A CIP-PSP funded pilot action
Grant agreement n°325188

Deliverable	**D4.2 Statistical evaluation of the impact of the Pilot**
Work package	WP4 Evaluating and Incentivizing Botnet Mitigation
Due date	31/07/2015
Submission date	31/07/2015
Revision	Final v1.1
Status of revision	
Responsible partner	Michel van Eeten (TUD)
Contributors	Qasim Lone (TUD) Hadi Asghari (TUD) Giovane Moura (TUD) Jan Kohlrausch (DFN-CERT) Christian Keil (DFN-CERT) Jonathan Chapman (reviewer, FKIE)
Project Number	CIP-ICT PSP-2012-6 / 325188
Project Acronym	ACDC
Project Title	Advanced Cyber Defence Centre
Start Date of Project	01/02/2013

Dissemination Level	
PU: Public	X
PP: Restricted to other programme participants (including the Commission)	
RE: Restricted to a group specified by the consortium (including the Commission)	
CO: Confidential, only for members of the consortium (including the Commission)	

Version history

Rev.	Date	Author	Notes
01	11/07/2015	Michel van Eeten, Qasim Lone, Hadi Asghari (TUD)	First partial draft
1.0	27/07/2015	Michel van Eeten, Qasim Lone, Hadi Asghari (TUD), Jan Kohlrausch, Christian Keil (DFN-CERT)	First complete draft
1.2	29/07/2015	Michel van Eeten, Qasim Lone, Hadi Asghari (TUD), Jan Kohlrausch, Christian Keil (DFN-CERT), Jonathan Chapman (reviewer, FKIE)	Final report

Glossary

ACDC	Advanced Cyber Defence Centre
AS	Autonomous System
CCH	Centralized Clearing House
CP	Community Platform
ISP	Internet Service Provider
NSC	National Support Center

Table of contents

Table of figures

Table of tables

1 Executive summary

In this report, we evaluate ACDC – more precisely, we ask to what extent the end-to-end approach has been realized and has been effective. To answer this question, Chapter 2 conceptualizes the end-to-end approach as a flow of inputs to outputs to outcomes. We focus the evaluation of ACDC on a core output (data quality) and outcome (impact on infection levels).

In Chapter 3 we evaluate the quality of the data (output) that ACDC is contributing to the fight against botnets. Does it enable better mitigation? The short answer is: yes. The volume of events being submitted each day is consistently around 500,000. This is a non-trivial amount, in the same range as major botnet sinkholes run by, for example, ShadowServer. Furthermore, the data that the ACDC sensors and experiments submit to the CCH seems to be relatively unique, with little overlap with the data feeds that Internet Service Providers and CERTS are receiving from organizations like ShadowServer and Spamhaus. This means he CCH can make an important contribution to enhanced mitigation by ISPs, CERTS and others.

Chapter 4 outlines a measurement methodology to look at the trends in infection levels across 60 countries and 265 Internet Service Providers. This allows us to undertake a qualitative evaluation of the changes in the infection levels in the countries that have set up national support centers.

In Chapter 5, we present rankings and time trends for the countries with ACDC National Support Centers (NSCs), compared to a population of 54 other countries. The countries with an ACDC NSC perform very different, as do the ISPs in those countries. They are similarly distributed as the 54 other countries in our analysis, as well as the ISPs in those countries. Some have infection rates that are substantially better (i.e., lower) than average. Others are average or worse than average.

We explored whether the setup of an NSC is part of an ongoing process of improvement in mitigation over time. In other words, we look at whether the countries with an NSC did cleanup faster than other countries. It turned out that they did not. The countries and ISPs that did well maintained their position. The ones who struggled with higher infection rates, continue to struggle with them. There does not seem to be a relationship with the presence of an NSC, nor with whether the NSC is operationally involved in mitigation.

None of this is to say that there is no improvement in cleanup as such. In most of our data sources, the number of infections are going down. This is partially related to the type of data sources we use (mostly sinkholes), and partially actual ongoing cleanup.

It is too early to declare a substantial verdict on the impact associated with a NSC. That being said, from a policy perspective, we highlight two implications. First is the issue of scale. According to Microsoft, and our own analysis, around one percent of all Windows PCs are infected at any moment in time. This number is two orders of magnitude larger than the number of customer notifications that the German and Spanish NSCs have done. Scaling up to the actual infection level seems necessary to have a substantial impact.

The second policy-relevant finding is that we found less variance among ISPs in the same country than we expected. In several countries, we saw surprisingly similar infection levels. This seems to suggest that ISPs' mitigation policies are less idiosyncratic and are in fact guided by what their peers and public authorities signal to them. This provides opportunities for the future of NSCs and the end-to-end approach of ACDC beyond the conclusion of the project.

2 Introduction

2.1 *Relation to other ACDC outputs*

ACDC provides an end-to-end approach to botnet mitigation, from detection to protection, in the form of services that are operated by partners in different countries: a Centralized Clearing House (CCH), support centers, detection and mitigation, and integrated tools for the removal of malware from the end-user devices.

To support the operational model, ACDC has developed the infrastructure described in D2.3 (see Figure 1). The infrastructure supports varying stakeholders in their work against botnets: providers and network operators, law enforcement agencies and CERTs, citizens, industry and researchers.

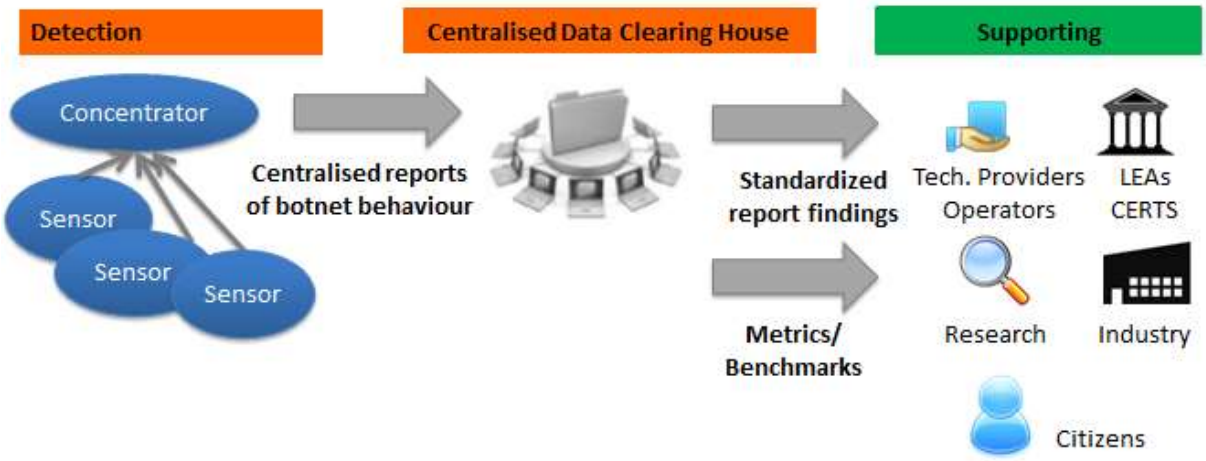

Figure 1: ACDC project anti-botnet operational model

Information about infected users and other botnet-related events is collected via sensors and stored in the CCH. From there, it is disseminated in basically two forms: (1) standardized reports that support mitigation activities by CERTs, providers and network operators; (2) metrics and benchmarks that provide information on infection levels, effectiveness of cleanup efforts and other issues that help all stakeholders evaluate their practices, policies and the situation in their countries.

The second form in which the data of the CCH is disseminated is at the heart of Work Package 4. WP4's core objective is to leverage the CCH for evaluating and incentivizing botnet mitigation. This is achieved by conducting quality control and processing tasks on the data in the CCH and then enriching this data in the form of comparative botnet metrics that can be used by a variety of stakeholders, via the Community Portal.

