

Preliminary Draft

**Broadband Industry Structure and Cybercrime
An Empirical Analysis**

**Draft 3.0¹
April 24, 2016**

**Carolyn Gideon, Tufts University
Christiaan Hogendorn, Wesleyan University**

Abstract

Prior studies have shown that while ISPs are well positioned to provide residential and SME users critical protection from cybercrime, their incentives to do so are often insufficient. The presence of competition in providing broadband service is a factor we might expect to impact such incentives as shown in prior theoretical work. We test this finding using data consisting of malware attacks sent by devices with botnet obtained with honeypot sensors. We conduct preliminary analysis of data collected to date. Using econometrics to analyze the frequency of infected devices found on residential ISP networks by county we find a significant and nonlinear relationship between the number of ISPs providing service in a county and the rate of infection. In counties with very small numbers of ISPs, entry of a new ISP will enhance security against botnet infections. Preliminary results suggest that in counties with larger numbers of ISPs, additional ISPs will decrease security.

¹ As data collection is still in process, there are limited results reported in this preliminary draft.

1. Introduction

Cybercrime continues to be a growing drain on the world economy. A widely cited recent report estimated the global cost of cybercrime at \$475 billion (CSIS 2015). Amid this continual game of cat-and-mouse between hackers and those minding data stored and in transit, Internet service providers (ISPs) can play a pivotal role. While individual users obviously bear responsibility for maintaining basic safety through good digital hygiene and other precautions, there is much that can be done by the more concentrated and knowledgeable ICT industry. The advantages of such were noted by US FCC Chairman Tom Wheeler, when he recently called for the communications sector to “create a new paradigm of cyber readiness ...” in which “...the network ecosystem must step up to assume new responsibility and market accountability for managing cyber risks,” (Wheeler, 2014). ISPs in particular, as providing what might be considered a gateway to the Internet for the common users, are well positioned to help secure against most forms of cybercrime. Also, ISPs generally possess the necessary expertise, usually lacking in their residential and SME customers.

We are beginning to understand the nature of ISP incentives to invest in cybersecurity for their residential and SME customers through a building body of economic modeling literature. This is augmented by empirical evidence, including a 2010 OECD study showing that ISPs have significant discretion, and variation, in how they address botnet mitigation. The authors of that study recognized and estimated many of the factors that can explain the sizable differences found in the security performance of ISPs, emphasizing the institutional

Preliminary Draft

and organizational characteristics that shape ISPs' incentives (van Eeten et. al., 2010). This paper builds on such existing theoretical and empirical work to further explore the role of broadband competition on the level of security provided by ISPs. This is a particularly relevant area of inquiry for an industry where fast-paced innovation can create new entrants, such as the increasingly mobile provision of broadband.

The relationship between ISP competition and security investment has multiple dimensions. Increased competition might lower the margins for ISPs, resulting in lower security investment. Alternatively, if users are interested in greater security and can discern the relative security levels of the competing ISPs, competition might lead to increased security investment. Yet another possibility is that competition provides an opportunity to free ride on the security provided by rival ISPs, again reducing security investment. Gideon and Hogendorn (2015) show theoretically how ISP incentives change in different competitive situations. In this paper we test those theoretical results by analyzing malware data and tracing the geographic locations of the infected devices sending the malware. The data is obtained by placing honeypot sensors on servers to analyze the infection rates of the counties where residential ISP users generate the attacks. Our analysis of the ISPs where attacks originate builds on previous work where data was obtained with spam traps (van Eeten 2010). Our use of honeypots enables us to capture a sampling of infections beyond spam-generating botnets.

Why ISPs

Preliminary Draft

In tracing the IP addresses responsible for generating spam, malware attacks captured by honeypots, and communicating with the Confiker botnet, van Eeten et. al. (2011) found that 80% were on ISP networks that serve residential and SME users. This tells us two things. First, there is significant insecurity existing in these users of ISPs. Since cybercriminals almost uniformly protect their identities by committing crimes using botnets of infected machines that are distributed throughout the world, we can consider each of these attacking IP addresses as roughly one infection that occurred on the ISPs' users' networks.² Second, once infected, the attacks they execute spread the infection to create more attacking devices. This lack of security in residential and SME ISP and user networks provides an army of infected machines to further advance the volume of cybercrime. ISP users are both victims and resources for cybercrime, making them a particularly productive target of mitigation and prevention efforts. Improving the security on these networks can be rewarding.

Existing research shows that ISPs vary significantly in their infection mitigation and prevention strategies and performance. In a study of the IP addresses of bot-generated spam, van Eeten et al (2010) find that most bots are concentrated on 50 ISPs worldwide. A later study that combined datasets generated by spam traps, honeypots, and sinkholes found significant variation in infection rates among the ISPs of several countries (van Eeten

² This is likely undercounting infections, as there are often more than one device attached to a residential or SME network, affiliated with the same IP address, and more than one may be infected. There are also some complications in the specific counting when dynamic IP addressing is used. See van Eeten et. al., 2011 for a detailed discussion of this.

