The role of hosting providers in fighting command and control infrastructure of financial malware

Tajalizadehkhoob, Samaneh; Hernandez Ganan, Carlos; Noroozian, Arman; van Eeten, Michel

DOI
10.1145/3052973.3053023

Publication date
2017

Document Version
Peer reviewed version

Published in
Proceedings of the 2017 ACM Asia Conference on Computer and Communications Security -ASIA CCS 2017

Citation (APA)
Tajalizadehkhoob, S., Hernandez Ganan, C., Noroozian, A., & Van Eeten, M. (2017). The role of hosting providers in fighting command and control infrastructure of financial malware. In Proceedings of the 2017 ACM Asia Conference on Computer and Communications Security -ASIA CCS 2017 (pp. 575-586). Association for Computing Machinery (ACM). DOI: 10.1145/3052973.3053023

Accepted Author Manuscript of http://dx.doi.org/10.1145/3052973.3053023

The Role of Hosting Providers in Fighting Command and Control Infrastructure of Financial Malware

Samaneh Tajalizadehkhoob, Carlos Gañán, Arman Noroozian, and Michel van Eeten
Faculty of Technology, Policy and Management, Delft University of Technology
Delft, the Netherlands
s.t.tajalizadehkhoob@tudelft.nl

Keywords

Hosting providers, financial malware, modeling abuse

ABSTRACT

A variety of botnets are used in attacks on financial services. Banks and security firms invest a lot of effort in detecting and combating malware-assisted takeover of customer accounts. A critical resource of these botnets is their command-and-control (C&C) infrastructure. Attackers rent or compromise servers to operate their C&C infrastructure. Hosting providers routinely take down C&C servers, but the effectiveness of this mitigation strategy depends on understanding how attackers select the hosting providers to host their servers. Do they prefer, for example, providers who are slow or unwilling in taking down C&Cs? In this paper, we analyze 7 years of data on the C&C servers of botnets that have engaged in attacks on financial services. Our aim is to understand whether attackers prefer certain types of providers or whether their C&Cs are randomly distributed across the whole attack surface of the hosting industry. We extract a set of structural properties of providers to capture the attack surface.

We model the distribution of C&Cs across providers and show that the mere size of the provider can explain around 71% of the variance in the number of C&Cs per provider, whereas the rule of law in the country only explains around 1%. We further observe that price, time in business, popularity and ratio of vulnerable websites of providers relate significantly with C&C counts. Finally, we find that the speed with which providers take down C&C domains has only a weak relation with C&C occurrence rates, adding only 1% explained variance. This suggests attackers have little to no preference for providers who allow long-lived C&C domains.

1. INTRODUCTION

Research into the disruption of botnets has mainly focused on two strategies: comprehensive takedown efforts of the command and control (C&C) infrastructure and the cleanup process of the infected end user machines (bots) [12,32,43]. The first strategy has the promise of being the most effective, taking away control of the botnet from the botmasters. In reality, however, this is often not possible. The second strategy is not about striking a fatal blow, but about the war of attrition to remove malware, one machine at a time. It has not been without success, however. Infection levels have been stable in many countries [5].

In practice, a third strategy is also being pursued. Similar to access providers cleaning up end user machines, there is a persistent effort by hosting providers to take down C&C servers, one at a time. This line of mitigation has been studied much less, perhaps because most botnets have been resilient to these efforts.

Could this strategy be made more effective? This depends on how attackers distribute their C&C domains. Do they randomly distribute them over many hosting providers? Or do they locate them predominantly in carefully selected providers, perhaps those who are negligent in terms of abuse handling or who offer bullet-proof services to actively support criminal activities [18,22,40]. Depending on the answer, there are different directions for improving mitigation.

This paper sets out to discover the strategies of attackers for the placement of their C&C servers across the hosting market. We focus on botnet families that have, in varying degrees, been used to attack financial services. Well-known examples are Zeus, Citadel and Dyre. These are widely understood to be among the most harmful botnets. Industry association M3AAWG has listed them as a top priority for abuse handling by providers [28]. This means that if providers do anything in terms of mitigation, it would be most visible for these botnet families. Put differently, if attackers care about the security practices of providers, we should see it first and foremost in the location of the C&C for these botnets.

Do attackers prefer providers with little or no abuse handling? Or are the C&C domains more or less randomly distributed across the overall attack surface of the hosting market? We study seven years of data on the location of C&Cs for 26 botnet families engaged in attacks against financial services.

We model the distribution of C&C domains across the overall landscape of the hosting market. Using several datasets for approximating the size and attack surface of providers, we can quantify the extent to which the number of C&C domains per provider can be explained as the outcome of a random selection process by attackers. We then analyze

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

whether there is a relation between the concentration of C&C in providers and the speed with which providers take down such domains.

Our contributions are as follows:

- We track the trends in the hosting locations of C&C for 26 different malware families that, to varying degrees, have been used in attacks on financial services. We find that, over time, C&Cs domains are spread out over more providers, diluting the concentrations of C&C;
- We model the distribution of C&Cs across providers and show that the mere size of the provider can explain around 71% of the variance in the number of C&Cs per provider, whereas the rule of law in the country only explains around 1%, suggesting a predominantly random selection process by the attackers for locating their C&C;
- Using a sample of hosting providers, we show that business model characteristics – such as pricing, popularity, time in business and the ratio of WordPress websites – all have a significant impact on the concentration of C&C domains;
- We demonstrate that there are statistically significant differences among providers in C&C takedown speed. Despite such differences, the take-down speed only has a weak relation with the concentration of C&Cs across providers, suggesting that attackers have little or no preference for hosting their domains in hosting providers that allow longer C&C uptime;

The remainder of this paper is organized as follows: Section 2 describes our data collection methodology. Section 3 provides a descriptive summary of our datasets and studies the concentrations of C&Cs in terms of malware and hosting types and across different geo-locations. Section 4 outlines a set of variables that capture different aspects of provider s' characteristics and next use them to model the C&C concentrations across providers. In this section we discuss our modeling approach and results at length. We then extend our model in Section 5 with taking the effect of provider take-down speed of C&C domains into account. Our finding are compared to the related work in Section 6. Finally, we discuss the main conclusions and limitations of our work in Section 7.

2. DATA COLLECTION METHODOLOGY

To understand the attacker's strategy for the placement of their C&C servers across the hosting market, we employ two types datasets: (i) data on C&C domains; and (ii) data on hosting providers. We first provide an overview of these datasets.

2.1 Command-and-Control Data

As stated earlier, we focus on C&Cs of botnets engaged, to varying degrees, in attacks on financial services. We make use of two datasets which in conjunction provide information on C&C domains located in 109 countries:

ZeusTracker: Provided by Roman Huessy from ZeusTracker [6], is a C&C panel tracker that contains meta data on C&C servers online at any point of time between 2009 and 2016 for the ZeuS malware family.

Private honeypots: Captured by a security company specialized in threat intelligence for banks and financial institutions using honeypots located all over the world, this dataset contains a list of botnet C&C domains from various botnets. Some of those botnets are predominantly used for attacks on financial services, like Citadel. Others are more generic malware families, but the security company has observed them as participating in attacks on financial services. The data is collected over a period of one year (2015Q1-2016Q1) using two methods: by running live malware samples and using honeypots.

The combined dataset contains 11,544 unique domain names associated with 8,528 IP addresses. A more detailed summary of our C&C data is shown in Table 1.