Overall, WP4 consists of four tasks:

- Task 4.1 Data processing and quality control
- Task 4.2 Design and production of comparative botnet metrics
- Task 4.3 Evaluating the impact of the Pilot
- Task 4.4 Evaluating the compliance with legal requirements of the Pilot
- Task 4.5 Publishing benchmarks to incentivize market actors

This deliverable (D4.2) presents the results of task 4.3: using the metrics that have been developed and published in D4.1 for evaluating the impact of the Pilot – that is, of the ACDC project.

2.2 *Evaluating ACDC*

How can a large and complex project like ACDC be meaningfully evaluated? The relatively straightforward concept of 'botnet mitigation' obscures the complexity of effective mitigation at a pan-European scale. There are a lot of 'moving parts' in such an undertaking; activities that are distributed across many different stakeholders and that need to be lined up for mitigation to be effective. ACDC has lined up these parts in an end-to-end approach, from the detection of infected resources with a variety of sensors run by different partners, to the aggregation at scale in the centralized clearinghouse, to the dissemination of this data to the right entities, to the actual cleanup of the affected resources.

The overarching question for the evaluation of ACDC is to what extent this end-to-end approach has been realized and has been effective. To answer this question, it helps to conceptualize the end-to-end approach as a flow of inputs to outputs to outcomes (see Figure 2).

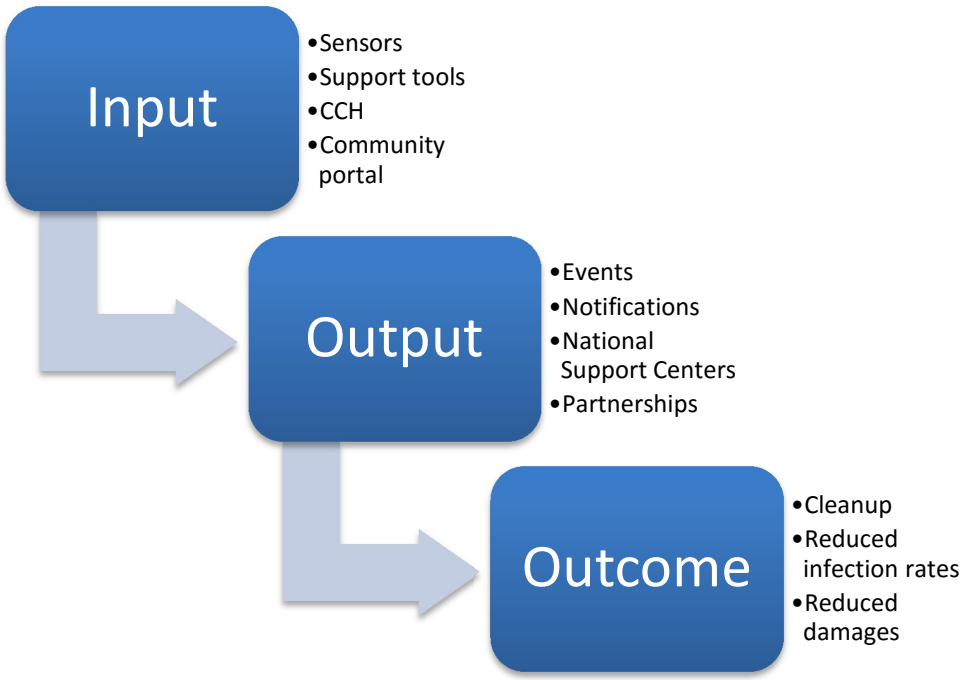

Figure 2: Evaluating ACDC

The input to the end-to-end process is the infrastructure that is needed to enable it: the sensors, support tools, CCH, and the community portal. The outputs are the direct results of this infrastructure: the botnet events that have been observed, the notifications that have been sent to CERTS and network operators, the National Support Centers (NSCs) that have been launched to support enhanced mitigation and the partnerships that have been established around the infrastructure.

Outputs are not goals in themselves, however. They are produced in order to achieve certain outcomes: the degree in which the intended impact has been realized. Did the intervention actually impact the problem that needed to be solved? For ACDC, the ultimate test of the effectiveness of any anti-botnet solution is in reducing botnet infection rates in real-world networks.

This original objective of this evaluation was mostly focused on outcomes. In the Description of Work, it was stated as follows:

> *Once comparative metrics are available, it becomes possible to evaluate the impact of the Pilot, the experiments as well as other mitigation strategies. So far, all claims about best practices and actual mitigation initiatives are based on anecdotal evidence, at best. Using a variety of statistical techniques, we can identify the impact of the Pilot.*

We know from decades of evaluation research that evaluating projects on the basis of outcomes is a harsh test under the best of circumstances, as there are so many factors are at play that can undermine the impact of the intervention. In the context of ACDC, this is even more difficult, due to the short time frame within which the project had to realize its objectives. Outcomes typically become visible over years, rather than months. Furthermore, there have been delays in the course of the project that made it even harder to evaluate ACDC based on its outcomes. We discuss these issues in the next section. In short, we cannot evaluate ACDC purely on outcomes.

In light of all of this, we can now scope the objectives of evaluation. It will have to be more modest than originally envisioned. The evaluation will include outcomes (metrics on infection rates in countries where a new national support center was set up) as well as outputs (events and notifications) that are a precursor for further mitigation in the future. In the next section, we articulate the objectives of the evaluation more precisely.

2.3 *Objectives of this document*

The original aim of this deliverable was to present the evaluation of the impact of ACDC on botnet mitigation. This overall objective still stands, but its scope had to be revised, however, because of developments elsewhere in the project.

The initial idea for D4.2 was that mitigation based on data from the CCH would get underway in the course of the project. The comparative metrics developed in D4.1 would then enable us to track and evaluate the impact of this mitigation. We would focus especially on the metrics that capture the relative infection rates of networks. The idea was to determine that providers who would use CCH data for mitigation would improve faster than other providers. We were planning to deploy a variety of statistical techniques to identify whether the networks of partners that used the CCH data for mitigation moved to lower infection rates compared to the networks of non-partners.

In the course of the ACDC project, delays have occurred, as has been documented extensively elsewhere and has been discussed with reviewers as well. A variety of factors, most notably the legal complications around data sharing, have led to delays in other work packages. This, in turn, has delayed the flow of infection data into the CCH and, thus, of notifications out of the CCH to support the mitigation actions of relevant partners.

Recently, progress has been made. Mitigation is currently underway. However, the requirements for the statistical analysis of the impact that we originally envisioned cannot be met with the data at hand. More particularly:

- Mitigation has only come into force in the past few months. This means that the time frame to evaluate the impact has become too short for any statistically analysis to detect a deviation of the autonomous trends in the infection metrics;
- Only a limited set of partners are currently using the CCH data for mitigation. This group is not large enough to draw meaningful statistical conclusions on the impact of the mitigation efforts supported by ACDC;
- Last, the volume of events and notifications that are being disseminated by the CCH is modest compared to other sources of such event data, such as ShadowServer. This means

that any reduction in infection levels would be difficult to associate with ACDC data being used for mitigation.

For these reasons, we have developed a qualitative approach to the objective of D4.2. First, we focus on assessing the quality of data collected and disseminated via the CCH (outputs). Second, we undertake an qualitative evaluation of the infection rates in the countries that have set up a National Support Center based on data external to ACDC, namely a set of infection metrics that have been developed by TU Delft over the past years. We will generate country and ISP rankings and see if the rankings of the countries which have set up a NSC improves over time. This approach cannot isolate the effect of the NSCs specifically, but rather sees them as part of a broader mitigation effort in those countries. These efforts were often already ongoing before the official launch of the NSC. This means, we can look at the impact of the NSCs as part of a mitigation strategy over a longer period of time.

In short, we have broken down the overall objective into two more feasible objectives to evaluate ACDC in its current stage:

- Evaluate the quality of data coming into the clearinghouse and disseminated for mitigation. This evaluation uses the following types of metrics, as defined in D4.4: data quality metrics; tool-based metrics; operational metrics.
- Qualitatively evaluate the changes in the infection levels in the countries that have set up national support centers. This part of the evaluation is based on external data from TU Delft and its partners. It uses this data to generate the botnet impact metrics, as defined in D4.1. and referenced in D4.4.

