Post-Mortem of a Zombie: Conficker Cleanup After Six Years

Hadi Asghari, Michael Ciere, and Michel J.G. van Eeten, ***Delft University of Technology***

https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/asghari

This paper is included in the Proceedings of the 24th USENIX Security Symposium

August 12–14, 2015 • Washington, D.C.

ISBN 978-1-939133-11-3

Open access to the Proceedings of the 24th USENIX Security Symposium is sponsored by USENIX

Post-Mortem of a Zombie: Conficker Cleanup After Six Years

Hadi Asghari, Michael Ciere and Michel J.G. van Eeten
Delft University of Technology

Abstract

Research on botnet mitigation has focused predominantly on methods to technically disrupt the command-and-control infrastructure. Much less is known about the effectiveness of large-scale efforts to clean up infected machines. We analyze longitudinal data from the sinkhole of Conficker, one the largest botnets ever seen, to assess the impact of what has been emerging as a best practice: national anti-botnet initiatives that support large-scale cleanup of end user machines. It has been six years since the Conficker botnet was sinkholed. The attackers have abandoned it. Still, nearly a million machines remain infected. Conficker provides us with a unique opportunity to estimate cleanup rates, because there are relatively few interfering factors at work. This paper is the first to propose a systematic approach to transform noisy sinkhole data into comparative infection metrics and normalized estimates of cleanup rates. We compare the growth, peak, and decay of Conficker across countries. We find that institutional differences, such as ICT development or unlicensed software use, explain much of the variance, while the national anti-botnet centers have had no visible impact. Cleanup seems even slower than the replacement of machines running Windows XP. In general, the infected users appear outside the reach of current remediation practices. Some ISPs may have judged the neutralized botnet an insufficient threat to merit remediation. These machines can however be magnets for other threats — we find an overlap between GameoverZeus and Conficker infections. We conclude by reflecting on what this means for the future of botnet mitigation.

1 Introduction

For years, researchers have been working on methods to take over or disrupt the command-and-control (C&C) infrastructure of botnets (e.g. [14, 37, 26]). Their successes have been answered by the attackers with ever more sophisticated C&C mechanisms that are increasingly resilient against takeover attempts [30].

In pale contrast to this wealth of work stands the limited research into the other side of botnet mitigation: cleanup of the infected machines of end users. After a botnet is successfully sinkholed, the bots or zombies basically remain waiting for the attackers to find a way to reconnect to them, update their binaries and move the machines out of the sinkhole. This happens with some regularity. The recent sinkholing attempt of GameoverZeus [32], for example, is more a tug of war between attackers and defenders, rather than definitive takedown action. The bots that remain after a takedown of C&C infrastructure may also attract other attackers, as these machines remain vulnerable and hence can be re-compromised.

To some extent, cleanup of bots is an automated process, driven by anti-virus software, software patches and tools like Microsoft's Malicious Software Removal Tool, which is included in Windows' automatic update cycle. These automated actions are deemed insufficient, however. In recent years, wide support has been established for the idea that Internet Service Providers (ISPs) should contact affected customers and help them remediate their compromised machines [39, 22]. This shift has been accompanied by proposals to treat large-scale infections as a public health issue [6, 8].

As part of this public health approach, we have seen the emergence of large-scale cleanup campaigns, most notably in the form of national anti-botnet initiatives. Public and private stakeholders, especially ISPs, collaborate to notify infected end users and help them clean their machines. Examples include Germany's Anti-Botnet Advisory Center (BotFrei), Australia's Internet Industry Code of Practice (iCode), and Japan's Cyber Clean Center (CCC, superseded by ACTIVE) [27].

Setting up large-scale cleanup mechanisms is cumbersome and costly. This underlines the need to measure whether these efforts are effective. The central question

of this paper is: What factors drive cleanup rates of infected machines? We explore whether the leading national anti-botnet initiatives have increased the speed of cleanup.

We answer this question via longitudinal data from the sinkhole of Conficker, one the largest botnets ever seen. Conficker provides us with a unique opportunity to study the impact of national initiatives. It has been six years since the vulnerability was patched and the botnet was sinkholed. The attackers have basically abandoned it years ago, which means that infection rates are driven by cleanup rather than the attacker countermeasures. Still, nearly a million machines remain infected (see figure 1). The Conficker Working Group, the collective industry effort against the botnet, concluded in 2010 that remediation has been a failure [7].

Before one can draw lessons from sinkhole data, or from most other data sources on infected machines, several methodological problems have to be overcome. This paper is the first to systematically work through these issues, transforming noisy sinkhole data into comparative infection metrics and normalized estimates of cleanup rates.

For this research, we were generously given access to the Conficker sinkhole logs, which provide a unique long term view into the life of the botnet. The dataset runs from February 2009 until September 2014, and covers all countries — 241 ISO codes — and 34,000 autonomous systems. It records millions of unique IP addresses each year — for instance, 223 million in 2009, and 120 million in 2013. For this paper, we focus on bots located in 62 countries.

In sum, the contributions of this paper are as follows:

1. We develop a systematic approach to transform noisy sinkhole data into comparative infection metrics and normalized estimates of cleanup rates.

2. We present the first long term study on botnet remediation.

3. We provide the first empirical test of the best practice exemplified by the leading national anti-botnet initiatives.

4. We identify several factors that influence cleanup rates across countries.

2 Background

2.1 Conficker timeline and variants

In this section we will provide a brief background on the history of the Conficker worm, its spreading and defense mechanisms, and some milestones in the activities of the Conficker Working Group.

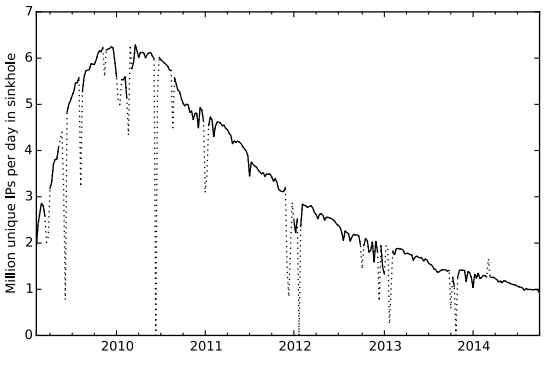


Figure 1: Conficker bots worldwide

The Conficker worm, also known as Downadup, was first detected in November 2008. The worm spread by exploiting vulnerability MS08-067 in Microsoft Windows, which had just been announced and patched. The vulnerability affected all versions of Microsoft Windows at the time, including server versions. A detailed technical analysis is available in [29]. Briefly put, infected machines scanned the IP space for vulnerable machines and infected them in a number steps. To be vulnerable, a machine needed to be unpatched and online with its NetBIOS ports open and not behind a firewall. Remarkably, a third of all machines had still not installed the patch by January 2009, a few months after its availability [11]. Consequently, the worm spread at an explosive rate. The malware authors released an update on December 29, 2008, which was named Conficker-B. The update added new methods of spreading, including via infected USB devices and shared network folders with weak passwords. This made the worm propagate even faster [7].

Infected machines communicated with the attackers via an innovative, centralized system. Every day, the bots attempted to connect to 250 new pseudo-randomly generated domains under eight different top-level domains. The attackers needed to register only one of these domains to reach the bots and update their instructions and binaries. Defenders, on the other hand, needed to block all these domains, every day, to disrupt the C&C. Another aspect of Conficker was the use of intelligent defense mechanisms, that made the worm harder to remove. It disabled Windows updates, popular anti-virus products, and several Windows security services. It also blocked access to popular security websites [29, 7].

Conficker continued to grow, causing alarm in the cybersecurity community about the potential scale of attacks, even though the botnet had not yet been very active at that point. In late January, the community — includ-

ing Microsoft, ICANN, domain registries, anti-virus vendors, and academic researchers — responded by forming the Conficker Working Group [7, 31]. The most important task of the working group was to coordinate and register or block all the domains the bots would use to communicate, staying ahead of the Conficker authors. The group was mostly successful in neutralizing the botnet and disconnecting it from its owners; however, small errors were made on two occasions in March, allowing the attackers to gain access to part of the botnet population and update them to the C variant.

The Conficker-C variant had two key new features: the number of pseudo-randomly generated domains was increased to 50,000 per day, distributed over a hundred different TLDs, and a P2P update protocol was added. These features complicated the work of the working group. On April 9, 2009, Conficker-C bots upgraded to a new variant that included a scareware program which sold fake anti-virus at prices between $50–$100. The fake anti-virus program, probably a pay-per-install contract, was purchased by close to a million unwitting users, as was later discovered. This use of the botnet prompted law enforcement agencies to increase their efforts to pursue the authors of Conficker.[1] Eventually, in 2011, the U.S. Federal Bureau of Investigation, in collaboration with police in several other countries, arrested several individuals associated with this $72-million scareware ring. [21, 19]

2.2 National anti-botnet centers

Despite the successes of the cybersecurity community in neutralizing Conficker, a large number of infected machines still remained. This painful fact was recognized early on; in its 'Lessons Learned' document from 2010, the Conficker Working Group reported remediation as its top failure [7]. Despite being inactive, Conficker remains one of the largest botnets. As recent as June 2014, it was listed as the #6 botnet in the world by anti-virus vendor ESET [9]. This underlines the idea that neutralizing the C&C infrastructure in combination with automated cleanup tools will not eradicate the infected machines; some organized form of cleanup is necessary.