Table 1: C&C data summary

Year	# Domains	IP addresses	Families
2009	934	771	1
2010	1016	806	1
2011	1071	638	1
2012	1189	922	4
2013	1761	1365	3
2014	2188	1768	4
2015	3897	1819	28
2016	3718	969	34

2.2 Hosting Provider Data

The next steps towards studying the location of C&Cs is to attribute them to their responsible service providers. To that end, we need to reliably identify hosting providers. Most of the existing work towards identifying hosting providers use BGP routing data to map IP addresses to their respective Autonomous Systems (ASes). This approach essentially equates ASes with hosting providers. However, ASes do not perfectly correspond to hosting providers. The same AS can contain multiple providers and, reverse, the same provider can operate multiple ASes. We define providers as the entities that own the IP addresses in question, rather than the entities that route traffic to and from it. This is more fully explored in prior work [9, 39]. We use `WHOIS` data to identify the organizations to which IP addresses are allocated.

Our starting point is the IP addresses and domain names in DNSDB, a passive DNS database that draws upon hundreds of sensors worldwide and generously provided to us by Farsight Security [1, 2]. We use passive DNS data to populate a global list of domain names and IP addresses used for web hosting. To our knowledge, DNSDB has the highest coverage of the domain name space available to researchers. We map the IP addresses and domains in passive DNS data to their corresponding organizations using `WHOIS` data from the MaxMind database [3, 26]. Next, by adopting some of the keywords and categories adopted from the previous work [14, 39], we filter out non-hosting organizations (e.g., educational and governmental). The final set consists of 45,358 hosting providers, representing the population of hosting services from all over the world. A more detailed description of this method is discussed in [39].

3. CHARACTERIZING C&C CONCENTRATIONS

Given our C&C and hosting provider datasets, we can examine the distribution of C&C domains across different hosting providers to gain insight into attacker C&C placement strategies. Do they prefer certain hosting providers? Do they prefer certain locations? In this section, we provide a descriptive summary of our data and examine such different aspects of C&C concentration.

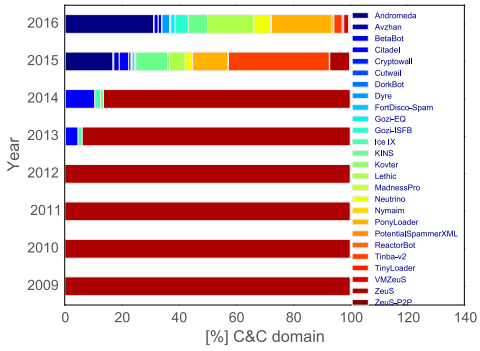

Figure 1: Distribution of malware types over years

3.1 Descriptive Summary of C&C Domains

Figure 1 displays the distribution and evolution of the financial malware families over years, given the first time a malware is seen in our data. The trend indicates the presence of Zeus as the main financial malware between 2009 and 2012. Starting from 2012, we observe the emergence of ZeuS-related families such as Citadel and Ice-IX and gradually other malware families such as Dyre, Cryptowall and Avzhan.

The portion of our C&C data that comes from ZeusTracker also includes information on the type of hosting for some of the C&C domains. The information about the hosting type is gathered by ZeusTracker based on manual analysis of a sample of C&Cs.

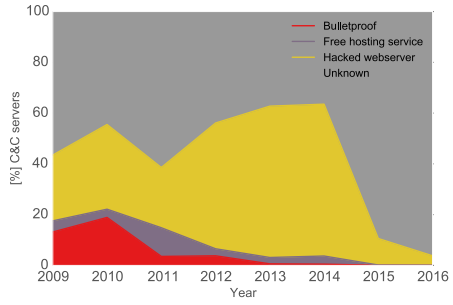

Figure 2: Distribution of malware hosting types over years

Figure 2 shows the distribution of these types over the measurement period. Since the hosting types are known only for a minority of the domains, it is not easy to make any substantive conclusions from the exact numbers. However, the plot suggests that the majority of C&Cs with known types are located on compromised servers, followed by a minority located at free or bulletproof hosting providers. This further highlights the importance of measures taken by providers to protect the machines they are hosting from getting compromised.

3.2 Concentration of C&Cs across Providers

Next, we examine the trends in concentration of C&C domains across providers, to examine if C&C domains are mostly concentrated in specific hosting providers. This could help us to gain a better understanding of attacker preferences.

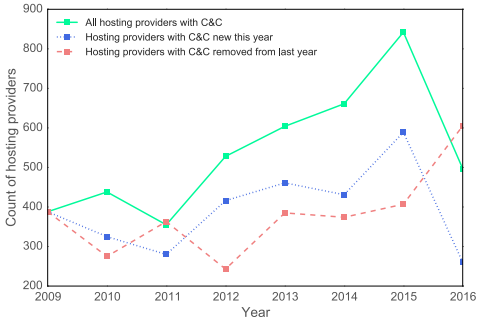

Figure 3: Time-series plot of providers hosting C&Cs

Figure 3 depicts the number of providers hosting C&C domains over time. The green line indicates the total number of providers hosting C&C domains in a given year. The blue line indicates the amount of newly observed providers hosting C&C domains for a specific year while the red line depicts providers that were no longer hosting C&Cs in comparison to the previous year. It should be noted that the removal of a hosting provider is not necessarily due to clean-up efforts, but could be the consequence of attackers' choices. The plot gives a better sense of the total number of hosting providers that are linked with hosting C&C domains.

Over time, we observe a general increase in the total number of providers. At the same time, the number of newly added and removed providers follow a similar upward trend which points to a relatively high entrance and exit rate of providers. The pattern also indicates that an attacker's choice of provider is highly dynamic and shifts from provider to provider over time.

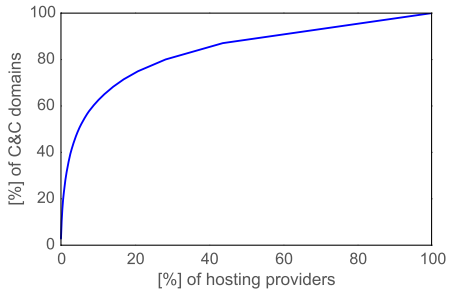

Figure 4: Cumulative percentage of C&C domains for the percentage of hosting providers

Figure 4 displays the cumulative percentage of C&C domains against the percentage of hosting providers. The blue line in the plot follows a power-law distribution: a large

number of C&C domains are concentrated in a small number of hosting providers, 80% of C&Cs are located in less than 30% of the hosting providers. This shows a clear concentration of C&C infrastructure. While the majority of C&Cs are hosted by a minority of providers, it is still unclear whether this concentration in caused by an attacker's preference to choose lax hosting providers in terms of security, or whether it is just an artifact of a provider's size and business model and therefore is randomly distributed. We further examine this question via modeling various provider characteristics in section 4.

3.3 Geography of Providers Hosting C&C Domains

We also examine the geographical distribution of the C&Cs and the providers who host them. Hosting providers operate from various jurisdiction and therefore specific geographical parts of their business could be prone to more abuse due to factors such as weak rule of law or enforcement institutions.

We map the C&C server to their geo-location using the MaxMind GeoIP API [3]. While the C&Cs in our data are located in 109 various countries around the globe, figure 5 suggests that the majority of C&C domains in the top-20 most abused hosting providers are located in US and western Europe. There are a few exceptions such as `Confluence Networks` that seem to operate in part from the Virgin Islands and `SoftLayer Technologies` that hosts domains in Panama.