2.4 *Structure of the document*

After this introductory chapter, the document is structured into four main chapters:

- Chapter 3 evaluates the quality of ACDC data based on a variety of metrics that have been defined and are currently being generated on an ongoing basis. It assesses the degree in which this data can meaningfully improve mitigation practices.
- Chapter 4 lays the groundwork for evaluating the impact of ACDC support centers. It outlines in detail the methodology that was developed to generate the infection metrics based on the external data.
- Chapter 5 evaluates the impact of the national support centers based on external data by comparing at the country and ISP level the countries with a new national center to other countries.
- Chapter 6 brings the main findings together and discusses their implications for ACDC.

3 Evaluating the Quality of Data for Mitigation

The goal of this chapter is to evaluate the quality of data that can be provided by the CCH. As explained in the previous chapter, this evaluates the output of the ACDC solution. We explore and describe the CCH data, provide descriptive statistics, and finally assess some quality features.

3.1 *CCH quality workflows*

Figure 3 shows the data flows and systems where the data is processed. The data generated by the experiments is submitted via HTTPS to the CCH which stores the data for further processing. The "Metrics VM" is a dedicated virtual machine that retrieves the data from the CCH and computes the metrics. The results are sent back to the CCH via the "research" workflow. Moreover, the VM provides data for the ACDC community portal where the results are visualised. The quality control is split into two parts:

1. [**Quality Control in the CCH**]: This part covers all steps that require raw data (e.g. IP addresses) and that do not require complex computational resources. The quality control includes checks to ensure the conformity of the data to the data formats specified.
2. [**Statistical Quality Control**]: This part is based on metrics that are being part of the research workflow computed on an external dedicated system (Metrics VM). Since the mathematical models are too complex to be computed in the CCH, the quality control is implemented on external systems. Their aim is to forecast the retrieved data for gaps and anomalies.

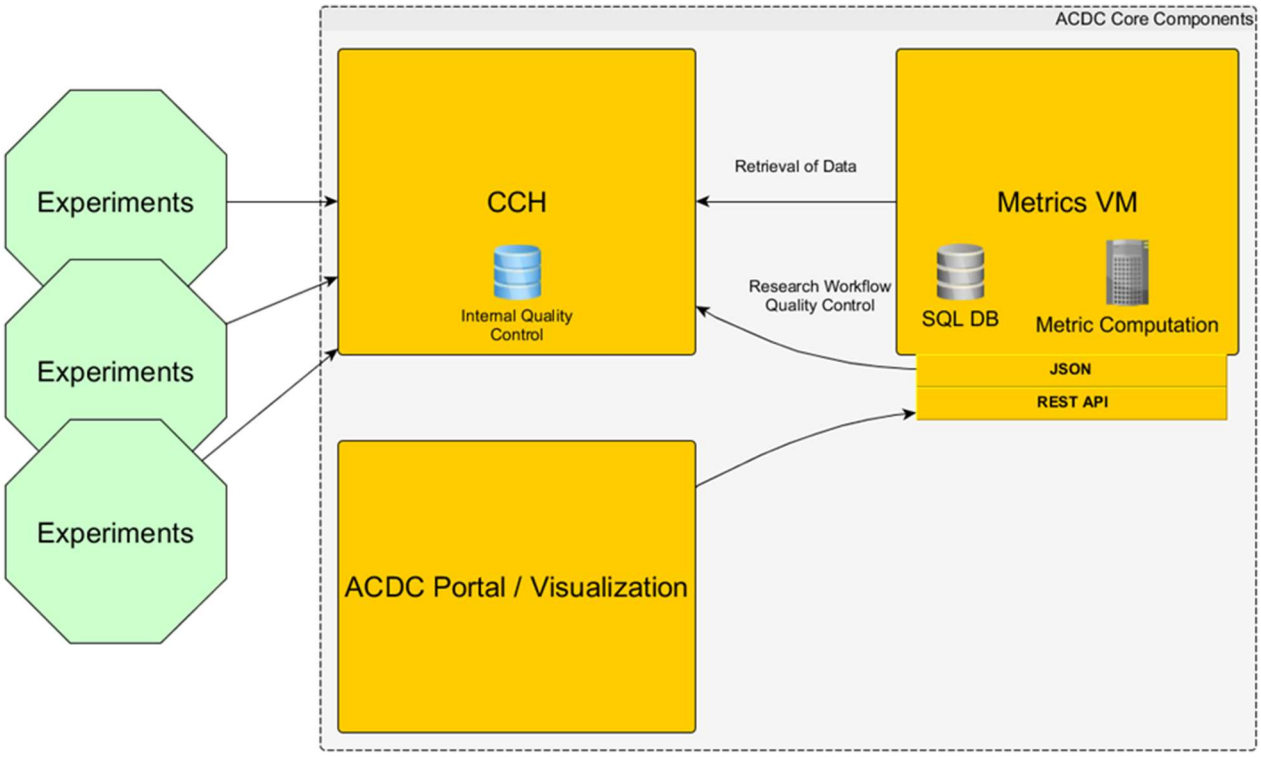

Figure 3: Data flows

The CCH receives the data via HTTPS requests from the participating sites, adds metadata (which includes, for example, the ASN and a DNS lookup of the IP address), optionally anonymizes or prseudonymizes it (to address data protection issues), and relays the data to the metrics VM. There, the data is stored in a database for a specific time frame (per default the data expires after one week). The metrics are computed on data retrieved from that database. The metrics VM offers two different ways to access the results. The results can be queried via a REST API, for instance for visualisation in the community portal. Moreover, the VM submits the results to the CCH where they are provided to the research workflow and further shared with some participants.

3.2 *CCH quality metrics*

Our evaluation is based on a number of metrics that were generated about the CCH data. In total, we received 17 metrics in a JSON schema, the details of which are outlined in D4.4. The term metrics is used rather broadly, with various forms, as we shall see. An example of the JSON schema with the metrics is given below.


```
{
"aggregation_type": "metric",
"measurement_window": 86400,
"metric_description": "Unique IPs per Second Level Domain",
"metric_id": 7,
"metric_result": {"anonhost.DE": {"attack": { “412": { "data": 1}}},
                  "anonhost.NET": {"attack": {"412": { "data": 1}}}},
"report_category": "eu.acdc.aggregated",
"report_subcategory": "ip_based_metric",
"report_type": "[METRIC][RDNS_METRICS][DFN-CERT] Unique IPs per Second Level Domain",
"timestamp": "2015-04-07",
"version": 1
}
```


According to the ACDC specifications, the design and specification of comparative botnet metrics is conducted in the context of Task 4.2 (“Design and production of comparative botnet metrics”) and as a result of this work, the following groups of metrics have been identified as suitable for implementation and benchmarking:

- **Data quality metrics**: assess the quality of the data submitted to the CCH in terms of gaps and anomalies (& geo-distribution), taking into account that it could distort the statistical stability of the data.
- **Botnet impact metrics**: assess the activity and distribution of botnets by comparing incidents related to Bots (IP-based, Proxy-based, RDNS-based) per ASN, per Country, per ISP-subscriber - helping assessing botnet presence in the world along time.
- **Tool-based metrics**: this category of metrics permits assessing technologies (i.e. network sensors, malware analysis tools, etc.) that provide data to the CCH in terms of quality and distribution.
- **Operational metrics:** this category of metrics permit evaluating the volume and quality of data reported for pilot experiments (like those conducted in WP3), or certain cybersecurity events (such as DDoS attacks, Malicious URIs detected or Malware samples analyzed).

The D4.4 report enumerates a large number of metrics possible that might be produced in each category. A subset those were selected to be produced. These metrics would be generated by the CCH every day. We compare the list of possible metrics and the metrics that were implemented in Table 1.