Preliminary Draft

et al, 2011).³ These differences among ISP security levels suggest they face different incentives (benefits) and/or possess different technology (costs) of mitigation and prevention of infection.⁴ It is generally understood that ISPs lack sufficient incentives to provide the socially optimal level of security on their networks, due largely to spillover effects (Anderson and Moore, 2006; Huang, Geng and Whinston, 2007; Gideon and Hogendorn, 2015). While some models of this market failure allow for the ISP to increase their revenue by providing greater security, based on the premise that this would attract customers (see, for example, Garcia and Horowitz, 2006), there is no evidence that customers consider the relative security of ISPs, when in fact they do have a choice of more than one ISP.

In this paper we explore the differences in security provided by ISPs with respect to ISP market structure and other demographic and technology intensity factors at the county level. The following section provides background on the role of ISPs in providing protection against malware and cybercrime, and on their incentives to do so. Section three describes the approach we use to answer the research question, and section four describes the data collected to date. Data collection is ongoing. The data, analysis and results presented herein are preliminary. In section five we present the econometric analysis of our data. We provide preliminary conclusions and discussion of next steps in section six.

³ Note that van Eeten et al 2011 finds the variance in ISP infection rates for the US is lower than most countries when measured based on the data from the spam traps, the honeypots, and the sink hole. This may limit the generalizability of our results using infected devices on ISP networks in the US.

⁴ This is based on a rather strong assumption that other non-related factors of infection rates, such as user hygiene, will be similarly distributed across the users of the different ISPs. We return to this in our data discussion and analysis.

2. Background

Potential role of ISPs

There is no obvious obligation on the part of ISPs to prevent infections and attacks on their users' networks. Certainly they are not the bad actors. Much of the focus on the ISPs is based on their expertise and structural position to provide greater security for much of the Internet ecosystem.⁵ However, as commercial entities, they lack incentives to incur costs that create a diffuse benefit not easily monetized. While we would expect users to value security in their broadband connections, they often cannot distinguish the relative security provided by different ISPs, or even by ISPs versus the software on their own devices and other elements of the Internet. Accordingly, users tend not to consider security a criteria when choosing a broadband connection (Rowe & Wood 2013).

In fact there is no consensus on the role of ISPs in cybersecurity nor on any obligation implied by their expertise and connection to less knowledgeable users. The debate on the responsibility of the ISPs for cybersecurity arguably originates in the even broader debate of whether any kind of communication provider should be responsible for activity that takes place using the services they provide. This reaches back before commercial use of the Internet. The first notable incidents of such activity in the US occurred during the 1980s, and the response was telling: each time, the affected party did little to alter their system's security, and instead called upon law enforcement. Lack of specific computer-related crime

⁵ However, one can argue that ISPs have less competence in providing security than the cybersecurity firms such as Norton, McAfee, Threatstream, and Trendmicro.

Preliminary Draft

laws at the time limited law enforcement. By 1990 US law enforcement was better equipped to prosecute computer crime, most notably in Operation Sundevil (“The History of Doom” 1990; Sterling, 1994). This early involvement of law enforcement, rather than network operators, was also the beginning of establishing norms of juridical responsibility regarding the ISPs in the US. One might interpret such widespread sting activity of federal law enforcement as a signal that protection against cybercrime is entirely in the domain of proactive government police action, not the establishment of greater security by organizations providing the network services that enable the threats.

The role of the ISP collaborator as informant continues to develop. In 2001, the Budapest Convention on Cybercrime primarily focused on imposing global penalties on virus authors, but did promote a role for ISPs by ensuring State Parties adopt legislation compelling ISPs to ‘cooperate and assist the competent authorities in the collection or recording of traffic data [and] keep confidential the fact of the execution of any power provided for in this article and any information relating to it,’ (Convention on Cybercrime, 2001). Similarly, the US established information sharing procedures between the US government and the ISPs with Presidential Decision Directive 63, further establishing a role for service providers as cooperating information providers to law enforcement, with no mention of minimal security standards or obligations (Moteff, 2015; Palfrey 2000).

Preliminary Draft

By the mid 2000s, with the proliferation of self-propagating botnets, ISP obligations were proposed, sometimes even with the insinuation of liability.⁶ A popular and continuing feature of these debates is the role of the informal inter-ISP cooperative institution in addressing cybercrime. Dourado (2012) argues that such institutions are more efficient than the proposed formal legal regimes, as ISPs can use peering agreements to enforce mechanisms against bad actors. Meanwhile, Bechtold and Perrig (2014) argue that with the possibilities offered by new Internet architecture, more accountability will have to be built in at every level, including for ISPs.

In a 2011 study of botnet detection and mitigation, the European Network Information Security Agency (ENISA) proposed three high-level objectives for reducing cybercrime: mitigation of existing botnets, prevention of new infections, and minimizing the profitability of botnets and cybercrime. They recommend a role for ISPs in the first of these objectives only. In their specific recommendations for ISPs, they recognize the conflict between the ISP's position to take a highly active role and the invasion of their customers' privacy this might entail, as well as their potential loss of reputation with customers if they become the bearers of bad news (in notifying users of infections). The resulting recommendation for the role of ISPs is the identification and notification of customers with malicious hosts, though with the provision of increased incentives to do so (Plohmann, Gerhards-Padilla, and Leder 2011). The resulting ISP role remains one of informant.