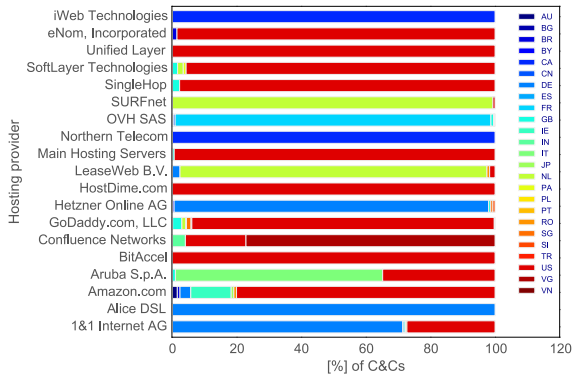

Figure 5: Geo-location of C&C domains for the top-20 providers hosting C&Cs

4. STATISTICAL MODEL OF C&C CONCENTRATIONS

As we explained earlier, we aim to have a better understanding of why C&C domains are concentrated in certain providers through building a statistical model that explains C&C counts from provider characteristics. In previous work, we proposed an approach to study phishing abuse counts across hosting providers using regression models that carefully decomposes different sources of variance in abuse counts for different characteristics [38]. Our current goal is to see whether we see similar patterns in attacker preferences for hosting C&C infrastructure. Contrary to phishing sites, one might expect C&C to be more selectively located.

We define a set of explanatory variables that capture structural characteristics of providers and their security effort, as defined in previous work [38]. In this section we study the relation between C&C abuse and structural characteristics of providers. In the next section, we examine the 'average C&C uptime' as a proxy for the security effort of providers. We categorize the variables characterizing structural properties of providers into those that capture *size*, *regulatory* aspects of the country in which providers operate and those that capture providers' *business model* characteristics. A summary of these variables is provided in Table 2.

4.1 Structural Characteristics of Providers

4.1.1 Size

Allocated IP space is the size of the IP address netblock(s) assigned to a hosting provider according to `WHOIS` data provided by Regional Internet Registries (RIR). We use this information as an indicator of the attack surface of a provider, assuming that the address range is a proxy for the amount of server infrastructure the provider is operating and that any machine in that infrastructure has a certain probability to be abused by miscreants – i.e., more servers means a higher count of C&C. This variable ranges from one IP address to many thousands of allocated addresses, suggesting a large heterogeneity in the market for hosting services, in terms of attack surface but also business models of providers.

Webhosting IP space is the number of IP addresses hosting a domain name. To collect information on this variable we make use of passive DNS data. We calculate this variable by summing up all the IP addresses associated with domains per provider that have been observed in our DNSDB passive DNS data. The combination of the allocated IP space and web hosting IP space not only indicate the size of a provider's infrastructure, but also reflect the kind of business a hosting provider is running. For instance, providers who use a large part of their allocated IP space for hosting domain names have a business model more focused on web hosting and are different from providers who use their allocated IP space for other services such as providing virtual private servers (VPS), collocation, or access services.

Domain name space is the number of domains hosted by a particular provider. Again we use passive DNS data to collect information on this variable. It is calculated by summing up the number of second-level domains hosted on the IP addresses of provider in the passive DNS. Note that due to the large variance and skewed distribution of the first three variables, we use a log-transformation of these variables (Log_{10}).

Proportion of shared hosting measures the ratio of domains that are hosted on shared IP addresses divided by the total size of domain name space. We consider an IP address shared if it hosts more than 10 domain names [39, 42]. This variable not only conveys information about the size of the shared hosting infrastructure of a provider, but also about the provider's business model, i.e., the degree to which a provider's business relies on low-cost shared hosting services.

4.1.2 Regulation

Rule of law is an index that we use as a proxy for law enforcement against illegal activity within a country. It is a well-established indicator relying on a large number of peri-

Table 2: Descriptive summary of variables in our model

variables	n	min	mean	median	max	sd
Allocated IP space size (log_{10})	45,363	0	3.08	3.19	8.35	1.16
Webhosting IP space size (log_{10})	45,363	0	1.78	1.66	6.24	0.76
Domain name space size (log_{10})	45,363	0	1.98	1.83	7.64	0.88
Portion of shared hosting (%)	45,363	0	50.99	58.99	100	37.13
Rule of law	46,269	−1.89	1.05	1.62	2.12	0.95
Best price (USD)	235	0	20.89	6.95	419	47.34
Popularity index	90	0	8,328.34	1,279.29	187,454.30	26,828.03
Time in business (years)	150	2	14.01	14.19	30	4.43
Vulnerable software ratio	86	0.01	0.19	0.16	0.48	0.11

odic surveys to measure how the rule of law is experienced in practical, everyday situations by the general public. The index is provided by the World Justice Project, a non-profit organization working to advance the rule of law around the globe and is based on indicators such as constraints on government powers, absence of corruption, order and security, civil and criminal justice, open government, fundamental rights, regulatory enforcement and justice experienced by ordinary people from 99 countries around the globe [7]. Lower index values represent a stronger rule of law.

4.1.3 *Business Model*

Most of the business model variables in this section cannot easily be collected at scale for the total population of hosting providers. While collecting price information requires manual inspection of the provider's webpage, collecting some other variables in scale such as vulnerable software information can perhaps even be considered unethical since it imposes a cost on the scanned network. Therefore, we collect information for variables in the *business model* category for only a sample of the providers.

Popularity index proxies the online popularity of a hosting provider. We use Alexa's one million top-ranked domains to calculate the popularity index. We assume a provider is more popular when more top-1M domains are on the list of domains that it hosts and speculate that more popular providers are exploited more often for setting up C&C domains. In order to reduce the bias towards the very large hosting providers, the index is calculated by summing up the base-10 logarithm of the reverse Alexa rank of all domains. The score communicates information about both website popularity (i.e., customers) and the density of popular domains in a hosting provider.

Time in business is a proxy for capturing the extent to which a provider can be exploited, given the amount of years it is operating in the hosting business. The expectation is for more experienced providers to be exploited less due to learning effects. The data for this variable is collected by querying the `WHOIS` databse for the registration date of the provider's website. We have cross-checked the results with the Internet Archive database [34] for all data points. Almost all domains in our sample were captured by Webarchive a couple of months after they were registered.

Best Price is basically the least expensive hosting plan on offer by the hosting provider. Our hypothesis is that providers with less expensive hosting plans are more popular to host C&C domains, not only for the case of malicious registrations but also in the case of compromised domains. The intuition being that providers with cheaper plans most probably dedicate less resources to the security of their services. All prices are converted to US dollars by taking the 2015 average exchange rate.

Vulnerable software ratio is the proportion of domains operating on vulnerable software installations hosted by the providers in our study. Previous research shows that popular software such as Content Management Systems (CMS), increase the chance of getting compromised [42]. We use Wordpress installations as a proxy for common vulnerable software. To scan for such installations, we use WPScan – a Wordpress vulnerability scanner developed and supported by Sucuri– to collect data for a random sample of 2% of a provider's hosted domains [4]. The 'vulnerable software ratio' is calculated by dividing the number of scanned domains with Wordpress installations by all scanned domains excluding those that we were unable to scan.

