The fight against botnets has been going on for more than a decade, but they still impose significant costs. ISPs have become increasingly central to the effort, as they can undertake mitigation more economically than end users. A study evaluates the role and performance of ISPs in botnet mitigation across 60 countries.

It’s been more than a decade since the start of the fight against botnets—networks of computers that are infected with malware and controlled by criminals. The many countermeasures have involved end users, ISPs, industry associations, and government. Despite some successes, botnets are still among our most urgent security threats. Botnet mitigation has occurred under the specter of dire predictions about how defenders are losing ground against innovative criminals. Security vendors tend to haunt us with stories about the exponentially growing number of malware variants. However, the empirical facts are less alarming. Microsoft publishes what’s arguably the best available data on malware infection rates. Roughly speaking, at each cleanup cycle, approximately 1 percent of all Windows computers running automatic updates were infected. Most of these machines were unique from one month to the next, which means that the average rate of 1 percent infected users per month translates to roughly one in 10 users experiencing an infection in the course of a year. In an earlier study using a completely different method, we came to a similar estimate. However, infection rates have been relatively stable since 2009 and look nothing like the dire predictions.

That said, even these infection rates imply significant direct cost to our economies. The cost of cleanup alone is estimated at a few billion US dollars worldwide. The overall direct and indirect costs of the criminal business models based on botnets are very hard to estimate. They likely run into the tens of billions per year. The scourge of botnets can’t be adequately understood as a mere technical problem. After all, malware has been around since the 1990s, well before botnets emerged. The rise and persistence of botnets reflect changes in the underlying economics of both attackers and defenders—for instance, end users don’t bear the full cost of infections. This has led to increasing pressure on ISPs to undertake mitigation and has prompted the launch of national initiatives to support ISPs in this effort.

Economic Incentives of Attackers and Defenders

The global malware outbreaks of the early 2000s, such as the ILOVEYOU and CODE RED computer worms,
were disruptive and highly visible. Their authors seemed motivated by the quest for notoriety. Then, as more economic transactions moved online and the cost of abusing vulnerabilities decreased, profit-driven criminals entered the scene and rapidly expanded their activities.\(^4\) Their incentives changed malware from visible and disruptive to stealthy code that kept victims’ machines up and running as part of a criminal infrastructure. Criminals discovered an expanding array of business models to monetize these infected machines: sending spam, performing distributed denial-of-service attacks, harvesting user credentials, committing financial fraud, hosting phishing sites, performing click fraud on advertising networks, and more.

As cybercrime expanded, the underworld’s economic organization also changed. An increase in specialization (for example, malware authors and botnet herders), the emergence of markets for attack tools and services, and a new complex system of monetizing online crime contributed to increases in the virility of attacks. The global migration to broadband, the ability to move attacks in an agile way across national borders, and the limited reach of national law enforcement further boosted the benefit–cost ratio for cybercriminals pursuing a broadening range of online crimes.\(^5\)

However, criminal incentives are only one side of the problem. Another crucial part of the problem relates to the incentives of the defenders—most notably the owners of the infected machines. Botnets typically attack third parties, not owners. This reduces the odds that owners will discover the infection and, more important, undermines their incentive to better secure their machines, as the damage is borne by others or society at large.\(^6\) At the same time, the benefits of investment in security partially accrue to other users. Hence, “private” cost and benefits of security as seen from an individual user’s perspective deviate from “social” costs and benefits to society at large. This is a classic form of an economic externality (the direct effect of the activity of one actor on the welfare of another that is not compensated by a market transaction)—a form of market failure.

ISP users, however, are increasingly called on to undertake mitigation. They’re seen as a natural control point because they’re the infected machines’ gateway to the Internet. Industry groups such as the Messaging Anti-Abuse Working Group and Internet Engineering Task Force, country regulators such as the American Federal Communications Commission, and international organizations such as the Organisation for Economic Cooperation and Development (OECD) have pushed for ISP best practices that include contacting and cleaning up infected customers’ machines.\(^7\)\(^\text{-}^9\)

Complementary to increasing pressure on ISPs, several countries have established national anti-botnet initiatives. These entail national call centers for infected ISP users, codes of conduct for ISPs, and joint mitigation schemes such as centralized clearinghouses that collect and channel infection data to ISPs and their customers.

We empirically assessed the effectiveness of these mitigation strategies. To do this, we first needed a sound way to estimate infection rates across ISPs.

