade. This section shows how the model can be used to study the impact of interventions such as botnet takedowns.

Although spam levels typically drop immediately following a takedown, there is anecdotal evidence that this effect is short-term, often returning to previous levels within a few weeks [43, 75, 62]. Given the high variance in the data, however, quantifying the short-term and long-term effects is challenging, and requires rigorous statistical testing. With only a small extension to the model, we can conduct such tests and consider the impact of takedowns on different regions of the world.

5.1 Modeling Takedowns

We model takedowns, which are a discrete event at the timescale of our data, by adding binary variables to the model:

$$B_k(t-j)=\begin{cases} 1 & \text{takedown } k \text{ occurred } j \text{ weeks ago} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Each $B_k(t-j)$ is incorporated into the model with its own coefficient, and the autoregressive model becomes:

$$\mathit{Equation}\ 4+\sum_k \sum_{j=0}^{l} \beta_{kj} B_k(t-j) \tag{6}$$

β_{kj} is the coefficient associated with $B_k(t-j)$. Using the log/linear form of Equation 6, we can estimate the general effect of a takedown using the estimates of β_{kj}. For each takedown, the fractional change in wickedness associated with the takedown during week j is $e^{\beta_{kj}}-1$. This process can be repeated to give $e^{\sum_{j=0}^{l}\beta_{kj}}-1$, which estimates the cumulative effect of the takedown over the time period l. If the MLE of any one of the β_{kj} is

[6] The autoregressive term accounts primarily accounts for the high R^2 in the model. However, without the autoregressive term the model still has an $R^2=0.58$, indicating moderate explanatory power.

Table 2: Effect of 12 historical botnet takedowns in the model. The recorded dates are the first date in our data set after the intervention. In column two *Communications Disruption* is the severing of communication between bots and the command and control (C&C) infrastructure, *C&C Takeover* refers to when control of the C&C infrastructure is gained without physical access, *Seizure* refers to the physical confiscation of C&C infrastructure, and *Arrest* refers to the arrest of individuals. The percent change columns report the percent change in global wickedness in the first week after the takedown (column three) and six weeks later including the first week of the takedown (column 4).

Botnet takedown (Date)	**Takedown Method**	**Initial % Change**	**6 Week % Change**
McColo (November 11, 2008)	Communication Disruption [35]	-17.4	44.6
Mariposa (December 24, 2009)	C & C Takeover [9]	35.8	34.8
Waledac (March 5, 2010)	Communication Disruption [38]	Not significant	-3.5
Spamit.com (October 1, 2010)	Self Shutdown [77]	Not significant	6.1
Bredolab/Spamit.com (October 29, 2010)	Seizure and Arrest [16]	-11.8	-17.2
Rustock (March 19, 2011)	Seizure [31]	-20.2	-13.9
Coreflood/Rustock (April 16, 2011)	Communications Disruption [40]	-7.3	13.8
Kelihos (September 17, 2011)	Communication Disruptions [20]	6.4	31.6
Kelihos Variant (April 1, 2012)	Communications Disruption [12]	Not Significant	30.1
Hermes-Carberp (June 24, 2012)	Arrest [57]	21.4	9.0
Grum/Hermes-Carberp (July 22, 2012)	Communications Disruption [52]	-11.3	49.4
Virut (January 22, 2013)	Communications Disruption [42]	-21.7	113.8

not statistically significant it is assumed to be 0. The statistical significance of the estimated coefficients provides a rigorous test of a takedown's effect.

We incorporated 12 different historical botnet takedowns into the extended model. We considered most major takedowns of botnets in the time span of our dataset that sent large amounts of spam. We allow i to vary from 0 (the week of the takedown) to $l = 6$ weeks. Beyond this time, we find no further statistically significant changes that can be attributed to the takedown, implying that the time horizon for the effect of a takedown is at most six weeks. In some cases, two botnet takedowns overlap the six-week windows, and we cannot separate the effect of the two takedowns.[7]. When this occurs we include both the initial effect of the first takedown and the combined effect of the second takedown.

The results are given in Table 2, which shows that the global effectiveness of these botnet takedowns varies significantly. Some takedowns were effective in the short run (6 out of 12), but over the six-week window only three of those showed any persistent significant decrease in spam.