4.2 Effect of Providers' Structural Characteristics

To disentangle the effects of the various structural characteristics of hosting providers which we have outlined previously on the concentration of C&Cs, we use a generalized linear model (GLM) with log-linear link-function of the form:

$$\ln(\lambda_i) = \beta_{1i} AllocatedIPSize + \beta_{2i} WebhostingIPSize + \beta_{3i} DomainSize + \beta_{4i} SharedHosting + \beta_{5i} RuleofLaw,$$

where the dependent variable – count of C&C domains – follows a Poisson distribution with parameter $\lambda \geq 0$ and βs are the estimated coefficients for the explanatory variables collected for all the hosting providers. Subscript i refers to measurements in different hosting providers.

We construct several models using the variables in the equation above, for the whole population of hosting providers. Our goal is both to maximize the amount of overall explained variance of the model and to find out which of these variables influence the concentration of C&C abuse in providers the most. The result of our regression models are displayed in Table 3.

It is important to note that the reason for building more than one model is to be able to compare goodness of fit values while adding new variables to each model. Hence, to asses how the models are performing in absolute terms and relative to the other models, we use the Log-likelihood, AIC

Table 3: Generalized Linear Regression Model (GLM) for the Population of Hosting Providers

	Response Variable: Count of C&C domains				
	Poisson with Log Link Function				
	(1)	(2)	(3)	(4)	(5)
Allocated IP space size		-0.991^{***}	-0.356^{***}	-0.358^{***}	-0.398^{***}
		(0.019)	(0.020)	(0.020)	(0.020)
Webhosting IP space size		2.725^{***}	0.711^{***}	0.868^{***}	0.931^{***}
		(0.020)	(0.027)	(0.031)	(0.032)
Domain name space size			1.465^{***}	1.301^{***}	1.300^{***}
			(0.014)	(0.020)	(0.021)
Portion of Shared hosting business				0.009^{***}	0.009^{***}
				(0.001)	(0.001)
Rule of Law					-0.213^{***}
					(0.013)
Constant	-1.380^{***}	-5.058^{***}	-6.834^{***}	-7.319^{***}	-7.130^{***}
	(0.009)	(0.039)	(0.049)	(0.066)	(0.068)
Observations	46,455	45,358	45,358	45,358	45,166
Log Likelihood	−50,777.110	−21,665.390	−15,485.920	−15,418.070	−15,253.920
Akaike Inf. Crit.	101,556.200	43,336.780	30,979.840	30,846.140	30,519.850
Dispersion	46.62	11.049	9.296	9.641	10.328
Pseudo R^2		0.587	0.717	0.719	0.722

Note: *p<0.05; **p<0.01; ***p<0.001
Standard errors in brackets

statistic, the Poisson dispersion parameter and the pseudo R-squared as measures of goodness-of-fit. We aim to minimize log-likelihood and AIC (the closer to 0 the better). The Poisson model assumes that $var[Y_i] = \phi E[Y_i] = \phi\lambda_i$, with $\phi \stackrel{!}{=} 1$, where ϕ is a dispersion parameter. The dispersion parameter hence captures the extent to which variance is different from the mean and more specifically the heterogeneity of the model. A model is over-dispersed when ϕ>1. The pseudo R-squared [17] is likewise calculated for our Poisson model given the dispersion parameter ϕ, using the following formula :

$$R^2 = 1 - \frac{D(\boldsymbol{y}, \hat{\boldsymbol{\lambda}}) + k \cdot \hat{\phi}}{D(\boldsymbol{y}, \bar{Y})}, \quad (1)$$

where $D(\boldsymbol{y}, \hat{\boldsymbol{\lambda}})$ is the deviance of the fitted model, $D(\boldsymbol{y}, \bar{Y})$ is the deviance of the intercept-only model, $\hat{\phi}$ is the estimated dispersion parameter and k is the number of covariates fitted, (excluding intercept). By building several models, we aim to maximize the value of the pseudo R-squared hence maximizing the amount of variance explain in C&C abuse counts by the dependent variables.

By inspecting Table 3, model 1 is the intercept-only model with count of C&C domains as dependent variables and no dependent variable. In Model 2, we take into account the size variables –'Allocated IP space size' and 'Webhosting IP space size'. The model indicates a significant negative relation between the variable 'Allocated IP space size' and C&C abuse counts, while 'Webhosting IP space size' correlates positively with C&C abuse counts. This is very much expected as pointed out earlier in the paper, these two variables together determine to what extent the provider is using its allocated IP space for web hosting services. In addition, our manual inspection of the hosting data shows providers with very large allocated IP space are normally not pure hosting providers but rather broadband providers who use a small portion of their IP space for hosting. Moreover, the value of our goodness-of-fit criteria shows that only by adding these two size variables, we have substantially reduced the log-likelihood, AIC and dispersion values and are able to explain approximately 58% of the variance in abuse counts.

We build on model 2 by including additional variables, namely 'Size of the domain names space' along with the extent to which a provider is hosting its domains on shared hosting services. Model 4 displays the estimated coefficients. The results indicate more domains in general and specifically shared hosting domains relate significantly with more C&C abuse. To put the value of the coefficients in perspective, by holding all other values constant in the model, a unit increase in the value of 'Size of the domain names space', multiplies the number of C&Cs by $e^(1.300) = 3.7$.

In addition to size variables analyzed in Models 1 to 4, we hypothesized that the rule of law index of a hosting provider's country might play a significant role in explaining the concentration of abuse in that country. Previous work has shown that the location of banks targeted by Zeus malware is not random [37]. Similarly, our dataset shows

Table 4: Generalized Linear Regression Model (GLM) a Sample of Hosting Providers

	Response Variable: Count of C&C domains					
	Poisson with Log Link Function					
	(1)	(2)	(3)	(4)	(5)	(6)
Best price		-0.004^{***}	-0.019^{***}	-0.043^{***}	-0.018^{***}	-0.084^{***}
		(0.001)	(0.002)	(0.006)	(0.005)	(0.015)
Time in business			0.063^{***}	0.075^{***}	0.060^{***}	0.070^{***}
			(0.006)	(0.008)	(0.009)	(0.014)
Vulnerable software ratio				1.463^{***}	1.462^{***}	2.035^{***}
				(0.327)	(0.366)	(0.508)
Popularity index					0.00001^{***}	0.00002^{***}
					(0.000000)	(0.000002)
Constant	-4.146^{***}	-2.363^{***}	-3.121^{***}	-3.533^{***}	-3.881^{***}	-20.624
	(0.011)	(0.024)	(0.095)	(0.145)	(0.160)	(2,103.363)
Country fixed-effects	No	No	No	No	No	Yes
Observations	45,363	230	144	85	85	85
Log Likelihood	$-21,854.350$	$-1,625.306$	$-1,133.003$	-715.212	-564.210	-343.260
Akaike Inf. Crit.	43,710.690	3,254.612	2,272.005	1,438.424	1,138.420	754.521

Note: *p<0.05; **p<0.01; ***p<0.001
Standard errors in brackets

that some C&Cs are hosted in several islands all around the globe which are mostly the so-called tax-heavens. We examine this effect by including the Rule of law index variable in addition to the other previous 4 variables in model 5. We see a clear negative relation between the rule of law index and the concentration of C&Cs abuse. Although the Rule of law index is a combination of several country-level regulation indicators, it provides valuable insight about the proportion of abuse in certain geographical locations.