⁶ See Lichtman and Posner 2004 as one such example in the US, arguing that ISPs fail to adequately disprove why they should not be held to the standards of indirect liability See also Harper 2005, a critical response and example of the controversy surrounding such proposals.

Preliminary Draft

In an empirical study of Dutch ISPs commissioned by the Netherlands Ministry of Economic Affairs, Agriculture and Innovation, van Eeten et. al. (2011) find that ISPs improve their detection of infected machines on their networks without unduly increasing their costs by collaborating with a common platform or clearinghouse to provide the necessary intelligence to all participants, as is done in Australia. They also note such a clearinghouse can serve to provide light-handed industry self-regulation. Huang et. al. (2007) also propose collaboration among ISPs, though of a different nature. They discuss how ISPs can engage in cooperative filtering and cooperative smoothing by caching and improve their security. In addition to Australia, collective efforts are underway in other countries, including the Netherlands Anti-Botnet Working Group. Public-private initiatives are also seen, including in Germany, Japan and Korea.

Incentives

After many years of limited implementation of technologies known to provide effective cybersecurity, attention has shifted to understanding the incentives of ISPs and other entities that prevent them from using the existing technologies to provide more secure networks. Huang et al. (2007) describe a broken incentive chain, illustrating how the ISPs are positioned to make the investment in greater security but are rarely compensated for the benefit it provides to Internet content providers and end users. Others also identify similar spillovers that result in underinvestment by the ISPs (Anderson and Moore, 2006; Garcia and Horowitz 2007). The impact of the spillovers are further complicated when there is competition in the ISP market (Gideon and Hogendorn, 2015). Current business models in the broadband value chain also impact the ISP incentives to invest in greater

Preliminary Draft

security. Huang et al (2007) show that the subscription payment model, which leads to the practice of maintaining residue bandwidth, as dampening any incentive to participate in cooperative filtering as the ISP is already protected against traffic surges. In this paper we focus on how the presence of competition in the ISP market impacts the level of security observed on the ISPs' networks.

3. Approach

We have developed two distinct and complementary strategies for collecting residential network botnet infection data using honeypot sensors. One of these strategies uses honeypot sensors placed on servers to detect and report malware attacks generated by infected devices on residential ISP networks. We trace the attacks to their ISP and county of origin to determine the rate of infection by county in the US. All data presented and analyzed in this paper is collected using this strategy. The second strategy involves placing honeypot sensors directly on residential networks to record attempted malware attacks. This strategy is not yet implemented. A description of this strategy and the obstacles to its implementation are included below.

Honeypot sensors

A honeypot is a device programmed with software to simulate a vulnerable user and so attract intrusion attempts and malicious code. The honeypot observes the attacks, recording data such as the source and nature of the attack. Honeypots are commonly used

Preliminary Draft

to enhance the security of a network by providing intelligence regarding potential attacks and existing vulnerabilities. In this study we simply use the honeypots to observe attacks and compile these observations into a dataset. Our sensors run software designed by Modern Honey Network.⁷ We deployed Kernel-based Virtual Machines (KVM) running Ubuntu 12.04.5 on servers physically located in Frankfurt, Singapore, or New York City. Each KVM was running a single form of honeypot software designed to make the sensor detect and record identifiable information about potential attackers, with some causing their KVM to emulate different vulnerabilities.⁸

Our sensors are configured to collect, for each attack, the source IP address, the destination IP address, the protocol used in the attack, the source port, the destination port which sensor was attacked, which honeypot was running on that sensor, the time, and any generated signature data. Combining this with other data sources allows us to use the IP address to identify the ISP of the source and its geographic location. This data is then used to determine the specific county of the infected device and the number of ISPs serving the county. We combine this with other county-level demographic and technology-intensity data to analyze the effect of the ISP market structure on the frequency of infected machines engaging in attacks.⁹

⁷ See <https://www.threatstream.com/blog/mhn-modern-honey-network> and <http://threatstream.github.io/mhn/> for more information on Modern Honey Network and its software.

⁸ Data since February 13, 2016 also incorporates KVMs that were running multiple honeypots in tandem (specifically two KVMs per city) to increase data collected. The combinations of honeypots are always the same per city to ensure regularity.

⁹ Our methodology of tracing the attack IP addresses back to the ISP of origin is based on the process used by van Eeten et al 2011 and 2010, described below.

Preliminary Draft

We placed sensors on Digital Ocean data center servers in New York City, Singapore, and Frankfurt. Data was collected from July 24 – November 12, 2015 and February 13 – April 20, 2016.

Attack sources

As in the prior empirical studies discussed above, the attack data collected provides a sample of infected devices. Intrusion attempts and malware attacks overwhelmingly originate from bots. Like van Eeten et al, (2010 and 2011), we can consider the source IP address of each attack to represent an infection that was not prevented by that user or their ISP.¹⁰ By tracing the IP addresses back to the ISPs and identifying their geographic market, we can evaluate the role of industry structure and other demographic and technology-intensity factors on the level of security provided for this sample.