During the past years, industry and regulatory guidelines have been calling for increased participation of ISPs in cleanup efforts. For instance, the European Network and Information Security Agency [1], the Internet Engineering Task Force [22], the Federal Communications Commission [10], and the Organization for Economic Cooperation and Development [27] have all called upon ISPs to contact infected customers and help them clean up their compromised machines.

[1]Microsoft also set a $250,000 bounty for information leading to arrests.

The main reason for this shift is that ISPs can identify and contact the owners of the infected machines, and provide direct support to end users. They can also quarantine machines that do not get cleaned up. Earlier work has found evidence that ISP mitigation can significantly impact end user security [40].

Along with this shift of responsibility towards ISPs, some countries have established national anti-botnet initiatives to support the ISPs and end users in cleanup efforts. The setup is different in each country, but typically it involves the collection of data on infected machines (from botnet sinkholes, honeypots, spamtraps, and other sources); notifying ISPs of infections within their networks; and providing support for end users, via a website and sometimes a call-center.

A number of countries have been running such centers, often as part of a public-private partnership. Table 1 lists the countries with active initiatives in late 2011, according to an OECD report [27]. The report also mentions the U.S. & U.K. as developing such initiatives. The Netherlands is listed as having 'ISP-specific' programs, for at that time, KPN and Ziggo — the two largest ISPs — were heading such programs voluntarily [39].[2] Finland, though not listed, has been a leader with consistently low infection rates for years. It has had a notification and cleanup mechanism in place since 2005, as part of a collaboration between the national CERT, the telco regulator and main ISPs [20, 25]. At the time of writing, other countries are starting anti-botnet centers as well. In the EU alone, seven new national centers have been announced [2]. These will obviously not impact the past cleanup rates of Conficker, but they do underwrite the importance of empirically testing the efficacy of this mitigation strategy.

Figure 2 shows the website of the German anti-botnet advisory center, *botfrei*. The center was launched in 2010 by eco, the German Internet industry association, and is partially funded by the German government. The center does three things. First, it identifies users with infected PCs. Second, they inform the infected customers via their ISPs. Third, they offer cleanup support, through a website — with free removal tools and a forum — and

[2]It has now been replaced by a wider initiative involving all main providers and covering the bulk of the broadband market.

COUNTRY	INITIATIVE
Australia	Internet Industry Code of Practice (iCode)
Germany	German Anti-Botnet Initiative (BotFrei)
Ireland	Irish Anti-Botnet Initiative
Japan	Cyber Clean Center / ACTIVE
Korea	KrCERT/CC Anti-Botnet Initiative
Netherlands	Dutch Anti-Botnet Initiative (Abuse-Hub)

Table 1: List of countries with anti-botnet initiatives [27]

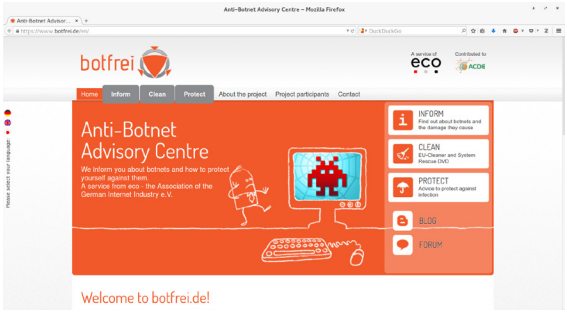


Figure 2: The German Anti-Botnet Advisory Center website - botfrei.de

a call center [17]. The center covers a wide range of malware, including Conficker. We should mention that eco staff told us that much of the German Conficker response took place before the center was launched. In their own evaluations, the center reports successes in terms of the number of users visiting its website, the number of cleanup actions performed, and overall reductions in malware rates in Germany. Interestingly enough, a large number of users visit botfrei.de directly, without being prompted by their ISP. This highlights the impact of media attention, as well as the demand for proactive steps among part of the user population.

We only highlight Germany's botfrei program as an example. In short, one would expect that countries running similar anti-botnet initiatives to have higher cleanup rates of Conficker bots. This, we shall evaluate.

2.3 Related Work

Similar to other botnets, much of the work on the Conficker worm has focused predominantly on technical analysis, e.g., [29]. Other research has studied the worm's outbreak and modeled its infection patterns, e.g., [42], [16], [33] and [41]. There have also been a few studies looking into the functioning of the Working Group, e.g., [31]. None of this work looks specifically at the issue of remediation. Although [33] uses the same dataset as this paper to model the spread of the worm, their results are skewed by the fact that they ignore DHCP churn, which is known to cause errors in infection rates of up to one order of magnitude for some countries [37].

This paper also connects to the literature on botnet mitigation, specifically to cleanup efforts. This includes the industry guidelines we discussed earlier, e.g., [1], [27], [10] and [22]; as well as academic work that tries to model different mitigation strategies, e.g., [6], [18] and [13]. We contribute to this discussion by bringing longitudinal data to bear on the problem and empirically evaluating one of the key proposals to emanate from this literature. This expands some of our earlier work.

In a broader context, a large body of research focuses on other forms of botnet mitigation, e.g., [14, 37, 26, 30], modeling worm infections, e.g. [35, 44, 43, 28], and challenges in longitudinal cybersecurity studies. For the sake of brevity we will not cite more works in these areas here (— except for works used in other sections).

3 Methodology

Answering the central research question requires a number of steps. First, we set out to derive reliable estimates of the number of Conficker bots in each country over time. This involves processing and cleaning the noisy sinkhole data, as well as handling several measurement issues. Later, we use the estimates to compare infection trends in various countries, identify patterns and specifically see if countries with anti-botnet initiatives have done any better. We do this by by fitting a descriptive model to each country's time-series of infection rates. This provides us with a specific set of parameters, namely the growth rate, the peak infection level, and the decay rate. We explore a few alternative models and opt for a two-piece model that accurately captures these characteristics. Lastly, to answer the central question, we explore the relationship between the estimated parameters and a set of explanatory variables.

3.1 The Conficker Dataset

The Conficker dataset has four characteristics that make it uniquely suited for studying large-scale cleanup efforts. First, it contains the complete record of one sinkholed botnet, making it less convoluted than for example spam data, and with far fewer false positives. Second, it logs most of the population on a daily basis, avoiding limitations from seeing only a sample of the botnet. Third, the dataset is longitudinal and tracks a period of almost six years. Many sinkholes used in scientific research typically cover weeks rather than months, let alone six years. Fourth, most infection data reflects a mix of attacker and defender behavior, as well as different levels (global & local). This makes it hard to determine what drives a trend – is it the result of attacker behavior, defender innovation, or just randomness? Conficker, however, was neutralized early on, with the attackers losing control and abandoning the botnet. Most other global defensive actions (e.g., patching and sinkholing) were also done in early 2009. Hence, the infection levels in our dataset predominantly reflect cleanup efforts. These combined attributes make the Conficker dataset excellent for studying the policy effects we are interested in.

Raw Data

Our raw data comes from the Conficker sinkhole logs. As explained in the background section, Conficker bots used an innovative centralized command and control infrastructure. The bots seek to connect to a number of pseudo-random domains every day, and ask for updated instructions or binaries from their masters. The algorithm that generates this domain list was reverse engineered early on, and various teams, including the Conficker Working Group, seized legal control of these domains. The domains were then 'sinkholed': servers were set up to listen and log every attempt to access the domains. The resulting logs include the IP address of each machine making such an attempt, timestamps, and a few other bits of information.

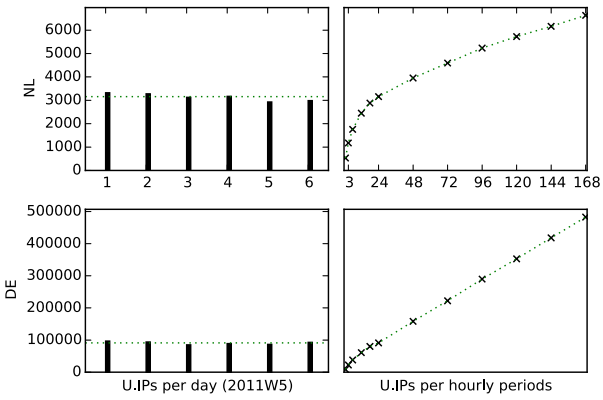


Figure 3: Unique IP counts over various time-periods

Processing Sinkhole Logs

The raw logs were originally stored in plain text, before adoption of the *nmsg* binary format in late 2010. The logs are huge; a typical hour of logs in January 2013 is around half a gigabyte, which adds up to tens of terabytes per year. From the raw logs we extract the IP address, which in the majority of cases will be a Conficker A, B, or C bot (the sinkholed domains were not typically used for other purposes). Then, using the MaxMind GeoIP database [23] and an IP-to-ASN database based on Routeviews BGP data [4], we determine the country and Autonomous System that this IP address belonged to at that moment in time. We lastly count the number of unique IP addresses in each region per hour.

With some exceptions, we capture most Conficker bots worldwide. The limitations are due to sinkholes downtime; logs for some sinkholed domains not being handed over to the working group [7]; and bots being behind an egress firewall, blocking their access to the sinkhole. None of these issues however creates a systematic bias, so we may treat them as noise.