With the fitted values in our final model – model 5 –, we are able to explain approximately 72% of the observed variance (i.e., Pseudo R-squared = 0.72) in C&C counts only through considering the size variables and the rule of law. This highlights a very important point: regardless of the security measures a provider has in place, certain characteristics, driven by the nature of a provider's business, are driving the majority of the abuse.

Note we hypothesize that there are additional factors that influence the concentration of C&C abuse in hosting providers such as variables that capture the business model of a provider, for example price of the hosting service. However, such variables are much harder to collect at scale for all hosting providers. In the next section, we assess the impact of such factors on the concentration of C&C abuse within a smaller sample of hosting providers.

4.3 Concentrations of C&Cs in a Sample of Providers

As explained earlier, we collected additional business model variables for a sample of providers 4.1. We initially started from a set of 235 randomly selected providers for which we collected price information, however due to missing values in other variables we ended up with 85 providers for whom we have data on all the four variables in this category.

Note that the downside of our sampling strategy is that we might end up with geographical biases. In order to control for such effects, we fit a "fixed-effects" GLM model with the count of C&C domains as dependent variable following a Poisson distribution. We add a country fixed effect, δ_i, by fitting a separate dummy variable as a predictor for each country. The country fixed effect prevents undue dependence of the residuals.

Similar to before, we add the variables one by one to the baseline model (model 1) to observe the extent to which a model is improved in comparison to others. The resulting models are shown in Table 4. Model 5 contains all of the 4 variables discussed before. In addition to those, in our final model, (model 6), we add the country fixed-effect variable as well.

Inspecting the estimated coefficients of model 6, we observe a significant negative relation between price of hosting and C&C counts. That is to say that if we were to increase price by one unit while holding all other variables constant, the C&C counts would be multiplied by $e^{(-0.084)} = 0.91$ as a result. The cheaper a providers price is, the more likely it is for the hosting provider to host C&C domains. As expected, the variable 'Best price' shows a weaker relation in model (5) where cross-country differences are not controlled by the fixed country effects of model (6). This is because the properties of hosting markets in different countries can differ substantially, which then eventually influence the cost of infrastructure in a country with respect to hosting services. Moreover, the cost of a hosting plan is in proportion to the economy of the provider's country. Hence, our conversion of

prices in different specific countries to USD, if not controlled for the country differences, can be very crude.

The Variables 'Vulnerable software ratio' and 'Popularity index' also show a significant positive relation with C&C counts. One unit increase in 'Vulnerable software ratio' while holding other variables constant, multiplies the C&C counts by $e^{(2.035)} = 7.652$. The 'Time in business' variable shows a significant positive relation with abuse as well, indicating that well-known providers or those who are in business for a longer time are attacked more. Please note that this can partially be caused by the fact that our data is longitudinal. In the following section, we will study the effect of C&C take-down speed on its concentration across providers.

5. EFFECT OF C&C TAKE-DOWN SPEED

Up to this point, we have demonstrated that the concentration of C&C domains can be explained by structural characteristics of providers, mostly related to their size and business model. Together, these factors form a proxy for the attack surface of the industry. The attack surface of providers accounts for at least 72% of the variance in the number of C&C in their networks. Providers with more infrastructure get more C&C. This does not indicate selective location choices by the attackers. Quite the opposite, in fact. The bulk of C&C can be explained from attackers randomly distributing their C&C domains across the overall global hosting infrastructure.

In this section, we investigate whether attackers prefer providers who are lax in taking down C&C servers. Longer uptime of C&Cs seems valuable for the attackers, so we would expect higher C&C counts in those networks. We examine if and how C&C uptimes influences the number of servers at that provider. C&C uptime has been used in previous security research as a standard metric for studying the lifetime of different attack types [15].

We define the "uptime" of a C&C domain as the number of days between the first and last time the C&C domain is observed online as reported by our datafeeds. Some of the C&C domains remain online beyond the measurement period, which unavoidably leaves their uptime unknown. The average uptime of C&C domains is depicted in Figure 6. There is no clear trend one way or the other. This also suggests that there are no learning effects among hosting providers that enables faster takedown over time. We first examine if the average C&C uptime is driven by a few providers, or whether it reflects the overall performance of hosting providers.

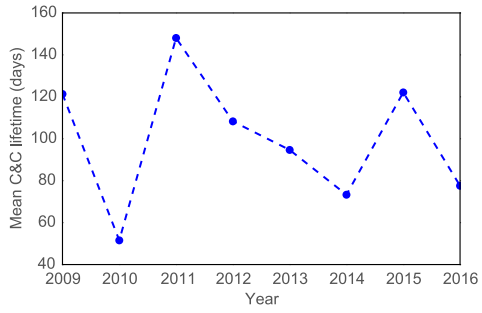

Figure 6: Mean uptime of C&C domains per year

5.1 Distribution of C&C Lifetime

Are long-lived C&Cs concentrated in certain providers? Figure 7 depicts the distribution of the average uptime of C&C domains per provider, for 2009-2016. Note that each individual plot is also indicative of the amount of C&C domains that are taken down in the corresponding year. What is clear from all the plots is that, there is always a majority of providers with a shorter uptime followed by a long tail of providers hosting C&Cs with very long uptimes. In some case there are examples of providers that hosts C&C domains for more than a year. Assuming no measurement errors are at play here, such examples could indicate ignorant or perhaps even bullet-proof hosting. This leads us to the next question: are these providers preferred by attackers?

5.2 Differences between C&C Take-down Speed of Providers

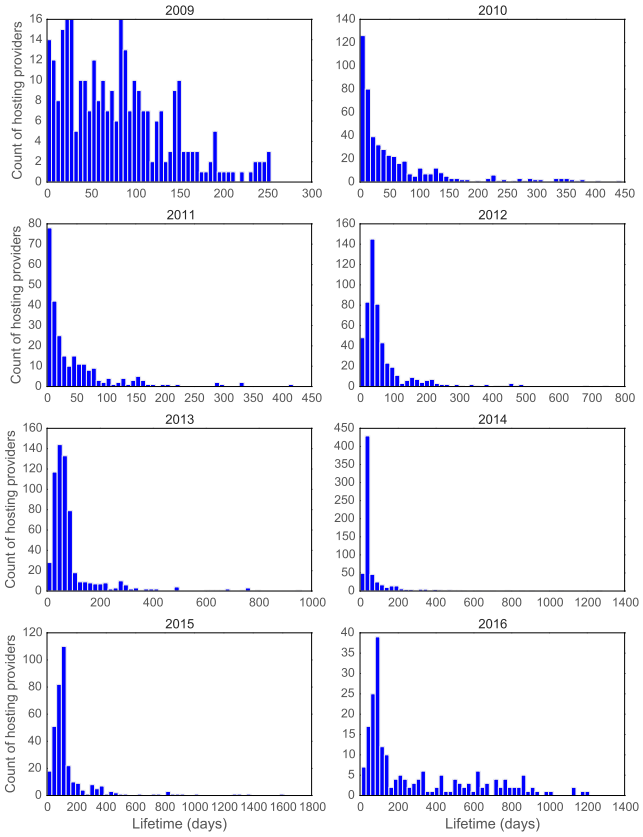

Figure 7: Distribution of C&C average uptime hosted by providers over years

To examine the differences between providers more carefully, we model the survival rate $S(t)$ of C&C domains using a Kaplan-Meier Survival Estimate which also allows to correctly account for the C&C domains that are not taken down by the end of our measurement period, i.e., right-censored data points [21]. The survival rate $S(t)$ basically expresses the probability that a C&C domain is online at a specific time during the observation period.