Considerations

There are some limitations to the conclusions we can draw based on this data collection strategy. First, as indicated by the minimal overlap of IP addresses found by three different datasets generated by (1) spam traps, (2) honeypots, and (3) a known botnet sinkhole in van Eeten et al. 2011, our use of honeypots will not provide a comprehensive picture of IP addresses with infected devices. This may impact the results of our analysis of source attacks. Another potential limitation is that we analyze only those IP addresses based in the US.¹¹ Thus our sample may exhibit less variance than if it included other countries.¹²

¹⁰ Data of attack sources collected by the honeypots is first cleaned to remove IP addresses that are not part of Tier 2 ISP networks, and corrected for dynamic IP addressing.

¹¹ We intend to expand this study to include international markets as noted below.

Preliminary Draft

It is also important to remember the dynamic nature of the cybercrime industries and their activities. The cybercriminals have proven themselves to be highly innovative and adaptive. Thus there are limits to any predictions we can make about attack data generated in 2015-16 based on results of studies done only a few years prior. This is seen in the observations of van Eeten et al 2011, where their results show a dramatic shift in the composition of cyber attacks, with the proportion of attacking IP addresses engaged in generating spam making a sudden surge to far exceed those captured by the honeypots from the fourth quarter of 2009 to the first quarter of 2010. Similarly, the attack activity observed in the current study cannot be considered predictive of future attack behavior. However, the sources of incentives for ISPs to invest in the security level they provide to their customers are less malleable, though this of course may also be responsive to changes in attack strategies, business models, and policies.

Malware Attack Detection on Residential Networks

This strategy collects data by placing honeypots designed to capture malware and other malicious intrusion attempts at multiple residential points on different ISP networks, some monopolists and some facing competition. We use this data to look at the difference in intrusion attempts across the destination ISPs, as well as the difference in the ISPs from which the attacks originated, as the attacks generally come from infected machines. Our study of intrusion attempt data by destination ISP is a new approach to understanding

¹² See van Eeten et al 2011 for a comparison of infections by country and variation of infection among ISPs by country. While there is variance in the infection rates of ISPs in the US, it is less than for most other countries in all measures shown.

Preliminary Draft

security investment incentives, enabling us to better identify the effect of ISP industry structure on protection against such intrusions. This strategy is designed to simply measure the number and nature of attack attempts that arrive on the users' networks of ISPs facing different competitive situations. Such attacks are detected and reported by honeypot sensors placed on numerous residential networks. We test this strategy by deploying laptops running Ubuntu 15.04 on residential networks.

Figure 1 illustrates honeypot placement within the relevant elements of the Internet structure.

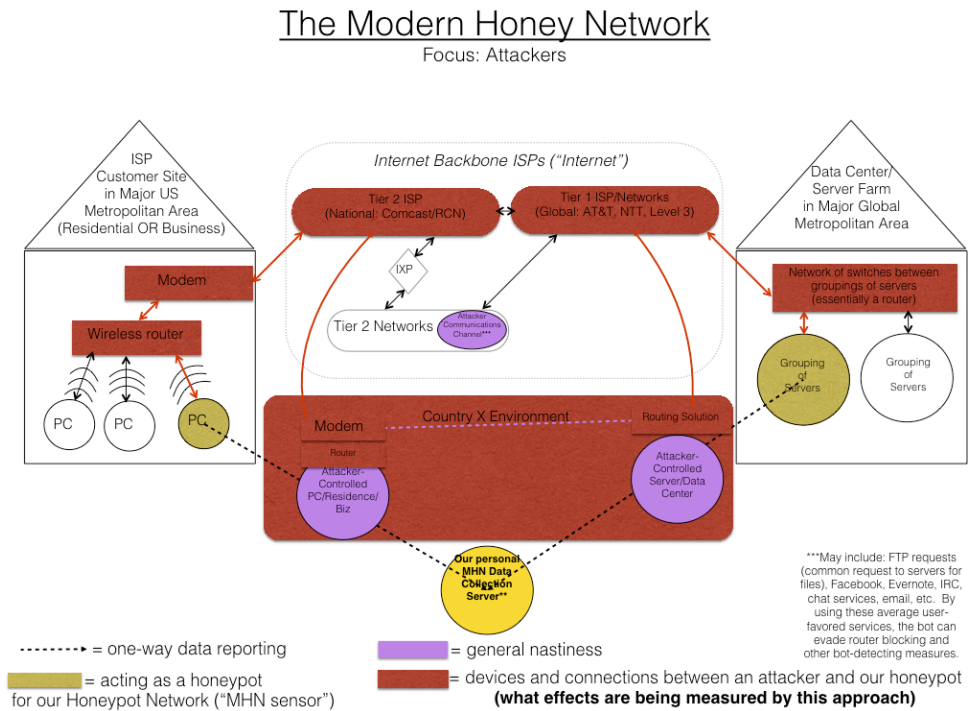


Figure 1: Honeypot Placement

Preliminary Draft

As most infections currently occur through email attachments or clicking on websites, an interactive honeypot sensor that can simulate website visiting and clicking activity will be most useful. The minimal overlap of IP addresses engaged in different forms of attacks (van Eeten 2011) will not present a problem in our analysis of attacks observed on the different ISP networks as they will all be monitored with the same types of sensors. Also there is no reason to expect the networks would be subject to different attacking populations.