After processing the logs we have a dataset spanning from February 2009 to September 2014, covering 241 ISO country codes and 34,000 autonomous systems. The dataset contains approximately 178 million unique IP addresses per year. In this paper we focus on bots located in 62 countries, which were selected as follows. We started with the 34 members of the Organization for Economic Cooperation and Development (OECD), and 7 additional members of the European Union which are not part of the OECD. These countries have a common development baseline, and good data is available on their policies, making comparison easier. We add to this list 23 countries that rank high in terms of Conficker or spam bots — cumulatively covering 80 percent of all such bots worldwide. These countries are interesting from a cybersecurity perspective. Finally, two countries were removed due to severe measurement issues affecting their bot counts, which we will describe later. The full list of countries can be seen in figure 8 or in the appendix.

3.2 Counting bots from IP addresses

The Conficker dataset suffers from a limitation that is common among most sinkhole data and other data on infected machines, such as spam traps, firewall logs, and passive DNS records: one has to use IP addresses as a proxy for infected machines. Earlier research has established that IP addresses are coarse unique identifiers and they can be off by one order of magnitude in a matter of days [37], because of differences in the dynamic IP address allocation policies of providers (so-called *DHCP churn*). Simply put, because of dynamic addresses, the same infected machine can appear in the logs under multiple IP addresses. The higher the churn rate, the more over-counting.

Figure 3 visualizes this problem. It shows the count of unique Conficker IP addresses in February 2011 over various time periods — 3 hours, 12 hours, one day, up to a week. We see an interesting growth curve, non-linear at the start, then linear. Not all computers are powered on at every point in time, so it makes sense to see more IP addresses in the sinkhole over longer time periods. However, between the 6th and 7th day, we have most likely seen most infected machines already. The new IP addresses are unlikely to be new infections, as the daily count is stable over the period. The difference is thus driven by infected machines reappearing with a new IP address.

The figure shows IP address counts for the Netherlands and Germany. From qualitative reports we know that IP churn is relatively low in the Netherlands — an Internet subscriber can retain the same IP address for months — while in Germany the address typically

changes every 24 hours. This is reflected in the figure: the slope for Germany is much steeper. Should one ignore the differences in churn rates among countries, and simply count unique IP addresses over a week, then a severe bias will be introduced against countries such as Germany. Using shorter time periods, though leading to under-counting, decreases this bias.[3] We settle for this simple solution: counting the average number of unique IPs *per hour*, thereby eliminating the churn factor. This hourly count will be a fraction of the total bot count, but that is not a problem when we make comparisons based on scale-invariant measures, such as cleanup rates.

Network Address Translation (NAT) and the use of HTTP proxies can also cause under-counting. This is particularly problematic if it happens at the ISP level, leading to large biases when comparing cleanup policies. After comparing subscriber numbers with IP address space size in our selection of countries, we concluded that ISP-level NAT is widely practiced in India. As we have no clear way of correcting such cases, we chose to exclude India from our analysis.

3.3 Missing measurements

The Conficker dataset has another problem that is also common: missing measurements. Looking back at figure 1, we see several sudden drops in bot counts, which we highlighted with dotted lines. These drops are primarily caused by sinkhole infrastructure downtime — typically for a few hours, but at one point even several weeks. These measurement errors are a serious issue, as they only occur in one direction and may skew our analysis. We considered several approaches to dealing with them. One approach is to model the measurement process explicitly. Another approach is to try and minimize the impact of aberrant observations by using robust curve-fitting methods. This approach adds unnecessary complexity and is not very intuitive. A third option is to pre-process the data using curve smoothing techniques; for instance by taking the exponentially weighted rolling average or applying the Hodrick-Prescott filter. Although not necessarily wrong, this also adds its own new biases as it changes data. The fourth approach, and the one that we use, is to detect and remove the outliers heuristically.

For this purpose, we calculate the distance between each weekly value in the global graph with the rolling median of its surrounding two months, and throw out the top 10%. This works because most bots log in about once a day, so the IP counts of adjacent periods are not independent. The IP count may increase, decrease, or slightly fluctuate, but a sudden decrease in infected machines followed by a sudden return of infections to the previous level is highly unlikely. The interested reader is referred to the appendix to see the individual graphs for all the countries with the outliers removed.[4]

[3] Ideally, we would calculate a churn rate — the average number of IPs per bot per day — and use that to generate a good estimate of the actual number of bots. That is not an easy task, and requires making quite a number of assumptions.

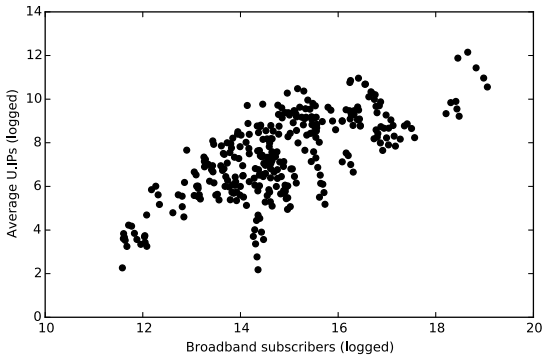


Figure 4: Conficker bots versus broadband subscribers

3.4 Normalizing bot counts by country size

Countries with more Internet users are likely to have more Conficker bots, regardless of remediation efforts. Figure 4 illustrates this. It thus makes sense to normalize the unique IP counts by a measure of country size; in particular if one is to compare peak infection rates. One such measure is the size of a country's IP space, but IP address usage practices vary considerably between countries. A more appropriate denominator and the one we use is the number of Internet broadband subscribers. This is available from a number of sources, including the Worldbank Development Indicators.

4 Modeling Infections

4.1 Descriptive Analysis

Figure 5 shows the Conficker infection trends for Germany, United States, France, and Russia. The x-axis is time; the y-axis is the average number of unique IP addresses seen per day in the sinkhole logs, corrected for churn. We observe a similar pattern: a period of rapid growth; a plateau period, where the number of infected machines peaks and remains somewhat stable for a short or longer amount of time; and finally, a period of gradual decline.

What explains these similar trends among countries, and in particular, the points in time where the changes

[4] An extreme case was Malaysia, where the length of the drops and fluctuations spanned several months. This most likely indicates country-level egress filtering, prompting us to also exclude Malaysia from the analysis.

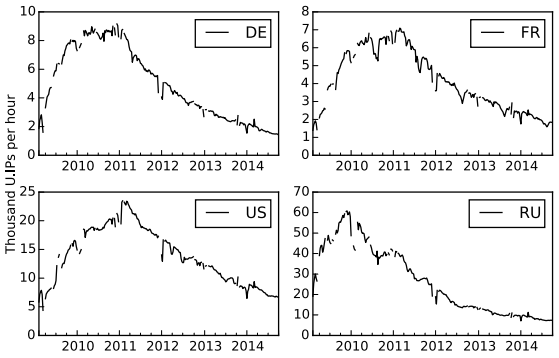


Figure 5: Conficker trends for four countries

occur on the graphs? At first glance, one might think that the decline is set off by some event — for instance, the arrest of the bot-masters, or a release of a patch. But this is not the case. As previously explained, all patches for Conficker were released by early 2009, while the worm continued spreading after that. This is because most computers that get infected with Conficker are "unprotected" — that is, they are either unpatched or without security software, in case the worm spreads via weak passwords on networks shares, USB drives, or domain controllers. The peak in 2010 – 2011 is thus the worm reaching some form of saturation where all vulnerable computers are infected. In the case of business networks, administrators may have finally gotten the worm's reinfection mechanisms under control [24].

Like the growth phase and the peak, the decline can also not be directly explained by external attacker behavior. Arrests related to Conficker occurred mid 2011, while the decline started earlier. In addition, most of the botnet was already out of the control of the attackers. What we are seeing appears to be a 'natural' process of the botnet. Infections may have spread faster in some countries, and cleanups may have been faster in others, but the overall patterns are similar across all countries.

4.2 Epidemic Models

It is often proposed in the security literature to model malware infections similarly as epidemics of infectious diseases, e.g. [28, 44]. The analog is that vulnerable hosts get infected, and start infecting other hosts in their vicinity; at some later point they are recovered or removed (cleaned, patched, upgraded or replaced).

This leads to multiple phases, similar to what we see for Conficker: in the beginning, each new infection increases the pressure on vulnerable hosts, leading to an explosive growth. Over time, fewer and fewer vulnerable hosts remain to be infected. This leads to a phase where the force of new infections and the force of recovery are locked in dynamic equilibrium. The size of the infected population reaches a plateau. In the final phase, the force of recovery takes over, and slowly the number of infections declines towards zero.

Early on in our modeling efforts we experimented with a number of epidemic models, but eventually decided against them. Epidemic models involve a set of latent compartments and a set of differential equations that govern the transitions between them — see [12] for an extensive overview. Most models make a number of assumptions about the underlying structure of the population and the propagation mechanism of the disease.

The basic models for instance assume constant transition rates over time. Such assumptions might hold to an acceptable degree in short time spans, but not over six years. The early works applying these models to the Code Red and Slammer worms [44, 43] used data spanning just a few weeks. One can still use the models even when the assumptions are not met, but the parameters cannot be then easily interpreted. To illustrate: the basic Kermack-McKendrick SIR model fits our data to a reasonable degree. However, we know that this model assumes no reinfections, while Conficker reinfections were a major problem for some companies [24].