Figure 8 displays the survival curves of C&C domains in

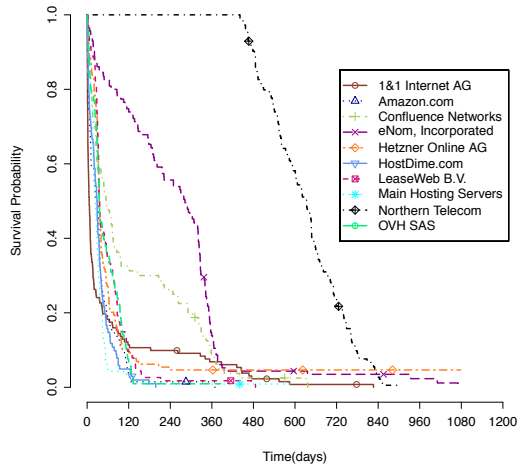

Figure 8: Kaplan-Meier estimated survival function of C&C uptime for top-10 most attacked hosting providers

the top-10 providers with the highest number of C&C domains in their network. Figure 9 depicts the χ^2 value of the Log-Rank test in which the providers are compared two by two in terms of their survival rate. Only the light blue tiles indicate non-significant differences at a 0.05 significance level.

As both plot suggests, hosting providers perform differently either in terms of survival probability or in terms of the total number of days that their C&C domains remain online. For example, more than 95% of the C&C domains in `Main Hosting Server` are taken down after approximately 60 days which is very similar to `HostDime.com`. However C&C domains hosted by `HostDime.com` are in total taken down after maximum of 4 months whereas this takes about more than a year for `Main Hosting Server`. On the extreme side are providers such as `Northen Telecom` and `eNom, incorporated` that host C&C domains that are online for more than 2 years.

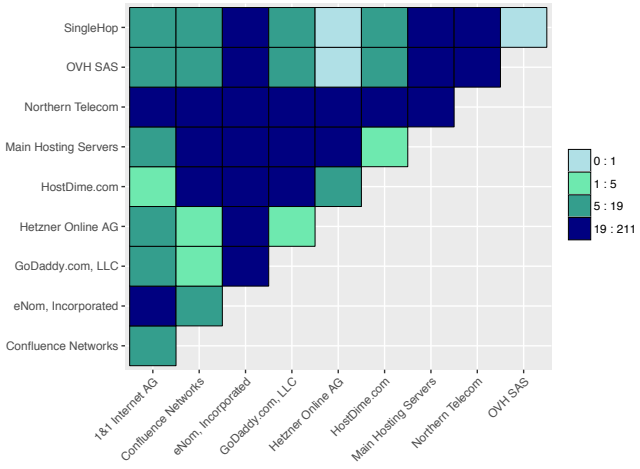

Figure 9: Log-rank test for pairs of providers (legend shows χ^2 value of Log-rank test)

5.3 Does C&C Uptime Explain Abuse Concentration?

Given that there are significant differences in the C&C uptime among hosting providers, we now analyze the impact of C&C uptime on the concentration of C&C abuse within providers. Do attackers prefer providers with long C&C uptimes?

We fit a similar GLM model to model 3 from Section 4.2 for the population of providers, having the count of C&C domains following a Poisson distribution. The result of our model with the addition of a C&C uptime variable is shown in Table 5.3.

In order to be able to make a relative comparison, model 2 is the final model from our inquiry into the structural properties of providers (see Table 3). We add the 'Average C&C uptime' variable to that model. The resulting estimated coefficients are observable in model 3. The model suggests that the variable 'Average C&C uptime' shows a statistically significant positive relationship with the number of C&Cs. Although the relationship is significant, we are only able to explain a total of 73% variance of C&C abuse counts, which is a 1% increase compared to the model with only structural provider variables. This rather indicates that there is very little or no preference by attackers for hosting their C&C domains at providers who allow long-living C&C domains.

Table 5: Generalized Linear Regression Model

	Response Variable: Count of C&C domains		
	Poisson with Log Link Function		
	(1)	(2)	(3)
Allocated IP space size		-0.398^{***}	-0.447^{***}
		(0.020)	(0.020)
Webhosting IP space size		0.931^{***}	1.054^{***}
		(0.032)	(0.032)
Domain name space size		1.300^{***}	1.162^{***}
		(0.021)	(0.020)
Portion of Shared hosting business		0.009^{***}	0.011^{***}
		(0.001)	(0.001)
Rule of Law		-0.213^{***}	-0.170^{***}
		(0.013)	(0.013)
Average C&C uptime			0.003^{***}
			(0.0001)
Constant	-1.380^{***}	-7.130^{***}	-7.077^{***}
	(0.009)	(0.068)	(0.069)
Observations	46,455	45,166	45,166
Log Likelihood	$-50,777.110$	$-15,253.920$	$-14,699.260$
Akaike Inf. Crit.	101,556.200	30,519.850	29,412.510
Dispersion	46.623	10.328	9.032
Pseudo R^2		0.722	0.733
Note:		*p<0.05; **p<0.01; ***p<0.001	
		Standard errors in brackets	

6. RELATED WORK

With the increasing number of attacks on financial services, efforts from industry and academia have focused on botnet evolution and mitigation strategies.

A first area of research aims to understand the functionality of the different malware families to develop countermeasures that could disrupt these botnets. Different studies have investigated the communication protocols of these botnets and their spreading techniques [20, 24, 27].

These studies have collected data on C&C and other botnet infrastructure. These are typically presented in a descriptive analysis, such as their distribution over countries. Rossow et al. [33] analyzed the lifetime and domain name characteristics of malware downloaders. They observed steady migrations of malware downloaders from domains and TLD registrars to others, and notice that attackers redundantly deploy their critical infrastructures across providers. Han et al. in [16] investigated the way cyber-criminals abuse public cloud services to host part of their malicious infrastructure, including exploit servers to distribute malware and C&C servers to manage infected terminals. Our work complements the insights obtained by these works by analyzing the factors that drive attackers to choose certain type of hosting provider.

A second strand of work has developed approaches to better detect botnet infrastructure. Cyberprobe [30] describes an active probing approach for detecting malicious servers and compromised hosts. ASwatch [22] aiming at detecting and identifying malicious ASes that exhibit "agile" control plane behavior (e.g., short-lived routes, aggressive rewiring). In this context, fast flux also appears as a technique that uses compromised computers to provide scalability, geographic diversity, anonymity and redundancy to organized cybercrime operators. The fast flux infrastructure relies on computing resources stolen from the unwitting users of infected endpoints. Cybercriminals rent these fast flux proxy networks to create a profitable black market hosting environment. The authors of [8, 44] have analyzed the structural relationships (domain, nameserver, IP connectivity) of fast-flux botnets and identified recurrent structural clusters across different botnet types. In [8], the authors have used a social network connectivity metric to show that {Command and Control and phishing} and {malware and spam botnets} have similar structural scores using the proposed metric. In this paper, we have defined metrics to capture not only the attacker behavior but also the hosting provider effort toward mitigating the malicious infrastructure located in their networks.