This strategy has proven difficult to implement. One interesting surprise in our test placement of sensors thus far is how difficult it is to get attacked on the residential networks, indicating that the level of security provided by the ISPs might be quite good. Why, then, are there so many bots? One possibility is that the number of bots in the US is quite low (as our sensors are currently limited to US networks). Indeed, van Eeten et al (2011) indicates that the US ranks fairly highly on an international comparison of three different types of infections. The persistent prevalence of infection, however, though lower than in most countries, implies that there are attacks that we were not detecting in our sensors. We are currently experimenting with different types of sensor software in order to find which will identify and capture the attacks that infect US ISP users. We are also experimenting with an alternative placement configurations of our residential sensors. The router in a residential network is often responsible for blocking potential attacks, as it will only accept what it is programmed to receive. A successful attack must 'trick' the router into thinking it is something it expects. This implies that an ISP with a larger volume and/or variety of attack attempt activity detected before going through the router is likely to have a higher probability of malware successfully gaining access to the users' premises.

Preliminary Draft

Thus we are experimenting with sensors placed between the ISP and the router to detect the activity attempting to enter the residential network.¹³

A potential limitation of this strategy is really one of our implementation of it, not the strategy itself. Due to resource constraints we are restricted to local placement of our sensors. Thus our samples for the different ISPs may show less variance than they would in other countries, or in a data set with sensors placed in multiple countries. This would be the ideal data collection design, and we plan expanding this study with international sampling at a later date. Thus any effect of industry structure found may be less than we would find in other countries.

4. Data¹⁴

Using sensors placed on servers hosted by Digital Ocean as described above, we collected data on 13,304,556 attacks as of April 20, 2016. We refined this data to find the attacks were executed from 154,955 unique attackers, as indicated by the IP addresses collected for each attack. Using a process similar to van Eeten et al 2010 and 2011, we then used a geolocation database to find the location of the ISPs associated with each IP address, finding 13,625 (8.79%) of the attacking IP addresses are located in the US. We then eliminated any IP addresses that were not from ISPs serving residential users in the US.¹⁵

¹³ As these tests are still in process, we do not report residential network sensor data in this draft.

¹⁴ As data collection is still in process, there are limited results reported in this preliminary draft.

¹⁵ For this study we analyze only ISPs in the US as we are using market structure data from the National Broadband Map created by NTIA and the FCC (found at

Preliminary Draft

Table 1 describes the results of this data cleaning process. The result is a dataset of 3,777 unique attacking US residential IP addresses, identified by county. With these IP addresses located by county, we create our dependent variables measuring infection rates by county. Attacking IP addresses are located in 820 of the total 3,230 US counties listed in the National Broadband Map.¹⁶ While this is only 25.39% of the counties, it includes 75.77% of the population and 75.3% of the households passed by any wireline broadband service. We then combine this with market and demographic data by county from the National Broadband Map and demographic and survey data used in Forman, Goldfarb and Greenstein, 2012. Table 2 provides a description of our market structure and demographic variables.

Table 1: Data Compilation

No. attacks recorded by Honeypot sensors	13,304,556
No. unique IP addresses recorded by Honeypot sensors	154,955
No. unique IP addresses in US	13,625
No. unique IP addresses in US residential ISPs	3,777

Using the collected attacking IP addresses and their locations, combined with National Broadband Map county data, we create variables to measure the extent of infected residential devices in each county. We consider three different infection rate variables: (1) number of attacking IP addresses per 100k population (`attack_rate_pop`), (2) number of attacking IP addresses per 100k housing units (`attack_rate_hh`), and (3) number of

<http://www.broadbandmap.gov/analyze>). Any attack source not from an ISP included in the National Broadband Map was excluded from the dataset.

¹⁶ For information about the National Broadband Map see <http://www.broadbandmap.gov/about>.

Preliminary Draft

attacking IP addresses per 100k housing units passed by any wireline technology (attack_rate_hhpassed).

Our primary focus is on the effect of ISP industry structure on the number of infections found. We create two variables to represent this, number of ISPs serving the county, obtained from the National Broadband Map, and a squared term for the number of ISPs to allow for a nonlinear relationship.¹⁷ These variables are imperfect as sometimes the ISPs listed for a county do not serve the entire county. Thus a county with more than one ISP may actually be a monopoly market if there is no overlap of territory served by the ISPs.. When using number of ISPs as our market structure independent variable, we may be including some very small ISPs that serve only a select small market within the county, though all ISPs included serve a minimum of 4% of the county.

We also consider other factors that might contribute to differences in infection rates among broadband users by county. The data from Forman, Goldfarb and Greenstein 2012 is particularly helpful as it contains, in addition to the usual demographic factors, variables created by surveys that indicate the level of technological intensity of a county, which may be interpreted as indicators of the level of sophistication of broadband users of a county. Using these factors we can approximately correct for the level of protection users can provide themselves. Table 3 provides descriptive statistics of all variables for all 3,230

¹⁷ We also attempted dummy variables indicating a monopoly market, a duopoly market, or more than three providers. This was not useful, though, as the National Broadband Map includes many wireless ISPs so the number of ISPs per county was often quite high. See Table 3.

Preliminary Draft

counties. Table 4 contains descriptive statistics for these variables in the 820 counties where attacks from residential ISP IP addresses are observed.