More complex models reduce assumptions by adding additional latent variables. This creates a new problem: often when solved numerically, different combinations of the parameters fit the data equally well. We observed this for some countries with even the basic SIR model. Such estimates are not a problem when the aim is to predict an outbreak. But they are showstoppers when the aim is to compare and interpret the parameters and make inferences about policies.

4.3 Our model

For the outlined reasons, we opted for a simple descriptive model. The model follows the characteristic trend of infection rates, provides just enough flexibility to capture the differences between countries, and makes no assumptions about the underlying behavior of Conficker. It merely describes the observed trends in a small set of parameters.

The model consists of two parts: a logistic growth that ends in a plateau; followed by an exponential decay. Logistic growth is a basic model of self-limiting population growth, where first the rate of growth is proportional to the size of the existing population, and then declines as the natural limit is approached (— the seminal work of Staniford, et al. [35] also used logistic growth). In our case, this natural limit is the number of vulnerable hosts.

Exponential decay corresponds to a daily decrease of the number of Conficker bots by a fixed percentage. Figure 6 shows the number of infections per subscriber over

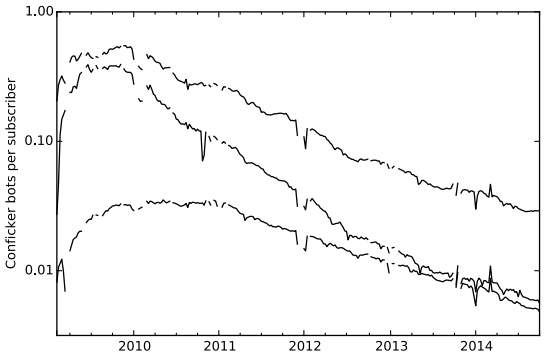


Figure 6: Conficker bots per subscriber on logarithm scale for (from top to bottom) Russia, Belarus, Germany.

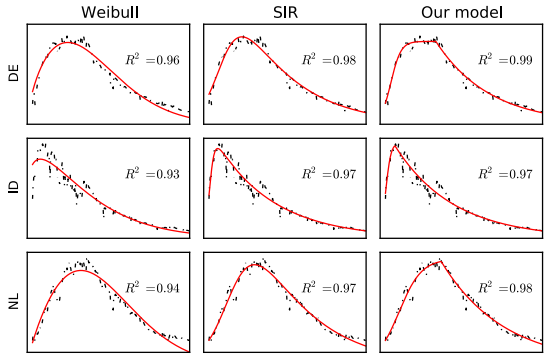


Figure 7: Comparison of alternative models

time for three countries on a logarithm scale. We see a downward-sloping straight line in the last phase that corresponds to an exponential decay: the botnet shrank by a more or less a constant percentage each day. We do not claim that the assumptions underpinning the logistic growth and the exponential decay models are fully satisfied, but in the absence of knowledge of the exact dynamics, their simplicity seems the most reasonable approach.

The model allows us to reduce the time series data for each country to these parameters: (1) the infection rate in the growth phase, (2) the peak number of infections, (3) the time at which this peak occurred, and (4) the exponential decay rate in the declining phase. We will fit our model on the time series for all countries, and then compare the estimates of these parameters.

Mathematically, our model is formulated as follows:

$$\text{bots}(t) = \begin{cases} \dfrac{K}{1+e^{-r(t-t_0)}}, & \text{if } t < t_P \\ \\ He^{-\gamma(t-t_P)}, & \text{if } t \geq t_P \end{cases} \quad (1)$$

where $\text{bots}(t)$ is the number of bots at time t, t_P is the time of the peak (where the logistic growth transitions to exponential decay), and H the height of the peak. The logistic growth phase has growth rate r, asymptote K, and midpoint t_0. The parameter γ is the exponential decay rate. The height of the peak is identified by the other parameters:

$$H = \frac{K}{1+e^{-r(t_P-t_0)}}.$$

4.4 Inspection of Model Fit

We fit the curves using the Levenberg-Marquardt least squares algorithm with the aid of the *lmfit* Python module. The results are point estimates; standard errors were computed by *lmfit* by approximating the Hessian matrix at the point estimates. With these standard errors we computed Wald-type confidence intervals (point estimate $\pm$ 2 s.e.) for all parameters. These intervals have no exact interpretation in this case, but provide some idea of the precision of the point estimates.

The reader can find plots of the fitted curves for all 62 countries in the appendix. The fits are good, with R^2 values all between 0.95 and 1. Our model is especially effective for countries with sharp peaks, that is, the abrupt transitions from growth to decay that can be seen in Hungary and South Africa, for example. For some countries, such as Pakistan and Ukraine, we have very little data on the growth phase, as they reached their peak infection rate around the time sinkholing started. For these countries we will ignore the growth estimates in further analysis. By virtue of our two-phase model, the estimates of the decay rates are unaffected by this issue.

We note that our model is deterministic rather than stochastic; that is, it does not account for one-time shocks in cleanup that lead to a lasting drop in infection rates. Nevertheless, we see that the data follows the fitted exponential decay curves quite closely, which indicates that bots get cleaned up at a constant rate and non-simultaneously.[5]

Alternative models: We tried fitting models from epidemiology (e.g. the SIR model) and reliability engineering (e.g. the Weibull curve), but they did not do well in such cases, and adjusted R^2 values were lower for almost all countries. Additionally, for a number of countries, the parameter estimates were unstable. Figure 7 illustrates why: our model's distinct phases captures the height of peak and exponential decay more accurately.

[5]The exception is China: near the end of 2010 we see a massive drop in Conficker infections. After some investigation, we found clues that this drop might be associated by a sudden spur in the adoption of IPv6 addresses, which are not directly observable to the sinkhole.

5 Findings

5.1 Country Parameter Estimates

Figure 8 shows the parameter estimates and their precision for each of the 62 countries: the growth rate, peak height, time of the peak, and the decay rate.

The variance in the peak number of infections is striking: between as little as 0.01% to over 1% of Internet broadband subscribers. The median is .1%. It appears that countries with high peaks tend to also have high growth rates, though we have to keep in mind that the growth rate estimates are less precise, because the data does not fully cover that phase. Looking at the peak height, it seems that this is not associated with low cleanup rates. For example, Belarus (BY) has the highest decay rate, but a peak height well above the median.

The timing of the peaks is distributed around the last weeks of 2010. Countries with earlier peaks are mostly countries with higher growth rates. This suggests that the time of the peak is simply a matter of when Conficker ran out of vulnerable machines to infect; a faster growth means this happens sooner. Hence, it seems unlikely that early peaks indicate successful remediation.

The median decay rate estimate is .009, which corresponds to a 37% decline per year ($100 \cdot (1 - e^{-.009 \cdot 52})$). In countries with low decay rates (around .005), the botnet shrank by 23% per year, versus over 50% per year on the high end.

5.2 National Anti-Botnet Initiatives

We are now in a position to address the paper's central question and to explore the effects of the leading national anti-botnet initiatives (ABIs). In figure 8, we have highlighted the countries with such initiatives as crosses. One would expect that these countries have slower botnet growth, a lower peak height, and especially a faster cleanup rate. There is no clear evidence for any of this; the countries with ABIs are all over the place. We do see some clustering on the lower end of the peak height graphs; however, this position is shared with a number of other countries that are institutionally similar (in terms of wealth for example) but not running such initiatives.

We can formally test if the population median is equal for the two groups using the Wilcoxon ranksum test. The p-value of the test when comparing the Conficker decay rate among the two sets of countries is 0.54, which is too large to conclude that the ABIs had a meaningful effect. It is somewhat surprising, and disappointing, to see no evidence for the impact of the leading remediation efforts on bot cleanup.

We briefly look at three possible explanations. The first one is that country trends might be driven by infections in other networks than those of the ISPs, as we know that the ABIs focus mostly on ISPs. This explanation fails, however, as can be seen in figure 2. The majority of the Conficker bots were located in the networks of the retail ISPs in these countries, compared to educational, corporate or governmental networks. This pattern held in 2010, the year of peak infections, and 2013, the decay phase, with one minor deviation: in the Netherlands, cleanup in ISP networks was faster than in other networks.

Country	ISP % 2010	ISP % 2013
AU	77%	74%
DE	89%	82%
FI	73%	69%
IE	72%	74%
JP	64%	67%
KR	83%	87%
NL	72%	**37%**
Others	81%	75%

Table 2: Conficker bots located in retail ISPs

A second explanation might be that the ABIs did not include Conficker in their notification and cleanup efforts. In two countries, Germany and the Netherlands, we were able to contact participants of the ABI. They claimed that Conficker sinkhole feeds were included and sent to the ISPs. Perhaps the ISPs did not act on the data — or at least not at a scale that would impact the decay rate; they might have judged Conficker infections to be of low risk, since the botnet had been neutralized. This explanation might be correct, but it also reinforces our earlier conclusion that the ABIs did not have a significant impact. After all, this explanation implies that the ABIs have failed to get the ISPs and their customers to undertake cleanup at a larger scale.