A third strand of work is the development of reputation systems for providers, especially focused on those that facilitate cybercrime [23, 31, 35]. For example, FIRE [35] introduced a ranking system using uptime of botnet hosting services to identify and expose providers that demonstrate persistent, malicious behavior. In [31] the authors propose various reputation metrics based on the concentration of abuse, while taking some structural hosting provider characteristics into account. During the explanatory analysis conducted in this paper, we use the structural properties of hosting providers to assess the impact of these on their security performance.

All these approaches help to identify and enumerate botnet C&C infrastructure and to describe their distribution across networks and countries. We extend this related work via explanatory analysis to determine the driving factors for the locations of the C&C infrastructure in the hosting market. We statistically model and explain the distribution of C&C from the structural properties of hosting providers, business models and factors like rule of law. We expand the work by Gañán et al. [15] by studying the properties of providers hosting C&C domains.

Hosting providers play a key role in the size and spread of these botnets. Different abuse reporting strategies have been proposed and evaluated to analyze the performance of hosting providers [11, 19, 29]. However, as shown by Canali et al. [10], hosting providers are often not taking appropriate measures, probably due to a lack of incentives. Millions of websites are often poorly managed by inexperienced users, shared web hosting providers have not developed reliable mechanism to keep their users safe. Moreover, with the emergence of cloud providers, attackers have a new platform to host their infrastructure. Current studies have shown that these type of providers are being used to launch long-tail spam campaigns because of their low cost [25,36]. Only a few specific providers have attempted to create added value by providing "add-on" security services. For instance, a Dutch web hosting provider [13] has added a free automated website vulnerability scanning, fixing and recovery service.

On the other end of the spectrum there are hosting providers acting as cybercrime facilitators [22, 40, 41]. Researchers and law enforcement agencies are searching better ways at squashing these providers. While these efforts are critical for the overall fight against cybercrime, our analysis suggests that the C&C of the botnets engaged in attacks on financial services do not depend on malicious hosting providers, nor do attackers seem to prefer these providers when locating their C&C.

7. CONCLUSIONS AND FUTURE WORK

Over the years, hosting providers have spent a great deal of effort taking down C&C infrastructure for botnets engaged in attacks on financial services.

This paper aimed to enlighten the strategies of the attackers using these botnets for the placement of their C&C servers across the hosting market. More specifically, we examined if attackers have shown a preference for providers with lax security efforts. Or, conversely, whether the placement choice of C&C domains is rather randomly distributed across the hosting space, as measured via the provider's structural properties.

We studied seven years of C&C data for 26 botnet families engaged in attacks on financial services and demonstrated a general increase in the total number of providers hosting C&C domains over time. We also found a dynamic pattern of providers who enter and exit the population of providers that host financial malware C&C.

Our results show that C&C abuse is highly concentrated in a small number of providers. That being said, this concentration can be explained from relatively large portion that these providers have of the overall attack surface of the hosting market.

To study the effect of hosting provider characteristics on C&C concentrations, we modeled the distribution of C&Cs using Generalized Linear Models (GLM), with C&C counts following a Poisson distribution. We showed that a provider's attack surface characteristics such as IP and domain space size and the proportion of shared hosting can explain around 71% of the variance in the number of C&Cs per provider. The rule of law in a country only explains an additional 1% of the variance, suggesting that the attackers do not prefer providers in jurisdictions with weak law enforcement. All in all, the selection process for C&C seems to be random: the probability of hosting C&C is highly proportional to the attack surface of the providers, as measured the by observed effect of indicators of size of the provider.

In addition, business model characteristics of providers

show a significant relation with C&C concentrations for a sample of hosting providers. While the pricing of a hosting plan negatively affects C&C concentrations, provider's popularity, time in business and the ratio of vulnerable software, have a significant positive relation with C&C concentrations.

Despite statistically significant differences in C&C take-down speeds among providers, when modeled in conjunction with attack surface variables, take-down speed shows only a very weak relation with the concentration of C&Cs across providers, suggesting that attackers are rather impervious to the take-down efforts of hosting providers.

On a more general level, our results suggest that the amount of C&C abuse in the network of a provider is a function of a provider's structural properties such as its size and its pricing strategy, rather than being driven by the effort they put in abuse handling.

Additionally, our approach helps in developing evidence-based policies in the hosting market. That is, we demonstrate an approach that enables better comparative abuse metrics by controlling for the structural differences among providers rather than relying on absolute counts.

Our work comes with a set of limitations as well. The dataset contains only malware families that have been used to attack financial institutions. Some are predominantly used for this purpose, like Citadel, but others are much more generic malware families. Although our methodology in generalizable, it is an open question whether the patterns we found are different for different kinds of abuse data. Future work could explore this. In addition, our uptime analysis can contain biases from unknown measurement errors in the first-seen and last-seen observations of C&C domains. Such observations are known to be quite noisy. We do however think that the effects would be negligible since the biases (if any) would be systematic. Finally, because we have used pooled data for the whole measurement period, our models do not account for changes of C&C counts over time. Future work can look into whether these patterns we discussed in this paper change over time.

Acknowledgments

The authors thank Farsight Security for providing access to DNSDB. We would like to thank Roman Huessy from ZeusTracker for generously sharing his data on Zeus and his methodology. This work was supported by NWO (grant nr. 12.003/628.001.003), the National Cyber Security Center (NCSC) and SIDN, the .NL Registry. Additionally, we thank our shepherd Juan Caballero for his support in improving the paper for the camera-ready version.

8. REFERENCES

[1] DNSDB. https://www.dnsdb.info.

[2] Farsight Security. https://www.farsightsecurity.com.

[3] MaxMind.

[4] WPScan. http://wpscan.org.

[5] Microsoft Security Intelligence Report. https://www.microsoft.com/security/sir/default.aspx, 2015.

[6] Zeus Tracker. https://zeustracker.abuse.ch, August 2016.

[7] Botero, J. C., and Ponce, A. Rule of law index. *The World Justice Project* (2010).

[8] Caglayan, A., Toothaker, M., Drapeau, D., Burke, D., and Eaton, G. Behavioral analysis of botnets for threat intelligence. *Information Systems and e-Business Management 10*, 4 (2012), 491–519.

[9] Cai, X., Heidemann, J., Krishnamurthy, B., and Willinger, W. Towards an AS-to-organization map. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement* (2010), ACM, pp. 199–205.

[10] Canali, D., Balzarotti, D., and Francillon, A. The Role of Web Hosting Providers in Detecting Compromised Websites. In *Proceedings of the 22Nd International Conference on World Wide Web* (2013), WWW '13, pp. 177–188.

[11] Cetin, O., Jhaveri, M. H., Gañán, C., van Eeten, M., and Moore, T. Understanding the role of sender reputation in abuse reporting and cleanup. *Journal of Cybersecurity 2*, 1 (2016), 83–98.

[12] Chang, W., Mohaisen, A., Wang, A., and Chen, S. Measuring botnets in the wild: Some new trends. In *Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security* (2015), ACM, pp. 645–650.

[13] de Vries, W. Hosting provider Antagonist automatically fixes vulnerabilities in customers' websites. https://www.antagonist.nl, 2012.

[14] Dimitropoulos, X., Krioukov, D., Riley, G., and claffy, k. Revealing the Autonomous System Taxonomy: The Machine Learning Approach. In *Passive and Active Network Measurement Workshop (PAM)* (2006), pp. 91–100.