Table 2: Variable Descriptions

Variable	Definition	Source*
attack_rate_pop	number of attacking IP addresses from residential ISPs / 100k county population	Authors and NBM
attack_rate_hh	number of attacking IP addresses from residential ISPs / 100k county housing units	Authors and NBM
attack_rate_hhpassed	number of attacking IP addresses from residential ISPs / 100k county housing units passed by anywireline technology	Authors and NBM
numISPs	number of ISPs in the county (including wireless) in 2014	Counted from NBM
surv_deeppost00	% businesses using advanced Internet in 2000	FGG
surv_pceremp00	PCs per employee in 2000	FGG
surv_shalpost00	% businesses using basic Internet in 2000	FGG
indivhomeinternet00_cty = anywline	% of households with Internet at home in 2000 % of population with access to any wireline broadband in 2014	CPS in FGG NBM
dem_race_black	% of black people in the geography in 2014	NBM
dem_educ_bachorgreater	% people with bachelor's degree or higher in 2014	NBM
dem_educ_hsgrad	% people with hs diploma in 2014	NBM
dem_inc_poverty_100	% pop under 100% poverty level in 2014	NBM
median_income	Median income for geography	NBM
carnegie1_enr	Per capita number of students enrolled in local PhD-granting institutions	Downes-Greenstein (2007) in FGG
frac_in_eng_pro	Per capita number of students enrolled in engineering programs at local universities	Downes-Greenstein (2007) in FGG US Patent Office in
npatent1980s	Total number of patents from inventors located in county, 1980-1989	FGG
frprof	% of county's work force employed in professional occupations in 2000	Census in FGG
dem_age_greater60	% of people >60 yrs. old in the county in 2014	NBM
netmig95	Net migration to county in 1995	Census in FGG

*

NBM = National Broadband Map

FGG = Forman, Goldfarb and Greenstein, 2012

Preliminary Draft

Table 3: Summary Statistics for All Counties:

Statistic	N	Mean	St. Dev.	Min	Max
numISPs	3,230	8.228	3.023	1	22
numattacks	3,230	1.169	5.965	0	223
dem_race_black	3,230	0.081	0.150	0.000	0.910
dem_inc_poverty_100	3,230	0.177	0.082	0.000	0.660
dem_educ_hsgrad	3,230	0.769	0.091	0.340	0.960
dem_educ_bachorgreater	3,230	0.168	0.079	0.030	0.640
dem_age_greater60	3,230	0.230	0.059	0.030	0.940
anywline	3,230	0.869	0.150	0.000	1.000
surv_shalpost00	2,742	0.720	0.219	0.000	1.000
surv_deeppost00	2,742	0.089	0.133	0.000	1.000
frprof	3,131	0.352	0.066	0.160	0.674
frac_in_eng_prog	3,131	0.001	0.006	0.000	0.112
carnegie1_enr	3,131	0.007	0.065	0.000	2.615
indivhomeinternet00_cty	3,131	0.031	0.116	0.000	0.765
npatent1980s	3,131	0.137	0.652	0.000	20.417
surv_pcperemp00	2,741	0.226	0.172	0.000	1.937
netmig95	3,131	0.252	3.628	-138.933	72.891
numISPs_squared	3,230	76.839	57.932	1	484
median_income	3,230	46.176	13.493	11.185	123.058
attack_rate_pop	3,230	0.806	4.162	0.000	179.533
attack_rate_hh	3,230	1.554	5.179	0.000	141.044
attack_rate_hhpassed	3,228	1.791	6.634	0.000	154.993
attacked	3,230	0.254	0.435	0	1

Table 4: Summary Statistics for Counties with Attacking IP addresses Observed

Statistic	N	Mean	St. Dev.	Min	Max
numISPs	820	9.482	3.084	3	22
numattacks	820	4.606	11.154	1	223
dem_race_black	820	0.093	0.128	0.000	0.730
dem_inc_poverty_100	820	0.156	0.056	0.030	0.400
dem_educ_hsgrad	820	0.805	0.079	0.340	0.950
dem_educ_bachorgreater	820	0.223	0.096	0.070	0.640
dem_age_greater60	820	0.219	0.044	0.090	0.500
anywline	820	0.945	0.087	0.180	1.000
surv_shalpost00	803	0.758	0.168	0.000	1.000
surv_deeppost00	803	0.099	0.084	0.000	1.000
frprof	820	0.387	0.071	0.221	0.666
frac_in_eng_prog	820	0.002	0.007	0.000	0.112
carnegie1_enr	820	0.015	0.055	0.000	0.612
indivhomeinternet00_cty	820	0.104	0.197	0.000	0.765
npatent1980s	820	0.447	1.211	0.000	20.417
surv_pcperemp00	803	0.290	0.156	0.000	1.017
netmig95	820	0.430	6.973	-138.933	72.891
numISPs_squared	820	99.401	68.816	9	484
median_income	820	54.511	15.220	26.828	123.058
attack_rate_pop	820	3.173	7.796	0.107	179.533
attack_rate_hh	820	6.121	8.818	0.267	141.044
attack_rate_hhpassed	820	7.050	11.674	0.267	154.993

5. Analysis

Using ordinary least squares econometric analysis we perform preliminary estimation of the effect of these different factors on the observed rate of infection by county. We consider three dependent variables as described above. Table 5 provides a description of the preliminary regression analysis results. Each observation in this analysis is a county. Our analysis includes 803 counties where attacking IP addresses were observed. This excludes the 27 counties that were eliminated because they were missing data. In these cases the data missing was usually for the Forman, Goldfarb and Greenstein 2012 survey variables.