**Measuring Botnet Infection Rates**

To develop relative infection rates for ISPs, we first processed global datasets of botnet activity and extracted infected machines’ IP addresses. Second, we attributed each IP address to an ISP at that point in time. Third, we counted the IP addresses seen in each ISP per day. Fourth, we calculated a normalized measure by dividing this count by the number of subscribers of each ISP. We designed all steps to ensure an effective metric.

**Datasets**

We processed infection data coming from a spam trap—an email server that hosts unused email addresses and hence receives only spam—as well as Conficker botnet and Gameover Zeus botnet sinkholes. The spam trap we used has been running for more than a decade. It logs the IP addresses of machines sending the spam. Because botnets have been the dominant spam distribution mechanism during recent years, many of these machines are bots. We reduced the false-positive rate by including only sources located in retail ISP networks. In 2010, the spam trap recorded an average of 888,000 unique IP addresses and 162 million spam messages per day. Triangulation with security vendors’ spam reports shows that the sample is representative.

Sinkhole servers are used to disrupt botnet command-and-control (C&C) infrastructures. Conficker started spreading in late 2008 and quickly infected millions of Windows machines. Security experts reverse-engineered the malware and sinkholed its C&C infrastructure shortly after. While this effectively neutralized the botnet, cleanup of machines has been slow. The sinkhole logs all bots that connect: in 2010, six million unique IP addresses appeared in the sinkhole every day; in 2014, this number was one million. The Conficker dataset doesn’t suffer from potential sampling biases or false positives when compared to the spam trap; the downside is that it’s an inactive botnet.

Gameover Zeus is considered one of today’s most dangerous botnets, partly owing to its sophistication and association with financial fraud. It has withstood multiple takedown attempts, including an international attempt in June 2014. Our third dataset contained IP addresses from a part of the botnet that has remained sinkholed from this attempt—1.9 million unique IP addresses in 42 days. This provided a recent snapshot to
compare with the longitudinal sets. In short, each of our datasets had its strengths and weaknesses, and there was little overlap among the sources.

Attributing IP Addresses
For each IP address, we looked up the autonomous system (AS) and country in which it was located. ASes are the technical entities that own IP addresses. Given our interest in both criminals’ and defenders’ economic incentives, we wanted to determine the actual legal entities that own the IP addresses. ASes need to be mapped to companies, so for this purpose, we built our own dataset. We started by selecting countries to include in the study. We ranked all countries based on their number of bots and selected those that cumulatively made up the top 80 percent—in total 39 countries. For comparison, we added all other OECD and EU countries and omitted four countries because of insufficient or incoherent market data. The final set contained 60 countries.

For each country, we obtained a list of major companies offering broadband access from the commercial TeleGeography GlobalComms dataset and a list of all top ASes as ranked by the number of bots. We then looked at the WHOIS records for each AS and mapped them to a company. In most cases, the relation was one to one. There were also cases in which one ISP mapped to multiple ASes, or vice versa, which were handled accordingly. In the end, we mapped 2,260 ASes; of these, 681 belonged to one of the 262 broadband ISPs. These ISPs controlled more than 85 percent of the broadband Internet market share in their respective countries. Hence, we can state that our dataset was generally representative of ISP behavior.

Aggregating IP Counts
We counted each ISP’s unique number of IP addresses over a particular time period as a security metric for each ISP. IP addresses have a well-known weakness when used as proxies for machines, as dynamic IP address allocation policies can make the same infected machine appear with multiple IP addresses—the so-called DHCP (Dynamic Host Configuration Protocol) churn.\textsuperscript{10,11} What’s more, these policies vary substantially across ISPs. We mitigated this problem by counting unique IP addresses in much shorter time frames (per day for the spam data and per hour for the Conficker data) and averaging these counts over a year to get the yearly metric. These aggregates—especially the hourly count—underestimate the total number of infected machines at any one point in time. However, this isn’t important as we sought to compare ISPs rather than determine the absolute number of infections.

Calculating Infection Rates
The final step toward acquiring comparative metrics across ISPs was to account for their size differences. ISPs with more subscribers are likely to have more bots simply due to their larger size (see Figure 1). Infection levels are a function of ISP subscribers, which necessitates using normalized metrics or rates. For this purpose, we divided the counts of each ISP’s infected machines by its number of subscribers according to the TeleGeography database. The resulting dataset provides a longitudinal and cross-sectional view on botnet infection rates in ISPs worldwide.
Do ISPs Make a Difference?