Table 1: Comparing possible and implemented metrics

Category	Listed metrics	Possible metrics	Implemented metrics
Data quality metrics	5	- Data sources per Submission Key - Data Distribution per Submission Key - Quality – Notification phase: Reports per ASN - Quality – false positives per partner - Quality - Reliable Data sources per Submission Key	#23 [QUALITY_METRICS][DFN-CERT] Subcategories per Key #24 [QUALITY_METRICS][DFN-CERT] Subcategories (unique IPs) per Key #21 [QUALITY_METRICS][DFN-CERT] Reliable Reports per Key and Category #11 [QUALITY_METRICS][ATOS] Reliable Reports per ASN #15 [QUALITY_METRICS][DFN-CERT] Reports per Key and Category
Botnet Impact Metrics	10	- Comparative - Daily BotIDs / country user - Comparative - Daily BotIDs / ASN-IP - Comparative - Daily BotIDs / ISP - IP-based - Unique daily IPs per country user - IP -based - Unique daily IPs per ASN-IP - IP -based - Unique daily IPs per ISP-Subscriber - Proxy-based - Daily events per country user - Proxy-based - Daily events per ASN-IP - Proxy-based - Daily events per ISP Subscriber - RDNS-based - Unique daily IPs with the same second level domain per day	#1 [IP_METRICS][TUD] Unique IPs per ASN #2 [IP_METRICS][TUD] Unique IPs per Country #3 [IP_METRICS][TUD] Unique IPs per Partner #4 [PROXY_METRICS][TUD] Unique Events per ASN #5 [PROXY_METRICS][TUD] Unique Events per Country #6 [PROXY_METRICS][TUD] Unique Events per Partner #7 [RDNS_METRICS][TUD] Unique IPs per Second Level Domain #22 [RDNS_METRICS][TUD] Unique events per Second Level Domain
Tool-based metrics	12	- Average Confidence Level - Volume of reports per ASN - Volume of reports per Country	#8 [QUALITY_METRICS][ATOS] Average Confidence Level per Key #9 [QUALITY_METRICS][ATOS] Number of Reports per ASN and Key #10 [QUALITY_METRICS][ATOS] Number of Reports per ASN and Key
Operational metrics	43 (in attack, malware, fast-flux, websites, mobile, spam)	- MALWARE Volume per Day - Mobile Malware - Volume per Day - DDOS Attacks Volume by Subcategory	#12 [QUALITY_METRICS][ATOS] DDOS Attack per Subcategory

There are certain differences between the listed and implemented metrics, as not all of the proposed metrics could be implemented. The naming conventions are slightly different and the delivered metrics had to respect certain limitations, most notably the lack of subscriber data in the CCH. This means that the CCH cannot normalize by the number of subscribers in the networks for which they are calculated. For the purpose of this evaluation, we will generate these metrics ourselves. We use

an approach to reliably normalize by the size of a network that we had developed as part of previous work [1]. Our approach cannot be replicated in the CCH because it relies on subscriber numbers from TeleGeography, a commercial market intelligence firm (further details are provided in Chapter 4). Because of intellectual property rights, we can use the size mappings for this evaluation, but we cannot store the subscriber data at the CCH, as it isn't ours. Hence, the CCH cannot provide normalized metrics. Like most other security metrics, these non-normalized are less suitable for comparison across ISPs or countries.

3.3 *Descriptive analysis*

We start by looking at the data provided to us from the CCH and analyzing each of the above metrics. The data spans from April 1st, 2015 till June 28th, 2015 and contains daily reports. For each day, the 17 metrics are provided with their daily values.

All of the data is at aggregate levels – aggregated at the level of Autonomous System (ASN), country or domain. It contains no personally identifiable information like IP addresses or bot IDs or email addresses. The metrics often cover the same data, sliced in different ways: events per ASN, per country, or partner, all add up to the same grand total number of events.

The metrics can be grouped into three major groups: the "event metrics" – generated by all sensors and experiments in ACDC; "IP metrics" – generated by some sensors; and "quality metrics", which is a meta-metric produced for most metrics. As an example, consider a spam-trap: each spam message would be recorded as one event and show up in the event metrics; the IP addresses of the spam-bots would show up in the IP metrics. The confidence of the message being a spam would show up under quality metrics.

The data is organized in each of the metrics around "feeds" or "submission keys". This is true both for cases where they are mentioned explicitly in the name of the metric (e.g. #9 Number of Reports per ASN and Key) as well as for the other cases (e.g. #1 Unique IPs per ASN is also in fact both per ASN and Key). Each feed refers to a specific form of data being submitted into the clearing house, for instance from a specific sensor or experiment. Many ACDC partners submit several feeds.

Interpreting the metrics for the feeds is not straightforward. Documentation from WP3 and the ACDC community portal provide many details about the experiments and sensors that make up the feeds; but they do not mention the feed-IDs directly, or what each value in each context precisely means. This makes comparing the values between feeds, or combining the values from the feeds into one single value for an ASN or country hard. For instance, one cannot add up the number of spam messages and bots. We can assume that it most cases higher numbers indicate a worse security situation.

Internally in the data, the feeds are tagged with certain category labels and a submission key that can be attributed to individual partners. This enables us to construct Table 2, which lists who is submitting feeds and what they contain. Please note that this covers only the data that has been provided by the ACDC experiments. External data source which have the sole purpose of being used for incident notification are not covered here. These sources are not reflected in the data set. However, as far as we know the data volume of these sources is very low compared to the provided data.

The types of data going into the clearinghouse are quite rich. On the other hand, the magnitude of these feeds are extremely different – some feeds have only two events, while others have hundreds of thousands. This makes combing them or comparing them difficult, especially when considering the fact that what they represent or measure is often not fully specified. Figure 4 shows these differences in the events stored in the CCH.

Table 2: Average daily events produced from experiments – grouped by partner and category

Partner ID + Name	Feeds	Attack	Bot	C2 server	Fast flux	Malicious URI	Malware	Spam campaign	TOTAL
1- Telecom Italia	2	282,920.1	-	-	-	-	7.6	-	282,927
2- CERT-RO	2	29,378.4	-	-	-	245.1	-	-	29,623
3- Garr.it/CERT.BE	3	39,156.1	-	-	-	10,768.5	5,485.6	-	55,410
4- ECO/Cyscon	1	-	-	-	-	10,872.5	-	-	10,873
5- Xlab	2	-	0.0	-	-	-	3.5	-	4
7- G-Data	2	-	-	-	-	450.2	3,777.1	-	4,227
10- Inteco	10	-	1480.2	-	102.6	1,354.5	20.1	-	2,957
11- Telefonica	4	3,171.0	-	-	-	80.6	1,046.9	-	4,299
12- Carmet	9	62.7	0.3	-	7.0	306.0	51.7	20.4	448
18- ATOS	6	333.2	19.8	-	33.1	46.7	15.4	-	448
21- De-Cix	1	171,277.2	-	-	-	-	-	-	171,277
22- ATOS	1	-	-	-	1.0	-	-	-	1
26- IFIS	1	-	-	1.6	-	-	-	-	2
27- Garr.it/CERT-BE	2	133.9	-	-	-	-	9.5	-	143
28- ATOS	1	-	-	-	81.4	-	-	-	81
TOTAL	47	526,432	1500	2	225	24,124	10,417	20	**562,721**

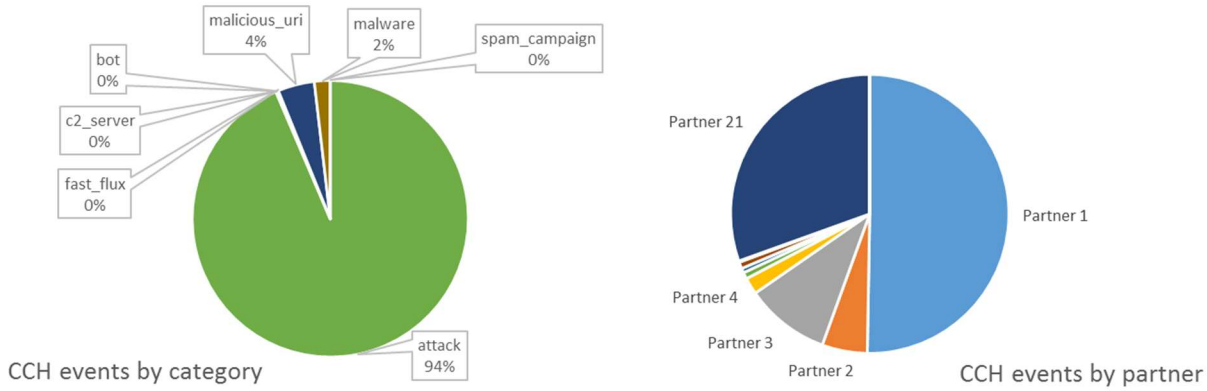

Figure 4: CCH events by category and partner

Table 3 shows statistics for the individual feeds: how many reports do they cover on average, how many unique IP addresses are included in them, as well as what number of ASNs and countries are covered by the feed.