The table shows that two takedowns (Bredolab and Rustock) had a relatively large long-term impact on spam in the six weeks following the takedown, while the third (Waledac) had a relatively minor impact. Both the Bredolab and Rustock takedowns involved physical seizure of offending infrastructure by law enforcement. Although this may not be directly related to the effectiveness of the takedowns, it is notable and is likely correlated with other external factors that have more lasting effect. Four takedowns that used communications disruption to shutdown the botnet showed a positive short-term impact (i.e. McColo [35], Coreflood [40], Grum [52], and Virut [42]) are followed by long-term increases in wickedness. The rest of the takedowns, such as the self shutdown of spamit.com [77], seemed to have little positive impact either initially or in the long-term. These values provide evidence that other interventions were likely the main driver of the decline in overall spam volumes, not botnet takedowns. We note that the two effective takedowns occurred at the end of era 1 and the beginning of era 2 respectively, however, without more data it is impossible to draw further conclusions about the relationship between takedown effectiveness and the era in which they occurred.

In the case of Mariposa, our results may reflect the historic details of the takedown. Shortly after the original takedown in December, during which control of command-and-control servers was obtained, attackers managed to regain control of the botnet and launched denial-of-service attacks against numerous ISPs [9], which may be related to the increased spamming activity.

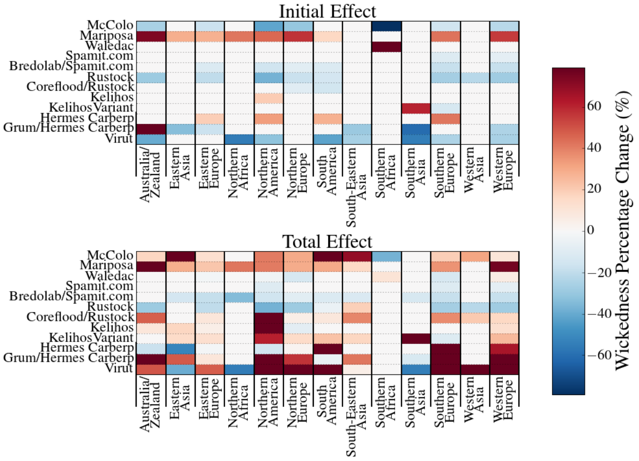


Figure 4: Regional effect of botnet takedowns. For each historical takedown studied the top panel shows the immediate effect by geographic region, and the bottom panel shows the effect after six weeks for the same geographic regions. The color shows the percent change in wickedness as indicated by the legend.

5.2 Regional effects of botnet takedowns

Bots are not uniformly distributed geographically [53], suggesting that takedowns might have different effects throughout the world. To investigate this hypothesis we re-applied our modeling approach, but at the regional level. Rather than creating a single model for all ISPs globally, we constructed one model for each geographic region defined in subsection 3.2, using only the ISPs in that region. We included regions that have at least two ISPs in our dataset to avoid over-fitting [79].

All takedowns showed varying effects for different regions (Figure 4). Some takedowns have effects regionally that resemble the global effect seen in Table 2, while others have differentiated behavior. For example, the McColo takedown initially appears successful, but in the long term wickedness increases across nearly

[7] An overlap results in two binary variables with the same value being included in the model (perfect collinearity), which would cause an ill-defined maximum likelihood calculation [79]

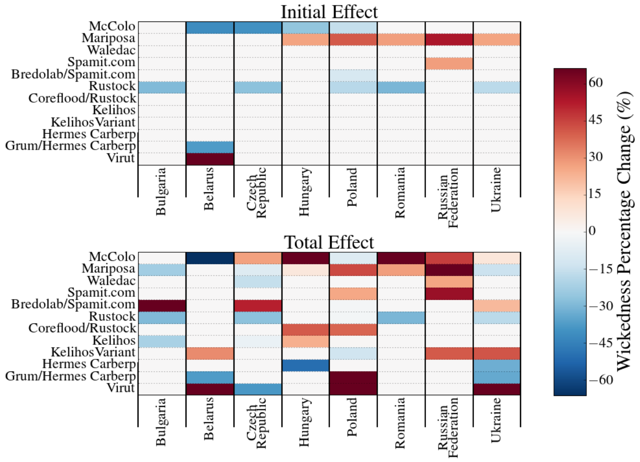


Figure 5: Country-specific effect of botnet takedowns in Eastern Europe. For each historical takedown studied, the top panel shows the immediate effect for each country, and the bottom panel shows the effect after six weeks for the same country. The color shows the percent change in wickedness as indicated by the legend.