Do legitimate ISPs in jurisdictions that care about botnets control the bulk of the infected population? The answer might seem trivial, but it isn’t. ISPs can identify the customer behind an IP address and, as such, are uniquely positioned to contact and quarantine infected customers. However, this doesn’t mean that they can control the bulk of the botnet problem. What portion of the infected population resides in ISP networks as opposed to, say, large corporate, educational, or even cloud networks? Furthermore, are the bots located in the networks of legitimate ISPs that are amenable to codes of conduct or regulatory efforts? Or are they in the networks of shady ISPs operators in jurisdictions with less mature governance structures, outside the reach of industry and the governments pursuing anti-botnet initiatives?

Figure 2 shows the proportion of all spam and Conficker bots located in the networks of 262 ISPs, for various years. We can see a relatively concentrated pattern: more than half of all infected machines are located in only 50 large ISPs. (The pattern holds for all years, except for spam in 2014. As overall spam levels have decreased, a larger proportion of the remaining spam-sending machines reside in non-ISP networks, especially hosting providers.)

Figure 2. Cumulative percentage of infected machines at top ISPs for (a) Conficker and (b) spambots. More than half of all infected machines are located in only 50 large ISPs. The pattern holds for all years, except for spam in 2014. As overall spam levels have decreased, a larger proportion of the remaining spam-sending machines reside in non-ISP networks, especially hosting providers.

Figure 2 shows the proportion of all spam and Conficker bots located in the networks of 262 ISPs, for various years. We can see a relatively concentrated pattern: more than half of all infected machines are located in only 50 large ISPs. (The pattern holds for all years, except for spam in 2014. As overall spam levels have decreased, a larger proportion of the remaining spam-sending machines reside in non-ISP networks, especially hosting providers.)

To put these findings into perspective, estimates of the total number of ISPs on the Internet run anywhere from 3,000 to 30,000. The top 50 are large firms that operate leading brands and are well-known to the regulators in their respective countries.

From a policy perspective, this provides an optimistic lesson: getting a limited number of large players to improve mitigation substantially affects the infection rate.

This brings us to a second question: Do the infection rates of these ISPs substantially differ, or are market incentives dictating a certain level of mitigation across ISPs? Earlier work has suggested that ISPs aren’t incentivized by the market to undertake mitigation and thus avoid its additional cost. If this is true, we should expect similar performance levels, especially among ISPs operating in the same price-competitive market.

As we can see in Figure 1, infection levels are a function of ISP size: more subscribers means more infections (note that both axes are plotted on a logarithmic scale). However, empirical evidence also shows dramatic variations in infection levels. Similarly sized ISPs can have up to two orders of magnitude difference in terms of numbers of infections. Even ISPs in the same country—under the same competitive pressures and regulatory framework—show differences of more than one order of magnitude, as illustrated by US and German ISPs in the figure. The pattern is consistent across all datasets and stable over time. Statistically, the coefficient of variation—a standardized measure of distribution dispersion—is 1.7 for Conficker infection rates and 2.7 for spambot infection rates, putting them both in the high-variance range. The correlation between infection rates in consecutive years is 0.7 for spambot and 0.9 for Conficker, indicating these differences are systematic and not driven by bad luck or transient circumstances.

The main lesson from this finding is that ISPs can and do make a difference. Whether a result of their policies, country policies, or just good luck in their share of Internet users, ISPs are critical control points and have discretion to undertake mitigation at higher levels. Regardless of the specific explanation, this finding is good news for regulators and enough to engage the ISPs. By determining what the better-performing ISPs are doing right, we can teach others how to improve their performance.

Why Do Some ISPs Perform Better?

How can we explain the fact that some ISPs have...
infection rates that are orders of magnitude worse than those of their peers? In principle, these differences can be driven by the ISPs’ policies and differences among their customer bases. ISP actions are guided by incentives and shaped by organizational factors, market conditions, and the regulatory environment (see Table 1). Using several proxies, we explored some of these relations in a regression model.

We captured regulatory efforts in two proxy variables. The first, lap_regulator, indicates whether a country’s regulator has joined the London Action Plan (LAP), a consortium that supports the development of anti-spam and anti-botnet policies. LAP has no commitment power over its members and can’t establish binding policies. We interpret membership as a proxy for regulatory attention to the issues of spam and botnets. Earlier reports found that LAP membership is correlated with lower infection rates.13 Of the 60 countries in our set, 24 had joined LAP. The second variable, anti_botnet, indicates whether a country has a national anti-botnet center. Since 2010, seven countries have had one: Australia, Finland, Germany, Ireland, Japan, Korea, and the Netherlands.8,14 Except for Germany, all these countries are also LAP members.