Figure 5 and Figure 6 show the time trends of the top 10 largest feeds from Table 3 (based on the average unique events per day and average unique IP addresses per day). The figures use a log scale so that the different feeds can be shown together. There are a few interesting observations: first, the feeds report a more or less stable number of events or IP addresses on most days. This suggests consistency in the data collection by the sensors. Second, there are periods where certain feeds drop to zero or there are gaps in the data. These typically relate to an experiment or sensor being offline or to errors in processing the logs in the backend for those days.

Table 3: Statistics per feed, sorted by average daily events

Feed nr.	Avg. events per day	Avg. UIPs	Days feed has data	Avg. # of ASNs	Avg. # of Countries	Avg. # of Domains
548	282,920.1	1,531.6	76	450.5	83.9	310.0
412	171,277.2	111,683.9	67	2,789.4	113.5	3,587.4
516	39,260.1	1,095.2	76	437.8	81.7	327.9
538	32,924.0	152.0	58	55.4	22.1	35.0
366	10,872.5	N/A	75	395.4	61.6	1,362.6
517	10,768.5	N/A	76	83.5	37.7	69.7
536	6,834.8	N/A	21	1.0	1.0	0.0
518	5,494.5	N/A	76	1.0	1.0	0.0
563	3,171.0	303.5	76	90.1	34.0	56.6
540	2,275.7	N/A	7	71.4	37.3	66.7
491	1,946.1	635.7	54	195.2	25.1	140.1
569	1,262.9	N/A	63	3.4	3.2	9.6
512	928.1	89.7	14	1.9	2.4	1.0
671	848.3	N/A	52	70.7	14.6	310.5
670	663.2	N/A	23	55.4	10.8	260.7
529	493.8	448.3	8	169.6	50.5	125.3
535	488.7	N/A	35	132.7	42.7	172.8
677	484.5	N/A	54	63.8	9.5	227.2
527	200.0	N/A	53	18.8	6.2	62.6
493	157.5	N/A	4	1.0	1.0	0.0
521	142.7	N/A	9	1.0	1.0	0.0
679	132.8	N/A	57	17.2	7.2	57.3
397	129.0	N/A	5	14.2	4.4	47.3
525	122.2	N/A	63	51.3	18.4	82.5
378	117.5	1.0	62	1.1	1.1	0.0
566	101.6	N/A	59	9.9	9.2	8.5
576	96.6	N/A	9	1.0	1.0	0.0
553	65.1	N/A	28	21.3	14.2	21.7
358	59.3	N/A	14	1.0	1.0	1.0
463	55.2	54.6	14	33.1	16.0	33.9
669	46.7	N/A	52	8.3	3.8	19.2
524	38.3	N/A	62	1.0	1.0	0.0
528	27.1	N/A	36	4.2	3.5	4.4
562	24.0	N/A	25	1.0	1.0	0.0
568	19.0	N/A	7	7.7	3.3	17.4
619	17.0	16.3	3	7.7	3.7	5.0
520	15.2	N/A	58	1.0	1.0	0.0
454	13.1	N/A	63	1.0	1.0	0.0
522	7.8	N/A	56	1.0	1.0	0.0
547	7.6	N/A	76	1.0	1.0	0.0
519	4.8	4.8	4	4.0	2.0	1.3
620	3.5	N/A	41	1.0	1.0	0.0
508	1.6	1.4	39	1.3	1.3	1.2

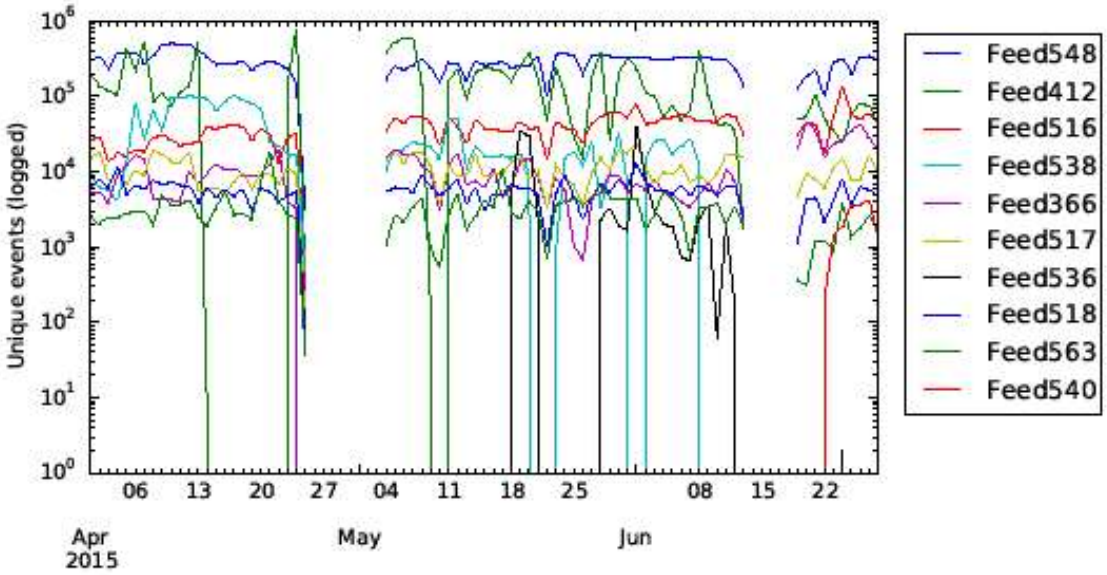

Figure 5: Number of events submitted to the CCH per day (log scale)

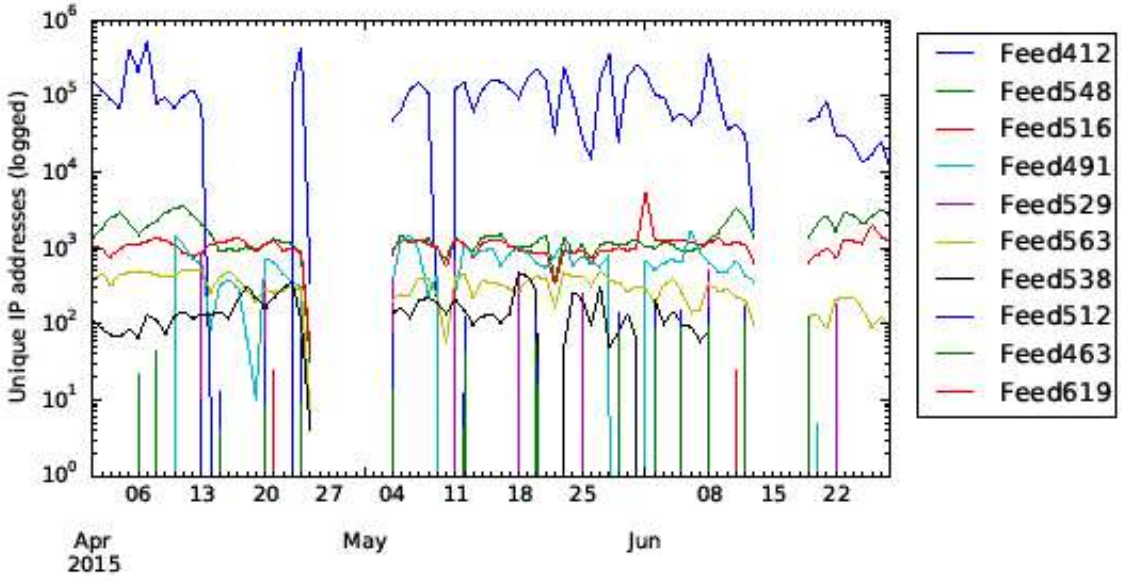

Figure 6: Number of unique IP addresses in the data submitted to the CCH per day (log scale)

3.4 *Statistical properties of the data*

Statistical properties of data can be used to identify gaps and anomalies. A time-proven approach is the use of Auto-Regressive-Moving-Average (ARMA) models to analyse the properties of time series data. As introduced in deliverable 4.1, ARMA models split up the data into deterministic intra-data correlations within time intervals and random fluctuations, allowing forecasting of future values.