The results of the preliminary regression analysis show that industry structure is a significant factor in the level of protection provided by ISPs. The coefficients for the number of ISPs and number of ISPs squared are both significant at the 1% level for all three measures of infection rate. The significance of the squared term shows that the relationship between the number of ISPs and the number of infections is nonlinear. The signs of the coefficients of the number of ISPs and the number of ISPs squared indicates a U-shaped relationship. This suggests that when there is a small number of ISPs, an increase in the number of ISPs brings a decrease in the number of infections. As the number of ISPs increases, the decrease in number of infection diminishes, and even increases at larger number of ISPs. This implies that the introduction of competition into a monopoly or duopoly ISP market can decrease the infections, i.e. improve the level of security provided. The increase in number of infections with additional ISPs in a market already served by

Preliminary Draft

several ISPs may reflect the presence of smaller ISPs who may lack the resources to provide adequate security or are disreputable in some other way in more heavily served areas. As shown on Table 4, the mean number of ISPs per county is 9, with as many as 22 ISPs in a county.

Many of the other market and demographic variables are significant in all three models.

The percent of population with access to any wireline broadband in 2014 is significant with a negative coefficient of large magnitude. This implies that the more population with access to wireline broadband the fewer attacking IP addresses observed in a county. This is a puzzling result as we would expect the opposite effect. The percentage of the county population that is black, the percentage of county population below the poverty level, and the percent of the population older than 60 are all significant with negative coefficients. This could be because in such counties a higher percent of the population are not using broadband and thus cannot be infected. Alternatively, there are fewer total devices or less total broadband use per population and household, reducing opportunities for infection. However, the median income is significant and negative, implying more income leads to less infection, contradicting this reasoning. This is also puzzling, and indicates there may be one or more missing explanatory variables. The education level variables are not significant.

Most of the correcting variables reflecting technological sophistication of the users in a county were not statistically significant. The percent of businesses using advanced Internet in 2000 (`surv_deeppost00`) is significant at the 5% level in models (1) and (2) and at the

Preliminary Draft

10% level in model (3). The negative coefficient in all models indicates infections are lower in counties with a higher percentage of businesses using advanced Internet in 2000. This may reflect a higher level of user sophistication, which would lead to better digital hygiene or possibly users demanding higher levels of security from their ISPs.

Table 5: Regression Results

	Dependent variable:		
	attack_rate_pop (1)	attack_rate_hh (2)	attack_rate_hhpassed (3)
numISPs	-0.685*** (0.157)	-1.084*** (0.304)	-1.254*** (0.355)
numISPs_squared	0.023*** (0.007)	0.035*** (0.013)	0.041*** (0.016)
surv_deeppost00	-3.154** (1.232)	-4.959** (2.397)	-5.260* (2.798)
surv_pcperemp00	0.903 (0.983)	1.117 (1.912)	1.623 (2.232)
surv_shalpost00	-0.209 (0.748)	-1.357 (1.454)	-1.677 (1.698)
indivhomeinternet00_cty	-1.065* (0.547)	-2.052* (1.064)	-1.260 (1.242)
anywline	-10.318*** (1.340)	-18.307*** (2.607)	-50.029*** (3.044)
dem_race_black	-3.035*** (0.879)	-6.491*** (1.710)	-5.927*** (1.996)
dem_educ_bachorgreater	5.106 (3.425)	7.404 (6.661)	9.715 (7.777)
dem_educ_hsgrad	-3.882 (2.389)	-8.408* (4.646)	-5.714 (5.424)
dem_inc_poverty_100	-19.804*** (3.893)	-36.409*** (7.572)	-39.874*** (8.840)
median_income	-0.097*** (0.016)	-0.166*** (0.031)	-0.171*** (0.036)
carnegie1_enr	-2.128 (2.872)	-4.102 (5.585)	-4.899 (6.521)
frac_in_eng_prog	-13.990	-18.330	-29.278

Preliminary Draft

	(19.476)	(37.876)	(44.220)
npatent1980s	0.027 (0.110)	0.078 (0.215)	0.234 (0.251)
frprof	2.436 (4.205)	6.268 (8.177)	5.585 (9.547)
dem_age_greater60	-2.637 (2.543)	-18.706*** (4.945)	-22.594*** (5.773)
netmig95	-0.010 (0.017)	-0.019 (0.033)	-0.013 (0.039)
Constant	27.195*** (2.927)	53.255*** (5.692)	84.152*** (6.645)

Observations	803	803	803
R2	0.245	0.217	0.381
Adjusted R2	0.227	0.199	0.366
Residual Std. Error (df = 784)	2.704	5.258	6.139
F Statistic (df = 18; 784)	14.120***	12.045***	26.770***