all regions (blue colors, upper panel in Figure 4, and red colors, lower panel in Figure 4, respectively), similar to the global effect. In contrast, two of the takedowns led to mixed effects throughout the world. Six weeks after the Hermes Carberp takedown, wickedness in Australia/New Zealand, Eastern Asia, and South-Eastern Asia decreased, but most other regions experienced increases. Similarly, six weeks following the Grum takedown, wickedness in South America had declined significantly, but the rest of the world experienced increases. These differentiated regional effects occur predominately in the second and third eras.

We can further analyze the effect of botnet takedowns on individual countries by constructing one model for each country, using the same procedure as we did for regions. Once again, we consider only countries with more than two ISPs in our dataset to avoid overfitting. Figure 5 shows the effect of various takedowns only on countries in Eastern Europe due to space constraints. We focus on Eastern Europe because it shows interesting variation among its countries. However, most other regions also showed significant variation.

Consistent with the earlier analyses, there are many countries for which a takedown initially has a positive effect, but where, in the long term, wickedness actually increases. One prominent example is the Czech Republic following the Bredolab/ Spamit.com takedowns, which saw a no significant decrease in wickedness the week of the takedown, but a near doubling in the following six weeks. Country-by-country there is little correspondence to the global takedown effect. For example, the McColo takedown saw initial success followed by global increases in spam on both a global and regional level. However, at the country level the results are mixed, with Belarus benefiting from the takedown while Romania, Hungary, and Russia experience increases at 6 weeks.

These regional results raise the interesting possibility that botnets can migrate in response to takedowns. That is, by reducing the number of infected hosts in one region, a takedown creates incentives for botnets to find new vulnerable hosts, thus moving the problem elsewhere. In this work we do not investigate this issue further; more advanced modeling techniques, such as vector autoregressive models [2], could shed light on this intriguing possibility.

6. RELATED WORK

This paper builds on the dataset of Van Eeten et al. [73], which investigated ISPs as control points for mitigating the spread of malware, using a comprehensive worldwide spam dataset. Here, we updated the dataset with 6 more years of data. The Van Eeten et al. analysis revealed that a country's development level is correlated with spam volume, and it analyzed how public policy initiatives might reduce infections. We extend this work by developing a data-driven statistical model, which the effect of different spam interventions and identifies temporal transitions in the dataset.

Other work locates infected hosts in IP address space. Moura et al. identified IP ranges with high concentrations of spam sending hosts [53]. Similarly Ramachandran et al. examined the network-level behavior of spammers, and showed that spam is concentrated in relatively small IP ranges [61]. Stone-Gross et al. studied ISPs with persistent malicious behavior [67], Chen et al. investigated malicious sources on the Internet over IPv4 [6], and Wilcox et al. studied the stability and availability of address space in spam and non spam networks [78]. Kokkodis and Faloutsos showed that spamming botnets have become more widely and thinly spread over IP space, a potential problem for filtering [33]. However, to our knowledge none of this work explores which topological features of the AS network correlate with infected hosts. Additionally, our model shows that previous regional concentrations of wickedness and malicious traffic correlate with future wickedness.

Collins et al. define *uncleanliness* as the probability that a host is vulnerable [8], while wickedness measures the concentration of active malicious hosts. They find that a network's past behavior is strongly correlated with its future behavior, which agrees with our finding that wickedness is autocorrelated.

Another related area proposes using economics to control malware and spam [28, 55, 44]. The idea of disrupting spammers' income by targeting the small number of banks that handle credit card payments [49, 29] may have helped reduce global spam levels. A related approach is the publication of infection rates of ISPs (measured by spam volumes) to provide incentives for them to control compromised customers in their networks [70].