We deploy the Box-Jenkins approach to compute and estimate the models, a combination of machine and manual model fitting (please see [2] for more information). We make use of the Python *statsmodels* module in the process. An indication for the applicability of a selected ARMA model and its parameters is given by the autocorrelation function (ACF) of the residuals. The residuals are the differences between the predicted values of the chosen ARMA model and the current time series. If the model is appropriate, all values of the ACF of the residuals are uncorrelated. Additionally, the Akaike Information Criteria (AIC) measure provides a trade-off between the likelihood of the model fitting and the complexity of the model. Although the accuracy of a model may benefit from more parameters, it also leads to overfitting. That is why AIC prefer a less complex model. In the end, it turns out that the ARMA(2,1) model provides for ACDC data a good compromise between complexity and accuracy of the ACDC data. We will apply this model to the different data feeds.

3.4.1 Unique Reports per Subcategory

For each ACDC report, a report category is assigned whereas the category “attack” overarches all different types of attack. The category itself is split into multiple different subcategories comprising, for example, “dos” for all reports related to denial of services attacks, “abuse” for all attacks abusing offered by the attacked system (e.g. Spam), and other which, for example, includes all unspecific attacks against honeypots (e.g. connection attempts).

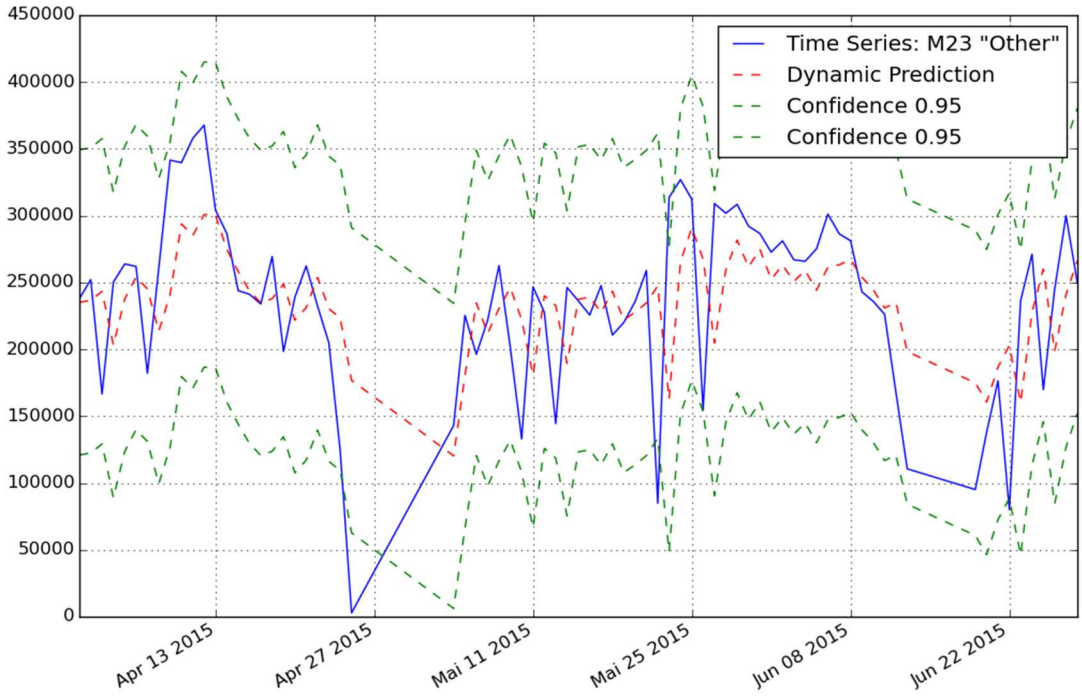

Figure 7: Reports per subcategory "other"

Figure 7 presents the graph for the overall number of reports per subcategory “other” from the 1st April to the 22nd June 2015 (blue line) that have been submitted by the ACDC experiments. The dotted red line represents the dynamic prediction given the parameters of the ARMA(2,1) model. Technically, after estimating the parameters of the recursive formula as presented above for the complete time series, this model has been used to predict all values y_t based on the previous values $y_{(t-1)}$ and $y_{(t-2)}$. The interval between the dotted green line shows the confidence interval for 0.95. The confidence interval states that at a confidence level of 95% the predicted value is between the lower and upper bound. If the observed value is outside of this interval (green dotted lines) it is reasonable to assume, that the anomaly has been caused by a failure of a technical component (sensor or CCH) or that a significant increase or decrease of the attacks has occurred.

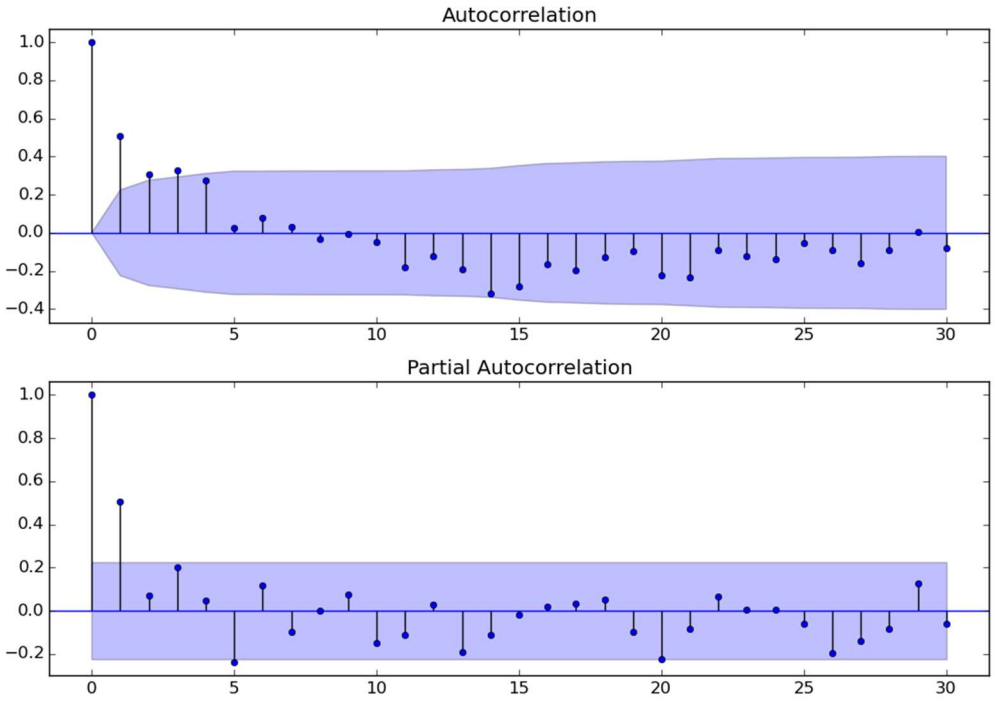

Figure 8 : ACF and PACF for the reports per subcategory “other”

The autocorrelation and partial autocorrelation function for the reports per subcategory “other” in Figure 8 indicates a correlation in the data. It can be seen that the value for a given day depends significantly of the value of the previous day.