Given that cleanup incurs cost for the ISP, one could understand that they might decide to ignore sinkholed and neutralized botnets. On closer inspection, this decision seems misguided, however. If a machine is infected with Conficker, it means it is in a vulnerable — and perhaps infected — state for other malware as well. Since we had access to the global logs of the sinkhole for GameoverZeus — a more recent and serious threat — we ran a cross comparison of the two botnet populations. We found that based on common IP addresses, a surprising 15% of all GameoverZeus bots are also infected with Conficker. During six weeks at the end of 2014, the GameoverZeus sinkhole saw close to 1.9 million unique IP addresses; the Conficker sinkhole saw 12 million unique IP addresses; around 284 thousand addresses appear in both lists. Given that both malware types only infected a small percentage of the total pop-

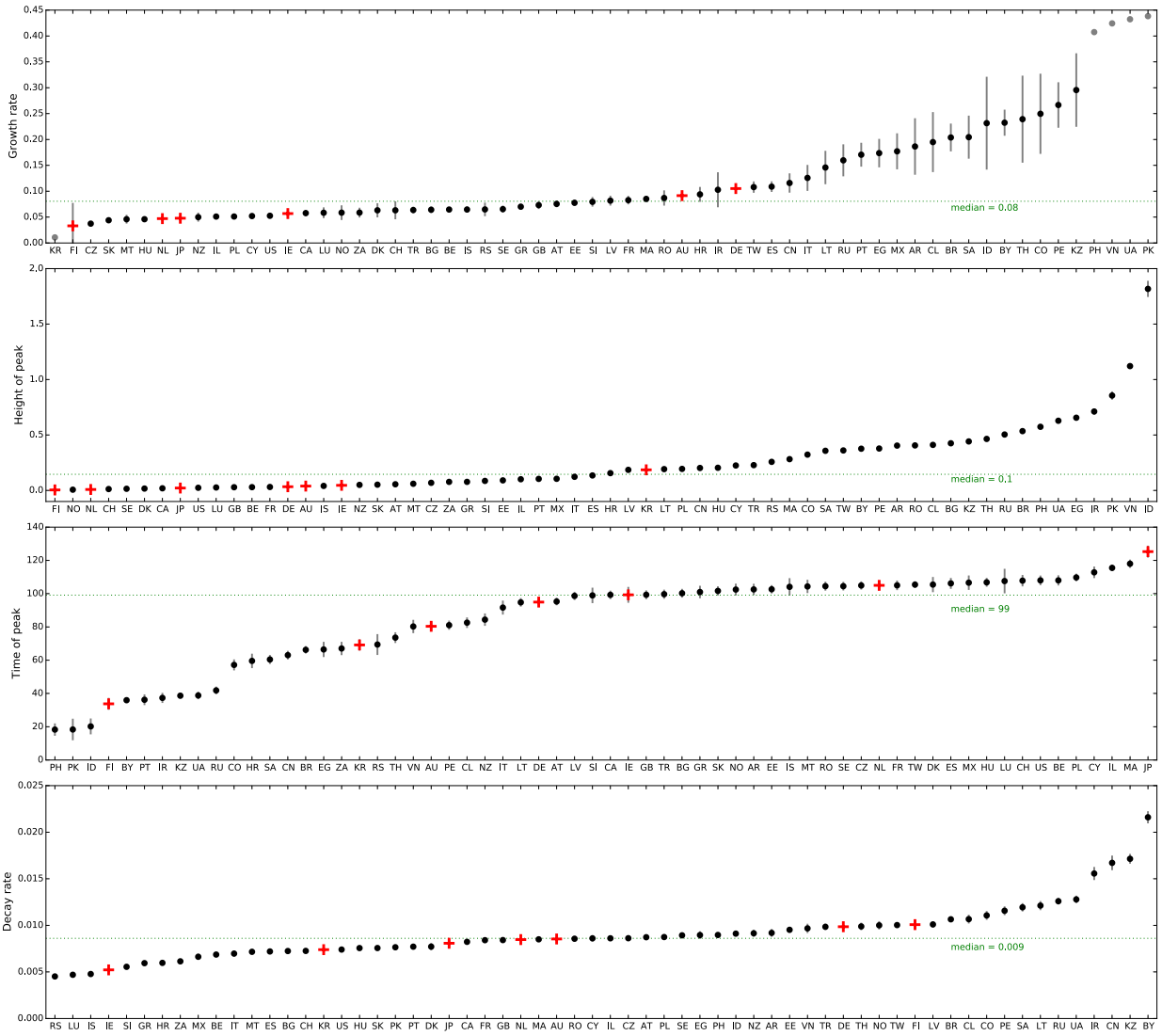


Figure 8: Parameter estimates and confidence intervals

ulation of broadband subscribers, this overlap is surprisingly large.[6] It stands in stark contrast to the findings of a recent study that systematically determined the overlap among 85 blacklists and found that most entries were unique to one list, and that overlap between independent lists was typically less than one percent [34]. In other words, Conficker bots should be considered worthwhile targets for cleanup.

[6]The calculated overlap in terms of bots might be inflated as a result of both NAT and DHCP churn. Churn can in this case have both an over-counting and under-counting effect. Under-counting will occur if one bot appears in the two sinkholes with different IP addresses, as a result of different connection times to the sinkholes. Doing the IP comparisons at a daily level yields a 6% overlap, which is still considerable.

5.3 Institutional Factors

Given that anti-botnet initiatives cannot explain the variation among the country parameters shown in figure 8, we turn our attention to several institutional factors that are often attributed with malware infection rates (e.g., see [40]). These are broadband access, unlicensed software use, and ICT development on a national level. In addition, given the spreading mechanism of Conficker, we also look at Operating System market shares, as well as PC upgrade cycles. We correlate these factors with the relevant parameters.

Correlating Growth Rate

Broadband access is often mentioned as a technological enabler of malware; in particular, since Conficker was a worm that spread initially by scanning for hosts to infect, one could expect its growth in countries with higher broadband speeds to be faster. Holding other factors constant, most epidemiological models would also predict this faster growth with increased network speeds. This turns out not to be the case. The Spearman correlation coefficient between average national broadband speeds, as reported by the International Telecommunication Union [15], and Conficker growth rate is in fact negative: -0.30. This is most probably due to other factors confounding with higher broadband speeds, e.g. national wealth. In any case, the effects of broadband access and speeds are negligible compared to other factors, and we will not pursue this further.

Correlating Height of Peak

As we saw, there is a wide dispersion between countries in the peak number of Conficker bots. What explains the large differences in peak infection rates?

Operating system market shares: Since Conficker only infects machines running Windows 2000, XP, Vista, or Server 2003/2008, some variation in peak height may be explained by differences in use of these operating systems (versus Windows 7 or non-Windows systems). We use data from StatCounter Global Stats [36], which is based on page view analytics of some three million websites. Figure 9 shows the peak height against the combined Windows XP and Vista market shares in January 2010 (other vulnerable OS versions were negligible). We see a strong correlation — with a Pearson correlation coefficient of 0.55. This in itself is perhaps not surprising.

Dividing the peak heights by the XP/Vista market shares gives us estimates of the *peak number of infections per vulnerable user*; we shall call this metric $\widetilde{\text{hp}}$. This metric allows for fairer comparisons between countries, as one would expect countries with higher market shares of vulnerable OS's to harbor more infections regardless of other factors. Interestingly, there is still considerable variation in this metric – the coefficient of variance is 1.2. We investigate two institutional factors that may explain this variation.

ICT development index is an index published by the ITU based on a number of well-established ICT indicators. It allows for benchmarking and measuring the digital divide and ICT development among countries (based on ICT readiness and infrastructure, ICT intensity and use, ICT skills and literacy [15]). This is obviously a broad indicator, and can indicate the ability to manage cybersecurity risks, including botnet cleanups, among

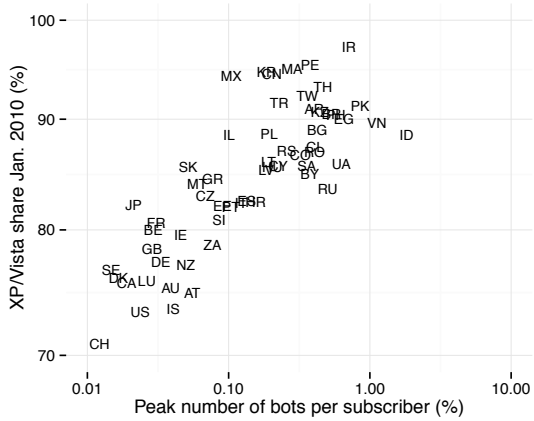


Figure 9: Bots versus XP & Vista use

both citizens and firms. Figure 10 shows this metric against $\widetilde{\text{hp}}$, and interestingly enough we see a strong correlation.

Unlicensed software use or piracy rates are another oft mentioned factor influencing malware infection rates. In addition to the fact that pirated software might include malware itself, users running pirated OS's often turn off automatic updates, for fear of updates disabling their unlicensed software — even though Microsoft consistently states that it will also ship security updates to unlicensed versions of Windows [38]. Disabling automatic updates leaves a machine open to vulnerabilities, and stops automated cleanups. We use the unlicensed software rates calculated by the Business Software Alliance [5]. This factor also turns out to be strongly correlated with $\widetilde{\text{hp}}$, as shown in figure 10.