There are few models of global malware dynamics. Venkataraman et al. model malicious activity as a decision tree over IP address space and infer the dynamics of the decision tree [74]. Their work focuses on IP address ranges rather than ISPs, but it reports some similar results as those observed in our model, for example, high variance in the data. Zhang et al. find that mismanagement of networks correlates with malicious behavior (measured using a quantity similar to our *wickedness*) in Autonomous Systems[82], but do not focus on how this behavior might evolve over time. Liu et al. use support vector machines trained on data from reputation blacklists to predict security incidents [46]. These predictions could be incorporated into our model to better predict some of the large changes in wickedness over time. A model of global malware dynamics was also proposed by Hofmeyr et al., which used an agent-based model to investigate the dynamics of malicious traffic flowing across the Internet at the AS level [25]. This model was significantly more abstract than ours, and did not incorporate actual data about spam, ISPs, demographic features, or intervention events such as takedowns.

Nadji et al. analyze botnet takedown efficacy [54], and other work considers raw measurements of spam volume [20]. Nadji et al. investigated three historical takedowns, performing post mortem analysis of each takedown's effectiveness, by measuring which malicious domains could still be resolved in the Domain Name Service (DNS). Contrary to our results, this work recommends DNS takedowns for a large fraction of current botnets. However, their results rely on relatively short time scales (two weeks), and it only considers the DNS, which may not be sufficient to identify rebounds once attackers establish new communication channels [18].

Mechanistic botnet models, e.g. [34, 80, 10, 7, 30], focus on specific infection mechanisms, while our model considers the security problem from a global perspective, with botnets being just one component. We find that most botnet takedowns have limited and transient impacts on global wickedness. This result agrees with other research, which found that botnets are surprisingly resilient [81], and in many cases they recover after a short time [54]. Other work has modeled malicious websites, noting the high variance of cybersecurity data, and investigates interventions through modeling [13].

General traffic filtering is an important intervention for reducing the number of infected hosts. There has been research into the effectiveness of various filtering techniques, e.g. [45, 27, 60], however this work focuses on the success of the filter itself and not whether the filter actually reduces the global distribution of infected hosts. Incorporating filtering interventions into our model is an area we plan to explore in future work.

7. DISCUSSION

Data-driven models such as the one presented here can potentially yield interesting and important insights, which in turn can inform policy makers about the utility of interventions or even how to prepare vulnerable regions of the world before they are applied. However, there are several pitfalls that any modeling approach needs to heed.

First, the model is built around statistical correlations, but it ignores precise mechanisms, e.g. by what process does a country's development and an ISP's position in the ISP network influence wickedness? Second, statistical models such as ours cannot determine causality, so detailed understanding of the data is needed to attribute cause and effect. Third, high variance data can hide significant changes, and also make it appear that significant change has occurred when it has not. Modeling global data is a powerful tool to address this issue, but the modeling methodology must take into account the variance (e.g., averages can be misleading). We are careful to use appropriate methodology to avoid this pitfall. Finally, any conclusions drawn from a statistical model depend on the quality of the data (although techniques do exist to help compensate for certain classes of data problems).

Any model is necessarily a simplification of reality. For example, our traffic model is simplistic given the complexity of the Internet. Future work could incorporate more realistic models, especially because our model shows that the traffic component is significant only during the third era. It could be that spam email was more likely to be used to spread infection during this era, whereas earlier it was used primarily for advertising, e.g. gray market pharmaceuticals. Similar to the traffic component, if other important features are identified, such as the type of service provided by an ISP, this information could easily be included in the model.

This paper focused on spam itself, but spam data have also been used to estimate the numbers of infected PCs [83, 70, 26]. By applying our methodology to other measures of infection, it should be possible to develop models that provide insight into the dynamics and global distribution of these other types of infections. In general, we are interested in the distribution of all malicious behavior (or wickedness), regardless of its source. In some cases, the definition of wickedness could be expanded, e.g to include the relative value of hosts in different regions—an infected machine in the US may be more valuable than one in India.

Cybersecurity is often viewed as an arms race, which complicates the task of predicting the impact of today's interventions against tomorrow's attackers. At least, however, we should evaluate the likely effect of new methods before embracing large-scale deployments or policy directives that enforce certain interventions, and models such as the one described here are one way to approach this.