Figure 9 presents the results for denial of service (dos) reports -- which also includes distributed attacks (ddos). It can be seen that statistic of dos reports differ from the statistic of the other reports. In contrast to the first subcategory (“other”), the number of detected attack exhibit larger variations. Moreover, it is not a constant activity, instead denial of service attack are conducted in temporal limited periods like bursts. It is important to note, that this characteristic hinders prediction with ARMA models.

It can be seen from the autocorrelation and partial autocorrelation in Figure 10 that the values of two adjacent days are still correlated, but it is much less significantly compared to the other subcategory. The correlation slightly exceeds the confidence level which is represented by the blue area in Figure 10.

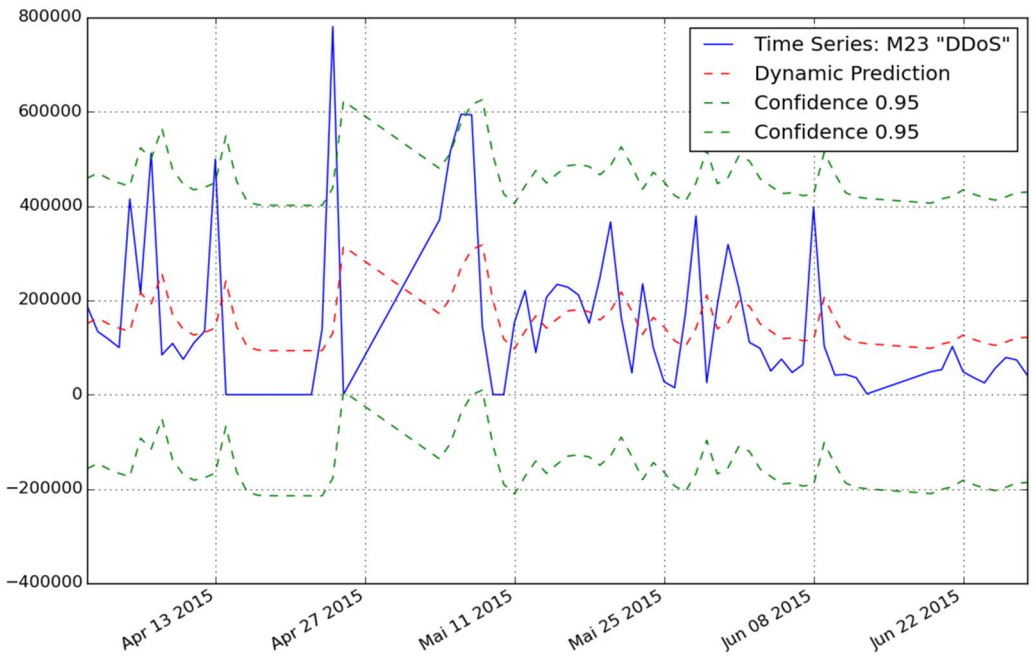

Figure 9: Number of reports per subcategory "dos"

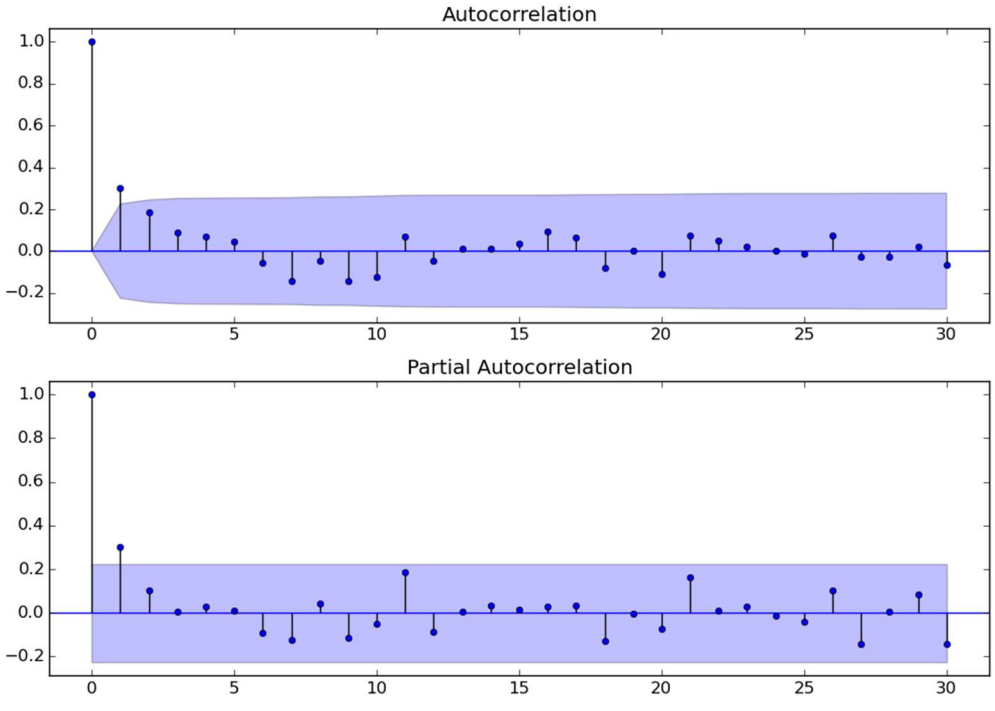

Figure 10: Autocorrelation and partial autocorrelation for the subcategory "dos"

3.4.2 Unique Reports per Category

While the previous section dealt with the subcategories of the report category “attack”, here we analyse the results for the different report categories comprising “attack”, “fast flux”, “malicious URI”, “malware”, and “bot”. As mentioned above, “attack” is an overarching format for different attack classes. The other categories affect the detection of “fast flux” domains, web-server attacking visitors (malicious URI), IPs serving malware, and other attacks related to bots.

The number of reports for the report category “attack” is displayed in Figure 11. Since the category aggregates both the predominant subcategories “other” and “dos”, the graph looks similar to those presented in the previous section.

The graphs for the other subcategories “fast flux”, “bot”, and “malware” are presented in Figure 12 to Figure 14. All graphs exhibit large fluctuations caused by a few outliers. To see their impact on the confidence interval, we also modelled the data after removing them (Figure 16). The upper and lower bound of the confidence interval (green dotted lines) are much closer to the predicted value. Thus the application of the model after the omission of outliers significantly improved the result.

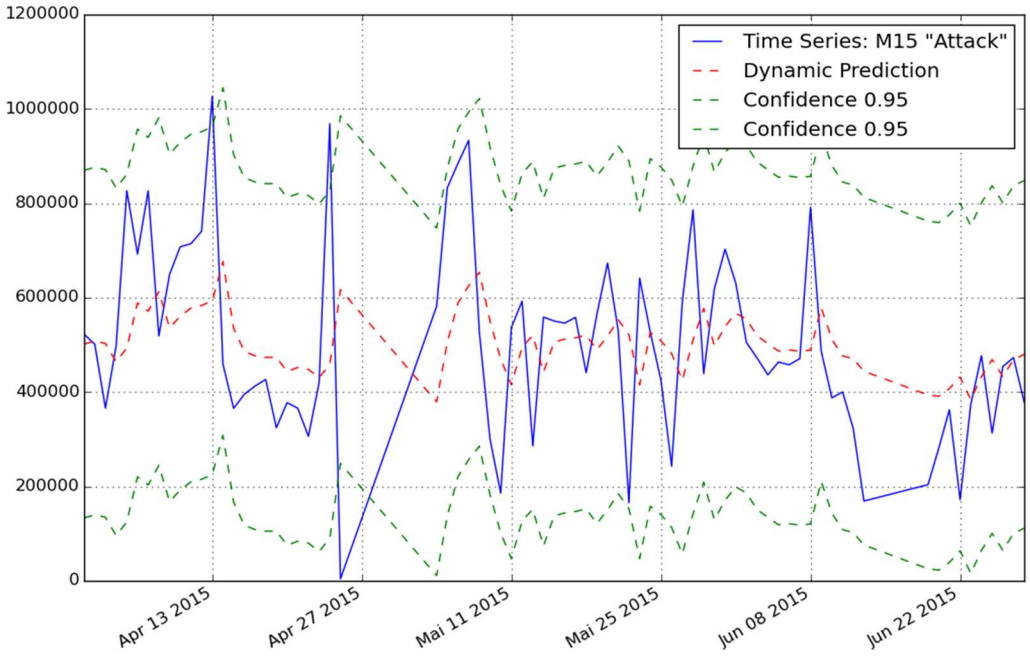

Figure 11: Number of reports per category "attack"