[15] Gañán, C., Cetin, O., and van Eeten, M. An Empirical Analysis of ZeuS C&C Lifetime. In *Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security* (2015), ASIA CCS '15, pp. 97–108.

[16] Han, X., Kheir, N., and Balzarotti, D. The role of cloud services in malicious software: Trends and insights. In *Proceedings of the 12th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment - Volume 9148* (New York, NY, USA, 2015), DIMVA 2015, Springer-Verlag New York, Inc., pp. 187–204.

[17] Heinzl, H., and Mittlböck, M. Pseudo R-squared measures for Poisson regression models with over-or underdispersion. *Computational statistics & data analysis 44*, 1 (2003), 253–271.

[18] Hostexploit. World Hosts Report. Technical report. http://hostexploit.com/downloads/summary/7-public-reports/52-world-hosts-report-march-2014.html, 2014.

[19] Jhaveri, M. H., Cetin, O., Gañán, C., Moore, T., and Eeten, M. V. Abuse reporting and the fight against cybercrime. *ACM Computing Surveys (CSUR) 49*, 4 (2017), 68.

[20] Kang, B. B., Chan-Tin, E., Lee, C. P., Tyra, J., Kang, H. J., Nunnery, C., Wadler, Z., Sinclair, G., Hopper, N., Dagon, D., and Kim, Y. Towards complete node enumeration in a peer-to-peer botnet. In *Proceedings of the 4th International Symposium on Information, Computer, and Communications Security* (New York, NY, USA, 2009), ASIACCS '09, ACM, pp. 23–34.

[21] Kaplan, E. L., and Meier, P. Nonparametric estimation from incomplete observations. *Journal of the American statistical association 53*, 282 (1958), 457–481.

[22] Konte, M., Perdisci, R., and Feamster, N. ASwatch: An AS reputation system to expose bulletproof hosting ASes. *ACM SIGCOMM Computer Communication Review 45*, 4 (2015), 625–638.

[23] Korczyński, M., Tajalizadehkhoob, S., Noroozian, A., Wullink, M., Hesselman, C., and van Eeten, M. Reputation metrics design to improve intermediary incentives for security of tlds. In *2017 IEEE European Symposium on Security and Privacy (Euro SP)* (April 2017).

[24] Li, Z., Goyal, A., Chen, Y., and Paxson, V. Automating analysis of large-scale botnet probing events. In *Proceedings of the 4th International Symposium on Information, Computer, and Communications Security* (New York, NY, USA, 2009), ASIACCS '09, ACM, pp. 11–22.

[25] Liao, X., Liu, C., McCoy, D., Shi, E., Hao, S., and Beyah, R. Characterizing Long-tail SEO Spam on Cloud Web Hosting Services. In *Proceedings of the 25th International Conference on World Wide Web* (2016), WWW '16, pp. 321–332.

[26] Liu, S., Foster, I., Savage, S., Voelker, G. M., and Saul, L. K. Who is .com? Learning to parse WHOIS records. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference* (2015), ACM, pp. 369–380.

[27] Lu, W., Tavallaee, M., and Ghorbani, A. A. Automatic discovery of botnet communities on large-scale communication networks. In *Proceedings of the 4th International Symposium on Information, Computer, and Communications Security* (New York, NY, USA, 2009), ASIACCS '09, ACM, pp. 1–10.

[28] M3AAWG. Anti-abuse best common practices for hosting and cloud service providers. https://www.m3aawg.org/sites/maawg/files/news/M3AAWG_Hosting_Abuse_BCPs-2015-03.pdf, 2015.

[29] Nappa, A., Rafique, M. Z., and Caballero, J. Driving in the cloud: An analysis of drive-by download operations and abuse reporting. In *Proceedings of the 10th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment* (Berlin, Heidelberg, 2013), DIMVA'13, Springer-Verlag, pp. 1–20.

[30] Nappa, A., Xu, Z., Rafique, M. Z., Caballero, J., and Gu, G. Cyberprobe: Towards internet-scale active detection of malicious servers. In *In Proceedings of the 2014 Network and Distributed System Security Symposium (NDSS 2014)* (2014), pp. 1–15.

[31] Noroozian, A., Korczynski, M., Tajalizadehkhoob, S., and van Eeten, M. Developing security reputation metrics for hosting providers. In *8th Usenix Workshop on Cyber Security Experimentation and Test (CSET 15)* (2015).

[32] Rossow, C., Andriesse, D., Werner, T., Stone-Gross, B., Plohmann, D., Dietrich, C. J., and Bos, H. Sok: P2pwned-modeling and evaluating the resilience of peer-to-peer botnets. In *Security and Privacy (SP), 2013 IEEE Symposium on* (2013), IEEE, pp. 97–111.

[33] Rossow, C., Dietrich, C., and Bos, H. Large-scale analysis of malware downloaders. In *Proceedings of the 9th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment* (Berlin, Heidelberg, 2013), DIMVA'12, Springer-Verlag, pp. 42–61.

[34] Solutions TotalBank. Internet archive. http://archive.org/web/, 2016.

[35] Stone-Gross, B., Kruegel, C., Almeroth, K., Moser, A., and Kirda, E. FIRE: FInding Rogue nEtworks. In *2009 Annual Computer Security Applications Conference* (Dec 2009), pp. 231–240.

[36] Stringhini, G., Hohlfeld, O., Kruegel, C., and Vigna, G. The Harvester, the Botmaster, and the Spammer: On the Relations Between the Different Actors in the Spam Landscape. In *Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security* (New York, NY, USA, 2014), ASIA CCS '14, ACM, pp. 353–364.

[37] Tajalizadehkhoob, S., Asghari, H., Gañán, C., and Van Eeten, M. Why them? extracting intelligence about target selection from zeus financial malware. In *Proceedings of the 13th Annual Workshop on the Economics of Information Security, WEIS 2014, State College (USA), June 23-24, 2014* (2014), WEIS.

[38] Tajalizadehkhoob, S., Böhme, R., Gañán, C., Korczyński, M., and van Eeten, M. Rotten Apples or Bad Harvest? What We Are Measuring When We Are Measuring Abuse.

[39] Tajalizadehkhoob, S., Korczynski, M., Noroozian, A., Ganán, C., and van Eeten, M. Apples, oranges and hosting providers: Heterogeneity and security in the hosting market. In *Network Operations and Management Symposium (NOMS)* (2016), IEEE/IFIP, pp. 289–297.

[40] TrendMicro. Criminal Hideouts for Lease: Bulletproof Hosting Services. http://www.trendmicro.fr/media/wp/wp-criminal-hideouts-for-lease-en.pdf.

[41] TrendMicro. Looking Into a Cyber-Attack Facilitator in the Netherlands. http://blog.trendmicro.com/trendlabs-security-intelligence/looking-into-a-cyber-attack-facilitator-in-the-netherlands/.

[42] Vasek, M., Wadleigh, J., and Moore, T. Hacking Is Not Random: A Case-Control Study of Webserver-Compromise Risk. *IEEE Transactions on Dependable and Secure Computing 13*, 2 (2016), 206–219.

[43] Welzel, A., Rossow, C., and Bos, H. On measuring the impact of DDOS botnets. In *Proceedings of the Seventh European Workshop on System Security* (2014), ACM, p. 3.

[44] Xu, W., Wang, X., and Xie, H. New trends in fastflux networks. In *Proceedings of the 16th BlackHat USA*.