We have studied the impact of botnet takedowns in some detail, but there are other interventions that would also be interesting to explore. For example, the traffic model provides a way to analyze the effect of blacklisting offending ISPs or different filtering strategies [24]. There is evidence that national and international initiatives against cybercrime can reduce wickedness [73]. Our model could, for example, be used to assess whether countries that are signatories to agreements such as the London Action Plan or Council of Europe's Convention on Cybercrime actually experience lower wickedness levels after ratifying the agreements. This could be studied by incorporating this information as an additional variable.

By looking for differential effects of takedowns geographically, we can identify *at risk* ISPs or countries, i.e. those that are likely to see little initial effect from the takedown but which could expect an increase in wickedness in the medium term. Our results to date have not identified any single factor that is consistently correlated (at a statistically significant level) with increased wickedness after a takedown. However, if we could identify at-risk countries and ISPs, they might make good candidates for targeted interventions, for example, ISPs on the periphery of the AS network which may have inadequate spam-fighting resources and lack automated methods to help customers clean up malware. Government interventions could focus on providing resources to those ISPs (or even countries), an approach that might prove more cost-effective than existing methods.

8. CONCLUSION

With unprecedented numbers of people now connected to and depending on the Internet (three billion in 2014 [72]), it is imperative that we understand and mitigate global cybersecurity threats. Further, we need to understand regional variations, and why some parts of the world and some corners of the Internet are disproportionately affected.

In this paper we studied an abstract quantity called *wickedness* (concentration of spam sending hosts) and showed that it clusters regionally, correlating with national demographics and certain properties of the ISP graph. Through the use of statistical modeling combined with a large dataset, we studied some of the factors affecting spam, a large-scale security problem distributed around the world. Leveraging a long-term historical view of data produced interesting insights about the effectiveness of certain cybersecurity interventions. We found that takedowns are only marginally effective in many cases, and in fact may be harmful to certain countries and ISPs.

Our model could serve as a starting point to predict future wickedness and test the likely effect of new interventions, both for spam and other similar problems. Our ultimate goal is to provide researchers and policy makers objective means to test intervention strategies and decide how best to mitigate global wickedness.

9. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the partial support of NSF CNS 1444500, DARPA (P-1070-113237), DOE (DE-AC02-05CH11231) and the Santa Fe Institute. The authors thank Hadi Asghari for preparing the data and providing comments on drafts of this work. BE and SF also thank Anil Somayaji, Robert Axelrod and members of the Adaptive Computation Laboratory for reading earlier drafts of this paper and providing helpful comments.

10. REFERENCES

[1] R. Anderson et al. Measuring the cost of cybercrime. In *WEIS*, 2012.

[2] D. Asteriou and S. G. Hall. *Applied Econometrics: a modern approach using eviews and microfit.* Palgrave Macmillan New York, 2007.

[3] W. Bank. World bank data. http://data.worldbank.org/, Mar. 2015.

[4] BotFrei. botfrei.de: The anti-botnet advisory centre. https://www.botfrei.de/, May 2014.

[5] H. Chang et al. An empirical approach to modeling inter-as traffic matrices. In *Proc. of ACM IMC*. USENIX Association, 2005.

[6] Z. Chen et al. Spatial-temporal characteristics of internet malicious sources. In *INFOCOM*. IEEE, 2008.

[7] C. Y. Cho et al. Inference and analysis of formal models of botnet command and control protocols. In *ACM CCS*. ACM, 2010.

[8] M. P. Collins et al. Using uncleanliness to predict future botnet addresses. In *ACM IMC*, 2007.

[9] L. Corrons. Mariposa botnet. *Panda Labs*, Mar. 2010.

[10] D. Dagon et al. Modeling botnet propagation using time zones. In *NDSS*, 2006.

[11] A. Dhamdhere and C. Dovrolis. Ten years in the evolution of the internet ecosystem. In *ACM IMC*, 2008.

[12] B. Donohue. Kapersky knocks down kelihos botnet again, but expects return. *Threatpost.com*, Mar. 2012.