To control for relevant differences in ISP subscriber populations, we used the unlicensed software rate as a proxy for differences among user populations and a general environment control. Previously known as the piracy rate, this statistic is collected and reported by the Business Software Alliance and indicates the percentage of software installed without a valid license. It reflects user attitudes and is correlated with software-patching practices—higher rates of unlicensed software means more infections. It’s also correlated with population properties such as education and wealth.

We used broadband prices (US dollars for a 1-Mbps connection) and broadband speeds as controls for the market conditions under which the ISP operates. The International Telecommunication Union gathers and reports these country averages. We hypothesized that lower prices reflect price pressure on ISPs, leaving less room for security expenditure. Higher broadband speeds might have an effect in both directions: as an enabler of malware spreading and as an indicator of a more mature and better-managed infrastructure. We left out several other typical country-level variables, such as information and communications technology (ICT) development, to avoid problems of multicollinearity.

We include the log of the number of an ISP’s subscribers as a proxy for its size. There are many reasons to include this variable as it relates to botnet mitigation. First, large ISPs are thought to react differently to regulatory pressure and peer pressure with regard to cleaning up their networks. Second, because of the scale of their operations, large ISPs are likely to have higher levels of automation in abuse handling and cleanup. This might reduce the cost per mitigation action.

We used fixed effect dummies as intercepts for each year. We wanted to control for the large variation in botnet activity across years driven by global attacker and defender behavior, beyond ISPs. Using fixed effects, we created a baseline level for each year against which we could then further distinguish ISP differences.

All these parameter estimates are partial effects under the assumption that all other factors remain unchanged. The variation among ISPs partially reflects differences among subscriber populations. Increased unlicensed

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (standard deviation)</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>P &gt;</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>P &gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>lap_regulator</td>
<td>0.47 (0.50)</td>
<td>0.033</td>
<td>0.059</td>
<td>0.580</td>
<td>–0.163</td>
<td>0.096</td>
<td>0.091</td>
</tr>
<tr>
<td>anti_botnet</td>
<td>0.16 (0.36)</td>
<td>–0.202</td>
<td>0.076</td>
<td>0.008</td>
<td>–0.276</td>
<td>0.134</td>
<td>0.049</td>
</tr>
<tr>
<td>unlicensed_sw</td>
<td>45.3 (19.9)</td>
<td>0.047</td>
<td>0.002</td>
<td>0.000</td>
<td>0.043</td>
<td>0.003</td>
<td>0.000</td>
</tr>
<tr>
<td>broadband_price</td>
<td>24.4 (9.6)</td>
<td>–0.016</td>
<td>0.003</td>
<td>0.000</td>
<td>0.000</td>
<td>0.005</td>
<td>0.933</td>
</tr>
<tr>
<td>broadband_speed</td>
<td>3.9 (6.4)</td>
<td>0.019</td>
<td>0.004</td>
<td>0.000</td>
<td>–0.001</td>
<td>0.006</td>
<td>0.875</td>
</tr>
<tr>
<td>isp_subs_l</td>
<td>5.6 (0.7)</td>
<td>–0.239</td>
<td>0.034</td>
<td>0.000</td>
<td>–0.264</td>
<td>0.052</td>
<td>0.000</td>
</tr>
<tr>
<td>FE_intercepts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
software use correlates with higher ISP infection rates. Differences in our proxies for market forces had no impact. Broadband price and speed coefficient were close to zero and had no effect.

Large ISPs have, on average, fewer infections per customer than smaller ISPs. This is in line with qualitative data from interviews with European and US ISPs about how they handle infected customers: large ISPs are more likely to have automated key parts of their botnet and abuse response process. Using automation helps reduce the cost of identifying, notifying, and mitigating infected customers, thus making it more economically efficient to mitigate on a larger scale.

In terms of policies, we found evidence that having a national anti-botnet center correlates with lower ISP infection rates (negative and significant). An active regulator seems to have significant impact.

An interaction between the policy variables and unlicensed software complicates our interpretation of the regression coefficients: countries that have an anti-botnet center also have considerably lower unlicensed software use. An intuitive way to disentangle such effects is to plot the regression predictor for different variable combinations (see Figure 3). Each curve shows the regression prediction given various combinations of regulatory activity and unlicensed software use. Although the presence of anti-botnet initiatives shifts the curves downward, the effect was relatively small compared to unlicensed software use in the country, which we know also correlates negatively with indexes of wealth and ICT development.

We should note that both regression models have unexplained variance because current models didn’t capture many ISP-level differences. This is an area for further research, for instance, looking at actual network management policies. The unexplained variance can also account for the differences among the spam rate and conficker rate models for the lap_regulator coefficient where they differ.