Since ICT development and piracy rates are themselves correlated, we use the following simple linear regression to explore thier joint association with peak Conficker infection rates:

$$\log(\widetilde{\text{hp}}) = \alpha + \beta_1 \cdot \text{ict-dev} + \beta_2 \cdot \text{piracy} + \varepsilon$$

where both regressors were standardized by subtracting the mean and dividing by two standard devia-

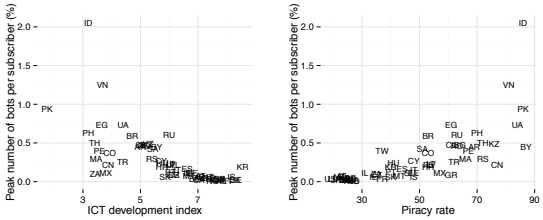


Figure 10: $\widetilde{\text{hp}}$ versus ICT development & piracy

tions. We use the logarithm of $\widetilde{\text{hp}}$ as it is a proportion. The least squares estimates (standard errors) are $\hat{\beta}_1 = -0.78(0.27)$, $p < 0.01$, and $\hat{\beta}_2 = 1.7(0.27)$, $p < 0.001$. These coefficients can be interpreted as follows: everything else kept equal, countries with low (one sd below the mean) ICT development have $e^{0.78} = 2.2$ times more Conficker bots per XP/Vista user at the peak than countries with high ICT development (one sd above the mean), and, similarly, countries with high piracy rates (one sd above the mean) have an $e^{1.7} = 5.5$ times higher peak than countries with low piracy rates (one sd below the mean). The R^2 of this regression is 0.78, which indicates that ICT development and piracy rates explain most of the variation in Conficker peak height.

Correlating Decay Rate

Although decay rates are less dispersed than peak heights, there are still noticeable differences among countries. Given the rather slow cleanup rates — the median of 0.009 translates to a 37% decrease in the number of bots after one year — one hypothesis that comes to mind is that perhaps some of the cleanup is being driven by users upgrading their OS's (to say Windows 7), or buying a new computer and disposing of the old fully.

For each country we estimated the **decay rate of the market share of Windows XP and Vista** from January 2011 to June 2013 using the StatCounter GlobalStats data. Figure 11 shows these decay rates versus Conficker decay rates. There is a weak correlation among the two, with a Spearman correlation coefficient of 0.26.

But more interesting and somewhat surprising is that in many countries, the Conficker botnet shrank at a slower pace than the market share of Windows XP / Vista (all countries below and to the right of the dashed line). Basically this means that the users infected with Conficker are less likely to upgrade their computers then the average consumer.[7]

6 Discussion

We found that the large scale national anti-botnet initiatives had no observable impact on the growth, peak height, or decay of the Conficker botnet. This is surprising and unfortunate, as one would expect Conficker bots to be among those targeted for cleanup by such initiatives. We checked that the majority of bots were indeed located among the networks of ISPs, and also observed that some of these machines have multiple infections. Turning away from the initiatives and to institutional factors that could explain the differences among

[7]This difference between users who remain infected with Conficker and the average user might be more extreme in countries with a higher level of ICT development. This can be observed in the graph.

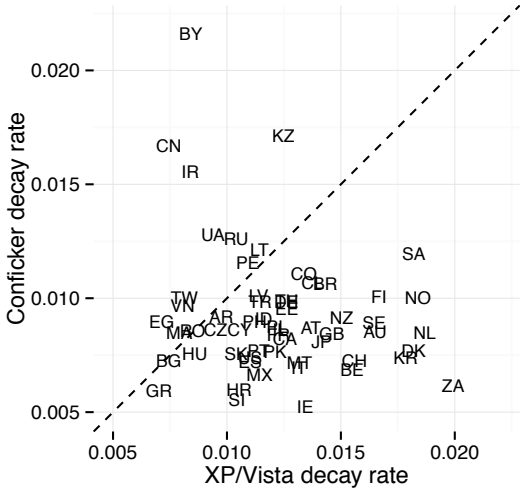


Figure 11: Conficker decay vs. XP/Vista decay

countries, we observed that the ICT development index and piracy rates can explain 78% of the variation in peak height, even after correcting for OS market shares. We also found that the Conficker cleanup rate is less than the average PC upgrade rate.

Perhaps not all security experts are surprised by these findings. They are nevertheless important in forming effective anti-botnet policies. We presented the research to an audience of industry practitioners active in botnet cleanup. Two North American ISPs commented that they informed their customers about Conficker infections — as part of the ISP's own policy, not a country-level initiative. They stated that some customers repeatedly ignored notifications, while others got re-infected soon after cleanup. Another difficulty was licensing issues requiring ISPs to point users to a variety of cleanup tool websites (e.g., on microsoft.com) instead of directly distributing a tool, which complicates the process for some users. Interestingly enough both ISPs ranked well with regards to Conficker peak, showing that their efforts did have a positive impact. Their challenges suggests areas for improvement.

Future work in this area can be taken in several directions. One is to test the various parameters against other independent variables — e.g., the number of CERTs, privacy regulation, and other governance indicators. A second avenue is to explore Conficker infection rates at the ISP level versus the country level. A random effects regression could reveal to what extent ISPs in the same country follow similar patterns. We might see whether particular ISPs differ widely from their country baseline, which would be interesting from an anti-botnet perspective. Third, it might be fruitful to contact a number of

users still infected with Conficker in a qualitative survey, to see why they remain unaware or unworried about running infected machines. This can help develop new mitigation strategies for the most vulnerable part of the population. Perhaps some infections are forgotten embedded systems, not end users. Forth and more broadly is to conduct research on the challenges identified by the ISPs: notification mechanisms and simplifying cleanup.

7 Conclusion and Policy Implications

In this paper, we tackled the often ignored side of botnet mitigation: large-scale cleanup efforts. We explored the impact of the emerging best practice of setting up national anti-botnet initiatives with ISPs. Did these initiatives accelerate cleanup?

The longitudinal data from the Conficker botnet provided us with a unique opportunity to explore this question. We proposed a systematic approach to transform noisy sinkhole data into comparative infection metrics and normalized estimates of cleanup rates. After removing outliers, and by using the hourly Conficker IP address count per subscriber to compensate for a variety of known measurement issues, we modeled the infection trends using a two-part model. We thereby condensed the dataset to three key parameters for each country, and compared the growth, peak, and decay of Conficker, which we compared across countries.

The main findings were that institutional factors such as ICT development and unlicensed software use have influenced the spread and cleanup of Conficker more than the leading large scale anti-botnet initiatives. Cleanup seems even slower than the replacement of machines running Windows XP, and thus infected users appear outside the reach of remediation practices. At first glance, these findings seem rather gloomy. The Conficker Working Group, a collective effort against botnets, had noted remediation to be their largest failure [7]. We have now found that the most promising emerging practice to overcome that failure suffers similar problems.

So what can be done? Our findings lead us to identify several implications. First of all, the fact that peak infection levels strongly correlate with ICT development and software piracy, suggests that botnet mitigation can go hand in hand with economic development and capacity building. Helping countries develop their ICT capabilities can lower the global impact of infections over the long run. In addition, the strong correlation with software piracy suggests that automatic updates and unattended cleanups are some of the strongest tools in our arsenal. It support policies to enable security updates to install by default, and delivering them to all machines, including those running unlicensed copies [3]. Some of these points were also echoed by the ISPs mentioned in section 6.

Second, the fact that long-living bots appear in a reliable dataset — that is, one with few false positives — suggests that future anti-botnet initiatives need to commit ISPs to take action on such data sources, even if the sinkholed botnet is no longer a direct threat. These machines are vulnerable and likely to harbor other threats as well. Tracking these infections will be an important way to measure ISP compliance with these commitments, as well as incentivize cleanup for those users outside the reach of automated cleanup tools.

Third, given that cleanup is a long term effort, future anti-botnet initiatives should support, and perhaps fund, the long-term sustainability of sinkholes. This is a necessity if we want ISPs to act on this data.

A long term view is rare in the area of cybersecurity, which tends to focus on the most recent advances and threats. In contrast to C&C takedown, bot remediation needs the mindset of a marathon runner, not a sprinter. To conclude on a more optimistic note, Finland has been in the marathon for a longer time than basically all other countries. It pays off: they have been enjoying consistently low infection rates for years now. In other words, a long term view is not only needed, but possible.

Acknowledgment

The authors would like to explicitly thank Chris Lee, Paul Vixie and Eric Ziegast for providing us with access to the Conficker sinkhole and supporting our research.

We also thank Ning An, Ben Edwards, Dina Hadziosmanovic, Stephanie Forest, Jan Philip Koenders, Rene Mahieu, Hooshang Motarjem, Piet van Mieghem, Julie Ryan, as well as the participants of Microsoft DCC 2015 and USENIX reviewers for contributing ideas and providing feedback at various stages of this paper.

References

[1] Botnets: Measurement, detection, disinfection and defence.

[2] ADVANCED CYBER DEFENCE CENTRE. Support centers - advanced cyber defence centre (ACDC).

[3] ANDERSON, R., BHME, R., CLAYTON, R., AND MOORE, T. Security economics and the internal market. 00068.

[4] ASGHARI, H. Python IP address to autonomous system number lookup module.