[13] B. Edwards et al. Beyond the blacklist: Modeling malware spread and the effects of interventions. In *NSPW*, 2012.

[14] B. Edwards et al. Internet topology over time. *arXiv preprint arXiv:1202.3993*, 2012.

[15] B. Edwards et al. Hype and heavy tails: A closer look at data breaches. In *WEIS*, 2015.

[16] T. Espiner. Dutch police take down bredolab botnet. *ZDNet*, Oct. 2010.

[17] D. Geer. Cybersecurity as realpolitik. http://geer.tinho.net/geer.blackhat.6viii14.txt, Aug. 2014.

[18] S. Goldberg and S. Forrest. Implications of security enhancements and interventions for core internet infrastructure. In *TPRC42*, 2014.

[19] D. Goodin. Waledac botnet 'decimated' by ms takedown. *The Register*, Mar. 2010.

[20] D. Goodin. "slain" kelihos botnet still spams from beyond the grave. *arstechnica*, Feb. 2012.

[21] Google. An update on our war against account hijackers. *Google Official Blog*, Feb. 2013.

[22] C. W. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 1969.

[23] D. Gudkova. Kaspersky security bulletin: Spam evolution 2013, Jan. 2013.

[24] S. Hird. Technical solutions for controlling spam. *Proc. of AUUG*, 2002.

[25] S. Hofmeyr et al. Modeling internet-scale policies for cleaning up malware. In *Economics of Information Security and Privacy III*. Springer, 2013.

[26] B. Johnson et al. Metrics for measuring isp badness: The case of spam. In *Financial Cryptography and Data Security*. 2012.

[27] P. Kalakota and C.-T. Huang. On the benefits of early filtering of botnet unwanted traffic. In *ICCCN*, 2009.

[28] C. Kanich et al. Spamalytics: An empirical analysis of spam marketing conversion. In *ACM CCS*, 2008.

[29] C. Kanich et al. Show me the money: Characterizing spam-advertised revenue. In *USENIX*, 2011.

[30] A. Karasaridis et al. Wide-scale botnet detection and characterization. In *Proc. of HotBots*. Cambridge, MA, 2007.

[31] G. Keizer. Rustock take-down proves botnets can be crippled, says microsoft. *Computer World*, July 2011.

[32] M. Kendall et al. Rank correlation methods. *Rank correlation methods.*, 1948.

[33] M. Kokkodis and M. Faloutsos. Spamming botnets: Are we losing the war. *Proc. of CEAS*, 2009.

[34] I. Kotenko et al. Agent-based modeling and simulation of botnets and botnet defense. In *Conf. on Cyber Conflict*, 2010.

[35] B. Krebs. Host of internet spam groups is cut off. *The Washington Post*, Nov. 2008.

[36] B. Krebs. Organized crime behind a majority of data breaches. *The Washington Post*, Apr. 2009.

[37] B. Krebs. The scrap value of a hacked pc. *The Washington Post*, May 2009.

[38] B. Krebs. Microsoft ambushes waledac botnet, shutters whistleblower site. *Krebs on Security*, Feb. 2010.

[39] B. Krebs. Takedowns: The shuns and stuns that take the fight to the enemy. *McAfee Security Journal*, 2010.

[40] B. Krebs. U.s. government takes down coreflood botnet. *Krebs on Security*, 2011.

[41] B. Krebs. Rogue pharma, fake av vendors feel credit card crunch. *Krebs On Security*, Oct. 2012.

[42] B. Krebs. Polish takedown targets 'virut' botnet. *Krebs On Security*, Jan. 2013.

[43] J. Leydon. Google: Botnet takedowns fail to stem spam tide, Apr. 2010.

[44] Z. Li et al. Botnet economics: uncertainty matters. In *Managing Information Risk and the Economics of Security*. 2009.

[45] X. Liu et al. To filter or to authorize: Network-layer dos defense against multimillion-node botnets. In *ACM SIGCOMM*, 2008.

[46] Y. Liu et al. Predicting cyber security incidents using feature-based characterization of network-level malicious activities. In *Proc. of ACM IWSPA*, 2015.

[47] M. LLC. Maxmind geoip, 2008.

[48] R. Mahajan et al. Understanding bgp misconfiguration. In *SIGCOMM*, 2002.