**Which Policies Are Effective?**

Our analysis suggests that ISP incentives partially drive infection levels and that these, in turn, can be influenced by policy efforts. In the model, we included two explicit policy factors: LAP membership—as a proxy for regulatory involvement—and presence of a national anti-botnet initiative. Only the latter was significant. However, it’s important to realize that the variables overlap: all countries that have a national initiative are also LAP members, except for Germany. So the impact of the anti-botnet variable is really a combined effect of LAP plus the initiatives. LAP members without a national initiative are probably less proactive.

National initiatives change ISP incentives in several ways. First, a national initiative demonstrates government involvement, which puts more pressure on ISPs to invest in security. Second, a national center reduces mitigation cost for ISPs, enabling them to increase their impact with the same resources. For example, the Netherlands’ centralized clearinghouse, called AbuseHUB, is partially government funded. It sets up relationships with suppliers of abuse data, such as the ShadowServer Foundation and Microsoft. It has also automated the parsing of this incoming data and feeds it directly into member ISPs’ automated abuse incident response processes. All this reduces ISP costs.
and scales up mitigation. Anti-botnet centers in other countries, such as Korea and Germany, provide actual customer support via a publicly funded call center. This shifts some of the mitigation cost to the tax payer, reducing the burden on ISPs.

In short, policies that incentivize ISPs appear effective, particularly when they take the form of national anti-botnet initiatives. However, the centers’ impact shouldn’t be overestimated. The extent to which ISPs respond to these reduced costs will differ. In an evaluation of the Dutch initiative, we found that, even though large ISPs received the same data feeds, if and when they acted on this data differed among providers, as evidenced by the fact that their relative infection rates continued to differ by a factor of three to five. We see similar variation in other countries with a national initiative. In the end, anti-botnet initiatives seem to nudge provider policies in the right direction but don’t dictate them.

We also see that policy impact is modest when compared to contextual factors such as the rate of unlicensed software use. Of course, policy can also try to influence piracy rates—and, in many countries, it does, as part of their intellectual property protections. This raises an interesting policy option for botnet mitigation: focusing on the ICT infrastructure’s general health might be the most effective way to reduce the societal burden of botnets.

The botnet battle hasn’t been lost. Infection rates have been relatively stable for several years. At a minimum, the dire predictions have been continually thwarted by the facts. That being said, the economic damage associated with botnets still runs into the tens of billions of dollars per year.

In the search for more effective mitigation, the focus has shifted from end users to Internet intermediaries, most notably ISPs. We empirically tested whether the assumptions behind this more recent strategy hold over time. They do. The problem of botnets isn’t located within the networks of shady ISPs in countries with poor governance structures. Well-known and well-established ISPs in relatively well-governed jurisdictions control the bulk of the problem.

In the ISP population, infection rates differ dramatically—more than two orders of magnitude. Even in the same country, infection rates can differ by more than one order of magnitude. This suggests that ISPs have discretion to enhance mitigation. Their economic incentives aren’t dictated by the need to operate in markets that are primarily driven by price competition.

The differences among ISPs’ infection rates can be understood from other incentives, most notably the cost of mitigation and the pressure of regulatory involvement. Large ISPs have lower infection rates, pointing to the benefits of automation in handling infection incidents, which lowers the cost per cleanup and allows ISPs to better scale up their mitigation. Regulatory involvement—that is, “soft regulation”—not only incentivizes ISPs to exert more effort but has also led to public–private initiatives, such as national anti-botnet centers, that reduce the mitigation cost for ISPs.

One sobering finding is that institutional factors, such as the level of software piracy in a country, might overwhelm the effects of anti-botnet policies. That said, some of these factors might also be the focus of more general cybersecurity policies. Such an approach might be more economically efficient, suggesting that we shouldn’t focus our efforts too myopically on the botnets themselves.

Acknowledgments

We thank Dave Rand, the Conficker Working Group, and Shadowserver for providing the data used for this research. We also thank Shirin Tabatabaie for her contribution to earlier stages of this research.

References


Hadi Asghari is an assistant professor at the Economics of Cybersecurity Group at Delft University of Technology. His research focuses on the economics of cybersecurity and online privacy. Contact him at h.asghari@tudelft.nl.

Michel J.G. van Eeten is a professor of governance of cybersecurity at Delft University of Technology, Faculty of Technology, Policy and Management. Contact him at m.j.g.vaneeten@tudelft.nl.

Johannes M. Bauer is a professor and chairperson of the Department of Media and Information at Michigan State University. His research, teaching, and writing focus on the digital economy. Contact him at bauerj@msu.edu.