[5] BUSINESS SOFTWARE ALLIANCE. BSA global software survey: The compliance gap: Home. 00000.

[6] CLAYTON, R. Might governments clean-up malware? 87–104.

[7] CONFICKER WORKING GROUP. Conficker working group: Lessons learned.

[8] EASTWEST INSTITUTE. The internet health model for cybersecurity. 00000.

[9] ESET. Global threat report - june 2014.

[10] FEDERAL COMMUNICATIONS COMISSION. U.s. anti-bot code of conduct (ABCs) for internet service providers (ISPs).

[11] GOODIN, D. Superworm seizes 9m PCs, 'stunned' researchers say.

[12] HEESTERBEEK, J. Mathematical epidemiology of infectious diseases: model building, analysis and interpretation. 02020.

[13] HOFMEYR, S., MOORE, T., FORREST, S., EDWARDS, B., AND STELLE, G. *Modeling internet-scale policies for cleaning up malware*. Springer, pp. 149–170.

[14] HOLZ, T., STEINER, M., DAHL, F., BIERSACK, E., AND FREILING, F. C. Measurements and mitigation of peer-to-peer-based botnets: A case study on storm worm. 1–9. 00375.

[15] INTERNATIONAL TELECOMMUNICATIONS UNION. Measuring the information society. 00002.

[16] IRWIN, B. A network telescope perspective of the conficker outbreak. In *Information Security for South Africa (ISSA), 2012*, IEEE, pp. 1–8.

[17] KARGE, S. The german anti-botnet initiative.

[18] KHATTAK, S., RAMAY, N. R., KHAN, K. R., SYED, A. A., AND KHAYAM, S. A. A taxonomy of botnet behavior, detection, and defense. 898–924.

[19] KIRK, J. Ukraine helps disrupt $72m conficker hacking ring.

[20] KOIVUNEN, E. *Why Wasnt I Notified?: Information Security Incident Reporting Demystified*, vol. 7127. Springer Berlin Heidelberg, pp. 55–70. 00000.

[21] KREBS, B. 72m USD scareware ring used conficker worm.

[22] LIVINGOOD, J., MODY, N., AND O'REIRDAN, M. Recommendations for the remediation of bots in ISP networks.

[23] MAXMIND. https://www.maxmind.com/en/geoip2-precision-country.

[24] MICROSOFT. Microsoft security intelligence report - how conficker continues to propogate.

[25] MICROSOFT. TelieSonera, european telecom uses microsoft security data to remove botnet devices from network.

[26] NADJI, Y., ANTONAKAKIS, M., PERDISCI, R., DAGON, D., AND LEE, W. Beheading hydras: performing effective botnet takedowns. ACM Press, pp. 121–132.

[27] OECD. Proactive policy measures by internet service providers against botnets.

[28] PASTOR-SATORRAS, R., CASTELLANO, C., VAN MIEGHEM, P., AND VESPIGNANI, A. Epidemic processes in complex networks. 00019.

[29] PORRAS, P., SAIDI, H., AND YEGNESWARAN, V. An analysis of confickers logic and rendezvous points.

[30] ROSSOW, C., ANDRIESSE, D., WERNER, T., STONE-GROSS, B., PLOHMANN, D., DIETRICH, C., AND BOS, H. SoK: P2pwned - modeling and evaluating the resilience of peer-to-peer botnets. In *2013 IEEE Symposium on Security and Privacy (SP)*, pp. 97–111. 00035.

[31] SCHMIDT, A. Secrecy versus openness: Internet security and the limits of open source and peer production.

[32] SHADOWSERVER FOUNDATION. Gameover zeus.

[33] SHIN, S., GU, G., REDDY, N., AND LEE, C. P. A large-scale empirical study of conficker. 676–690.

[34] SPRING, J. Blacklist ecosystem analysis. 00000.

[35] STANIFORD, S., PAXSON, V., WEAVER, N., AND OTHERS. How to own the internet in your spare time. In *USENIX Security Symposium*, pp. 149–167.

[36] STATCOUNTER. Free invisible web tracker, hit counter and web stats. 00000.

[37] STONE-GROSS, B., COVA, M., CAVALLARO, L., GILBERT, B., SZYDLOWSKI, M., KEMMERER, R., KRUEGEL, C., AND VIGNA, G. Your botnet is my botnet: Analysis of a botnet takeover. In *Proceedings of the 16th ACM Conference on Computer and Communications Security*, CCS '09, ACM, pp. 635–647.

[38] TOM'S HARDWARE. Microsoft: Pirated windows 7 will still get updates. 00000.

[39] VAN EETEN, M. J., ASGHARI, H., BAUER, J. M., AND TABATABAIE, S. Internet service providers and botnet mitigation: A fact-finding study on the dutch market.

[40] VAN EETEN, M. J., BAUER, J. M., ASGHARI, H., TABATABAIE, S., AND RAND, D. The role of internet service providers in botnet mitigation: An empirical analysis based on spam data.

[41] WEAVER, R. A probabilistic population study of the conficker-c botnet. In *Passive and Active Measurement*, Springer, pp. 181–190.

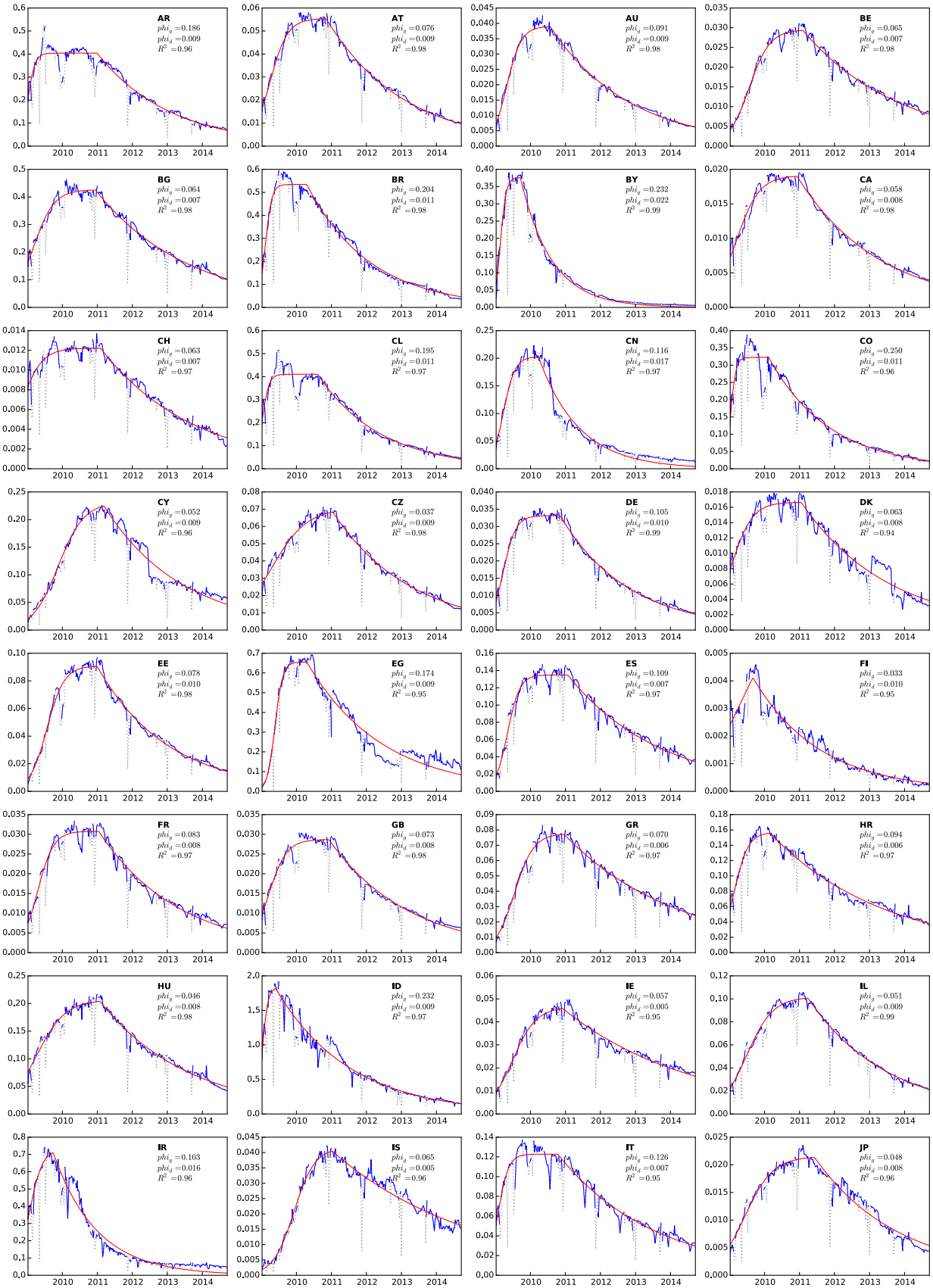
[42] ZHANG, C., ZHOU, S., AND CHAIN, B. M. Hybrid spreading of the internet worm conficker.