[49] D. McCoy et al. Priceless: The role of payments in abuse-advertised goods. In *Proc. of ACM CCS*, 2012.

[50] Microsoft. Microsoft security intelligence report, Aug. 2011.

[51] T. Moore and R. Clayton. Discovering phishing dropboxes using email metadata. In *Proc. of eCrime*, 2012.

[52] T. Morrison. Spam botnets: The fall of grum and the rise of festi. *SpamHaus Blog*, Aug. 2012.

[53] G. C. Moura. *Internet bad neighborhoods*. Number 12 in University of Twente Dissertation. Giovane Cesar Moreira Moura, 2013.

[54] Y. Nadji et al. Beheading hydras: performing effective botnet takedowns. In *SIGSAC*. ACM, 2013.

[55] Y. Namestnikov. The economics of botnets. *Analysis on Viruslist.com, Kapersky Lab*, 2009.

[56] U. Nations. Un geographic division. http://millenniumindicators.un.org/unsd/methods/m49/m49regin.htm, Oct. 2013.

[57] S. F. News. A young botnet suspect arrested by russian authorities. *SpamFighter*, July 2012.

[58] R. Oliveira et al. The (in) completeness of the observed internet as-level structure. *IEEE/ACM Transactions on Networking (ToN)*, 2010.

[59] R. Path. The global email deliverability benchmark report, 2h 2011, Mar. 2012.

[60] A. Ramachandran et al. Filtering spam with behavioral blacklisting. In *ACM CCS*, 2007.

[61] A. Ramachandran and N. Feamster. Understanding the network-level behavior of spammers. In *ACM SIGCOMM*. ACM, 2006.

[62] F. Y. Rashid. Grum botnet: Down one month, no impact on spam, Aug. 2012.

[63] Y. Rinott and M. Tam. Monotone regrouping, regression, and simpson's paradox. *The American Statistician*, (2), 2003.

[64] M. Roughan. Simplifying the synthesis of internet traffic matrices. *ACM SIGCOMM*, (5), 2005.

[65] F. Roveta et al. Burn: Baring unknown rogue networks. In *Proc. of VizSec*, 2011.

[66] SpamHaus. Spamhaus composite blocking list. http://cbl.abuseat.org/totalflow.html, May 2015.

[67] B. Stone-Gross et al. Fire: Finding rogue networks. In *ACSAC*, 2009.

[68] Symantec. 2012 internet security threat report, Apr. 2013.

[69] Symantec. 2014 internet security threat report, Apr. 2015.

[70] Q. Tang et al. Reputation as public policy for internet security: A field study. In *Proc. of ICIS*, 2012.

[71] B. Taylor. Sender reputation in a large webmail service. In *CEAS*, 2006.

[72] I. T. Union. Ict facts and figures. Technical report, International Telecommunications Union, May 2014.

[73] M. van Eeten et al. The role of internet service providers in botnet mitigation: An empirical analysis based on spam data. Technical report, OECD Publishing, 2010.

[74] S. Venkataraman et al. Automatically inferring the evolution of malicious activity on the internet. In *NDSS*, 2013.

[75] B. Violino. Spam levels creep back after rustock botnet takedown, Apr. 2011.

[76] Visa. Visa international operating regulations summary of changes, Oct. 2011.

[77] S. Walsh. Canadian pharmacy spam group reinvents self as "world pharmacy". *All Spammed Up*, Dec. 2010.

[78] C. Wilcox et al. Correlating spam activity with ip address characteristics. In *INFOCOM WKSHPS*, 2010.

[79] C. Wong et al. A student t-mixture autoregressive model with applications to heavy-tailed financial data. *Biometrika*, (3), 2009.

[80] P. Wurzinger et al. Automatically generating models for botnet detection. In *ESORICS 2009*. 2009.

[81] T.-F. Yen and M. K. Reiter. Revisiting botnet models and their implications for takedown strategies. In *Principles of Security and Trust*. 2012.

[82] J. Zhang et al. On the mismanagement and maliciousness of networks. In *NDSS*, 2013.

[83] L. Zhuang et al. Characterizing botnets from email spam records. *LEET*, 2008.