Note: *p<0.1; **p<0.05; ***p<0.01

6. Conclusion

In this paper we present preliminary results based on cyberattack data gathered with honeypots on server farms. Each attacking IP address is considered to be a botnet-infected device. Tracing the IP addresses back to their originating ISPs and counties, we identify those that are part of residential networks in the US. Our preliminary analysis of data collected to date suggests that industry structure in the ISP market is a significant predictor of botnet infections on residential networks. The results suggest that when there is a small number of ISPs serving a market, the entry of an additional ISP to the market can bring a reduction in the number of residential machines infected to perform attacks with malware. In other words, a market with an adequate level of ISP competition will experience better security. This may be due to the increased incentives to provide better security when

Preliminary Draft

competing for users. Alternatively, it may reflect the increased probability that infections will be deflected from more ISPs providing security in different ways. This could result from the spillover of the benefits of good protection from the ISP that provides the strongest protection. At larger numbers of ISPs per county, we see this decrease in infections abate, or even reverse.

Further research is needed to better evaluate and understand the relationships found in this paper. The research discussed in this paper is an ongoing project. Data collection activity continues as we do not yet have a sufficiently representative sample. We also continue to explore alternative specifications of the econometric model. One particular concern is that the variables created to capture competition are insufficiently precise, and may be capturing county characteristics. We continue to consider ways of addressing this, including alternative measures of broadband competition, additional explanatory variables, and instrumental variables for possible endogeneity. We also plan to expand this research project on two dimensions. One expansion is the inclusion of international ISP markets in our analysis. Our collected data contains many attacking IP addresses from outside the US that can be included in the analysis with comparable market data. Another planned expansion is the implementation of the residential network honeypot sensor data collection strategy.

Preliminary Draft

References

- Anderson, Ross and Moore, Tyler. 2006. "The economics of cybersecurity." *Science* 314:610-13.
- Bechtold, Stegan and Perrig, Adrian. 2014. "Accountability in Future Internet Architectures." *Communications of the ACM*. 57:9 (September).
- Center for Atrategic and International Studies. 2014. *Net Losses: Estimating the Global Cost of Cybercrime*. Intel Security. June.
- Convention on Cybercrime, November 23, 2001. Available at <<http://conventions.coe.int/Treaty/en/Treaties/Html/185.htm>>, accessed July 16, 2015.
- Dourado, Eli. 2012. "Internet Security Without Law: How Service Providers Create Order Online." Working Paper No. 12-19. George Washington University Mercatus Center. June.
- Forman, Chris, Goldfarb, Avi, and Greenstein, Shane. 2012. "The Internet and local wages: A puzzle." *American Economic Review* 102(1): 556-575.
- Garcia, Alfredo and Horowitz, Barry. 2007. "The potential for underinvestment in internet security: implications for regulatory policy." *Journal of Regulatory Economics*. 31:37-55.
- Gideon, Carolyn and Hogendorn, Christiaan. 2015. "Safety in numbers? The effect of network competition on cybersecurity." Presented at ASSA meetings January 4, 2015, Boston, MA.
- Harper, Jim. 2005. "Against ISP Liability." *Regulation* 28.
- The History of the Legion of Doom," *Phrack Inc.*, published May 28, 1990, <<http://phrack.org/issues/31/5.html>>, accessed July 16, 2015.
- Huang, Yun, Geng, Xianjun, and Whinston, Andrew B. 2007. "defeating DDoS attacks by fixing the incentive chain." *AMC Transactions on Internet technology* 7(1): Article 5, February.
- Lichtman, Doug and Posner, Eric. 2004. "Holding Internet Service Providers Accountable," *John M. Olin Law & Economics Working Paper No. 217*. The Law School of the University of Chicago. July.
- Moteff, John D. 2015. "Critical Infrastructures: Background and Early Implementation of PDD-63," *Congressional Research Service*, June 10.

Preliminary Draft

Palfrey, Terry. 2000. "Surveillance as a Response to Crime in Cyberspace," *Information and Communications Technology Law*. 9(3): 173-193.

Plohmann, Daniel, Gerhards-Padilla, Elmar, and Leder, Felix. 2011. *Botnets: Detection, Measurement, Disinfection & Defence*. European Networks and Information Security Agency.

Rowe, B., & Wood, D. (2013). "Are home internet users willing to pay ISPs for improvements in cyber security?." In *Economics of Information Security and Privacy III* (pp. 193-212). Springer New York.

Sterling, Bruce. 1994. *The Hacker Crackdown: Law and Disorder on the Electronic Frontier*. Available at < <http://www.farcaster.com/sterling/part3.htm> >, accessed July 16, 2015.

Van Eeten, Michel J.G., Asghari, Hadi, Bauer, Johannes, and Tabatabaie, Shirin. 2011. *Internet Service Providers and Botnet Mitigation: A fact –finding study of the Dutch market*, Report prepared Netherlands Ministry of Economic Affairs, Agriculture and Innovation. January.

Van Eeten, M., Bauer, J. M., Asghari, H., Tabatabaie, S., & Rand, D. (2010). "The role of internet service providers in botnet mitigation: An empirical analysis based on spam data." (No. 2010/5). OECD Publishing

Wheeler, Tom. 2014. Remarks at American Enterprise Institute. June 12. Available at <http://www.fcc.gov/document/chairman-wheeler-american-enterprise-institute-washington-dc>. Accessed August 12, 2015.