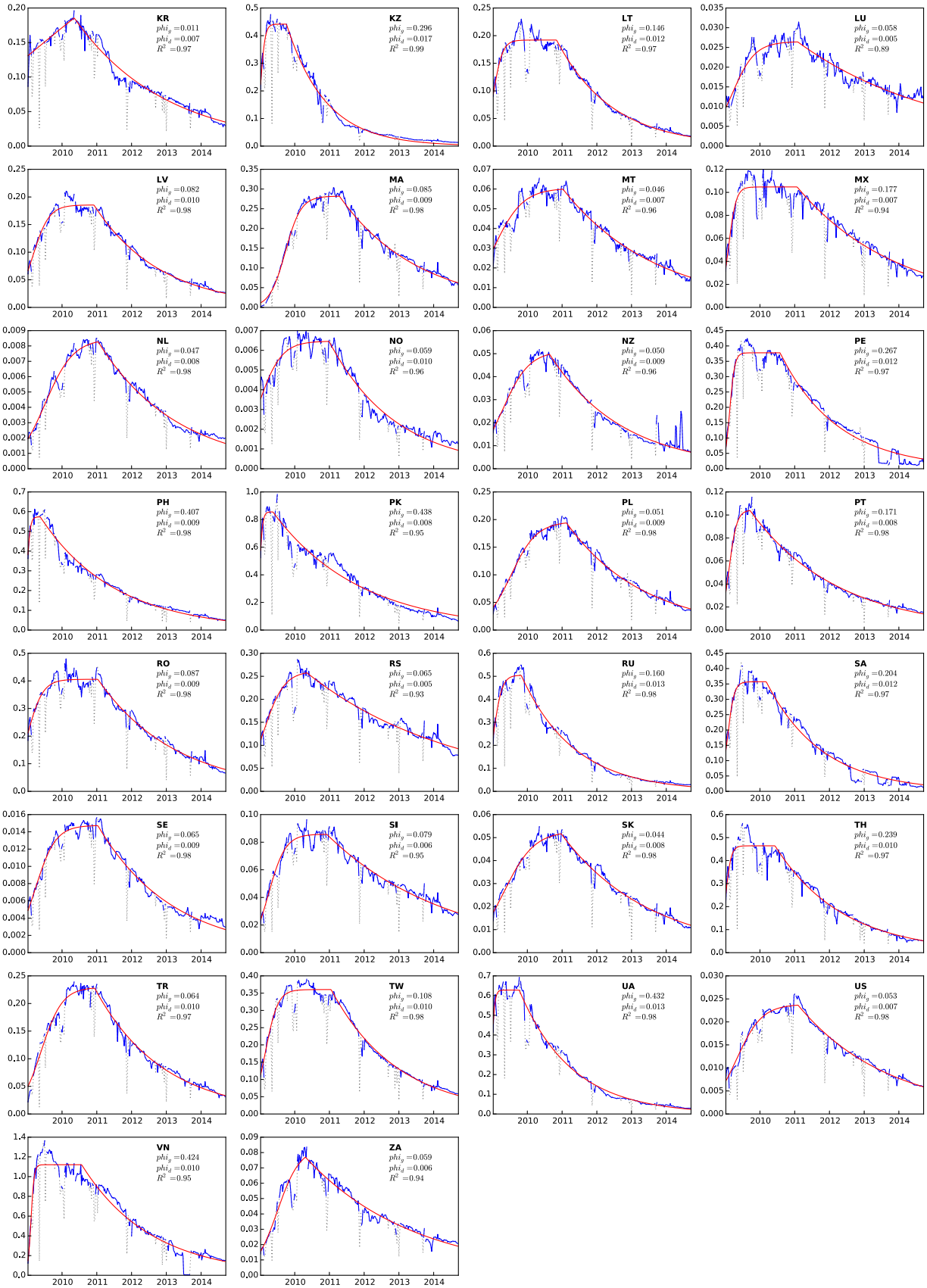
[43] ZOU, C. C., GAO, L., GONG, W., AND TOWSLEY, D. Monitoring and early warning for internet worms. In *Proceedings of the 10th ACM conference on Computer and communications security*, ACM, pp. 190–199.

[44] ZOU, C. C., GONG, W., AND TOWSLEY, D. Code red worm propagation modeling and analysis. In *Proceedings of the 9th ACM conference on Computer and communications security*, ACM, pp. 138–147.

Appendix - Individual Country Graphs

In this appendix we provide the model fit for all the 62 countries used in the paper. The graphs show the relative number of Conficker bots in each country - as measured by average unique Conficker IP addresses per hour in the sinkholes, divided by broadband subscriber counts for each country. (Please refer to the methodology section for the rationale). In each graph, the solid line (in blue) indicates the measurement; the dotted line (in gray) is removed outliers; and the smooth-solid line (in red) indicates the fitted model. The model has four parameters: growth and decay rates — given on the graph — and height and time of peak infections — deducible from the axes. The R^2 is also given for each country.


AR phi_g =0.186 phi_d =0.009 R^2 =0.96
AT phi_g =0.076 phi_d =0.009 R^2 =0.98
AU phi_g =0.091 phi_d =0.009 R^2 =0.98
BE phi_g =0.065 phi_d =0.007 R^2 =0.98
BG phi_g =0.064 phi_d =0.007 R^2 =0.98
BR phi_g =0.204 phi_d =0.011 R^2 =0.98
BY phi_g =0.232 phi_d =0.022 R^2 =0.99
CA phi_g =0.058 phi_d =0.008 R^2 =0.98
CH phi_g =0.063 phi_d =0.007 R^2 =0.97
CL phi_g =0.195 phi_d =0.011 R^2 =0.97
CN phi_g =0.116 phi_d =0.017 R^2 =0.97
CO phi_g =0.250 phi_d =0.011 R^2 =0.96
CY phi_g =0.052 phi_d =0.009 R^2 =0.96
CZ phi_g =0.037 phi_d =0.009 R^2 =0.98
DE phi_g =0.105 phi_d =0.010 R^2 =0.99
DK phi_g =0.063 phi_d =0.008 R^2 =0.94
EE phi_g =0.078 phi_d =0.010 R^2 =0.98
EG phi_g =0.174 phi_d =0.009 R^2 =0.95
ES phi_g =0.109 phi_d =0.007 R^2 =0.97
FI phi_g =0.033 phi_d =0.010 R^2 =0.95
FR phi_g =0.083 phi_d =0.008 R^2 =0.97
GB phi_g =0.073 phi_d =0.008 R^2 =0.98
GR phi_g =0.070 phi_d =0.006 R^2 =0.97
HR phi_g =0.094 phi_d =0.006 R^2 =0.97
HU phi_g =0.046 phi_d =0.008 R^2 =0.98
ID phi_g =0.232 phi_d =0.009 R^2 =0.97
IE phi_g =0.057 phi_d =0.005 R^2 =0.95
IL phi_g =0.051 phi_d =0.009 R^2 =0.99
IR phi_g =0.103 phi_d =0.016 R^2 =0.96
IS phi_g =0.065 phi_d =0.005 R^2 =0.96
IT phi_g =0.126 phi_d =0.007 R^2 =0.95
JP phi_g =0.048 phi_d =0.008 R^2 =0.96


KR
phi_g =0.011
phi_d =0.007
R^2 =0.97
KZ
phi_g =0.296
phi_d =0.017
R^2 =0.99
LT
phi_g =0.146
phi_d =0.012
R^2 =0.97
LU
phi_g =0.058
phi_d =0.005
R^2 =0.89
LV
phi_g =0.082
phi_d =0.010
R^2 =0.98
MA
phi_g =0.085
phi_d =0.009
R^2 =0.98
MT
phi_g =0.046
phi_d =0.007
R^2 =0.96
MX
phi_g =0.177
phi_d =0.007
R^2 =0.94
NL
phi_g =0.047
phi_d =0.008
R^2 =0.98
NO
phi_g =0.059
phi_d =0.010
R^2 =0.96
NZ
phi_g =0.050
phi_d =0.009
R^2 =0.96
PE
phi_g =0.267
phi_d =0.012
R^2 =0.97
PH
phi_g =0.407
phi_d =0.009
R^2 =0.98
PK
phi_g =0.438
phi_d =0.008
R^2 =0.95
PL
phi_g =0.051
phi_d =0.009
R^2 =0.98
PT
phi_g =0.171
phi_d =0.008
R^2 =0.98
RO
phi_g =0.087
phi_d =0.009
R^2 =0.98
RS
phi_g =0.065
phi_d =0.005
R^2 =0.93
RU
phi_g =0.160
phi_d =0.013
R^2 =0.98
SA
phi_g =0.204
phi_d =0.012
R^2 =0.97
SE
phi_g =0.065
phi_d =0.009
R^2 =0.98
SI
phi_g =0.079
phi_d =0.006
R^2 =0.95
SK
phi_g =0.044
phi_d =0.008
R^2 =0.98
TH
phi_g =0.239
phi_d =0.010
R^2 =0.97
TR
phi_g =0.064
phi_d =0.010
R^2 =0.97
TW
phi_g =0.108
phi_d =0.010
R^2 =0.98
UA
phi_g =0.432
phi_d =0.013
R^2 =0.98
US
phi_g =0.053
phi_d =0.007
R^2 =0.98
VN
phi_g =0.424
phi_d =0.010
R^2 =0.95
ZA
phi_g =0.059
phi_d =0.006
R^2 =0.94