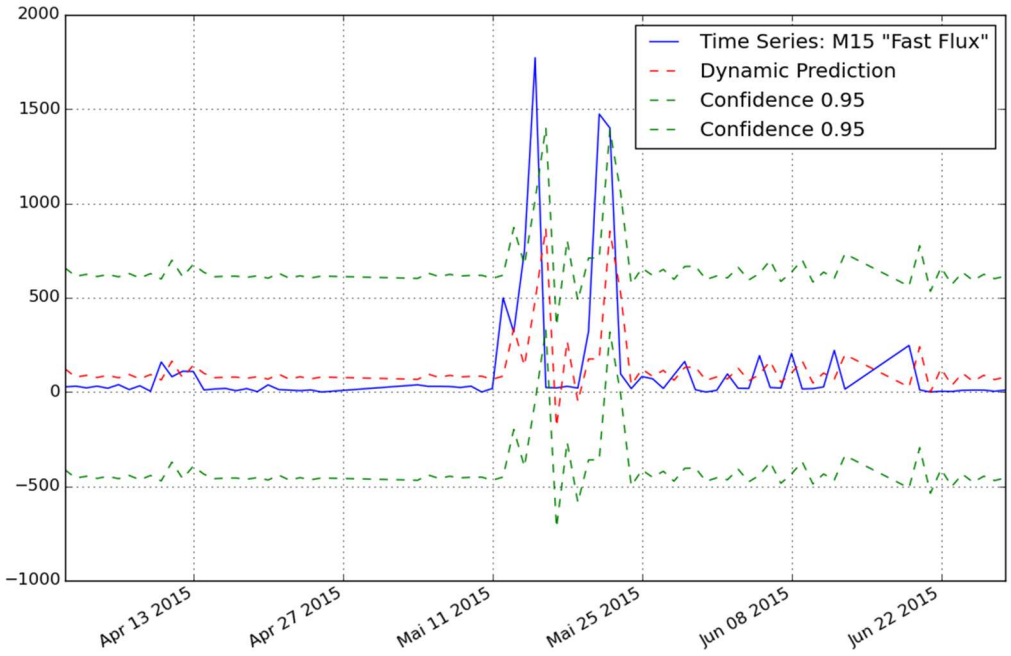

Figure 12: Number of reports per category "fast flux"

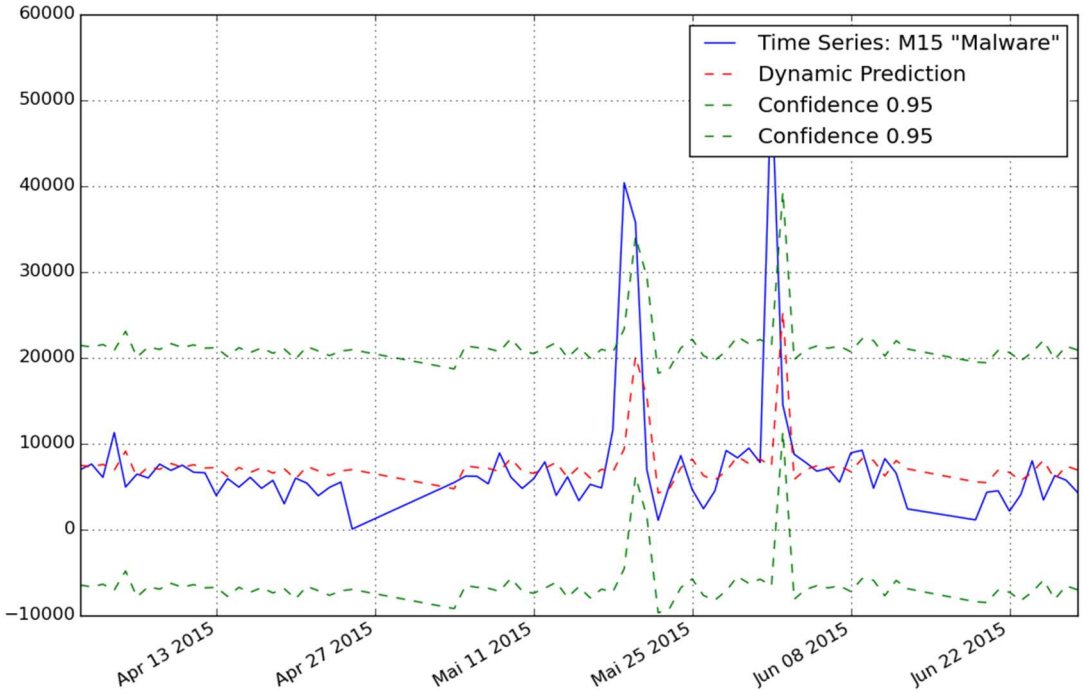

Figure 13: Number of reports per category "malware"

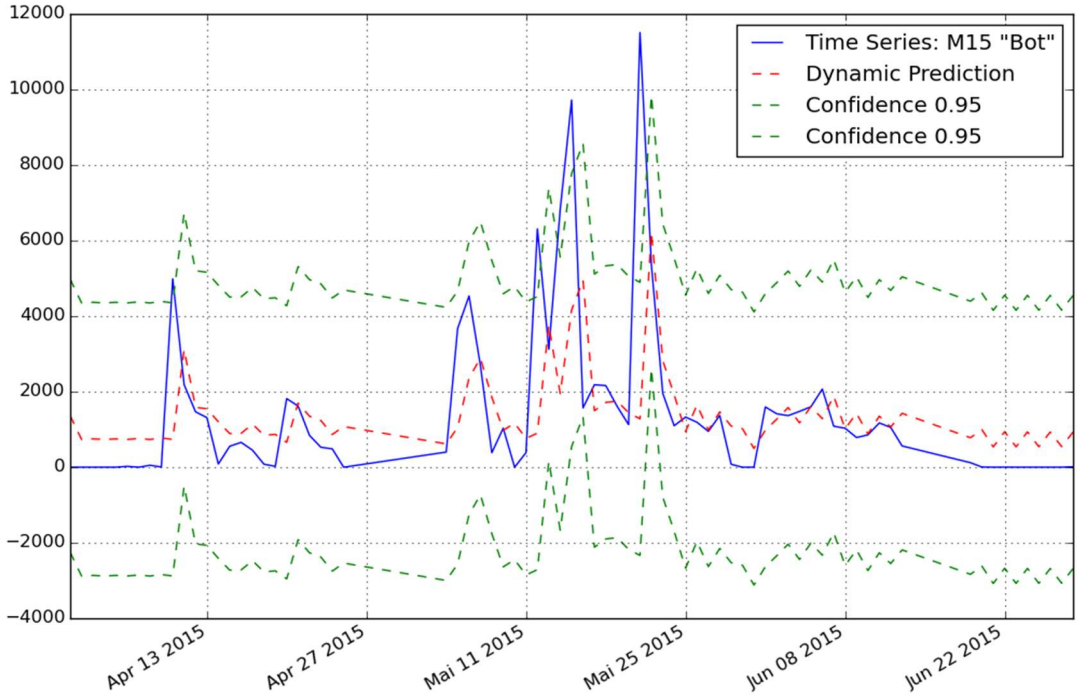

Figure 14: Number of reports per category "bot"

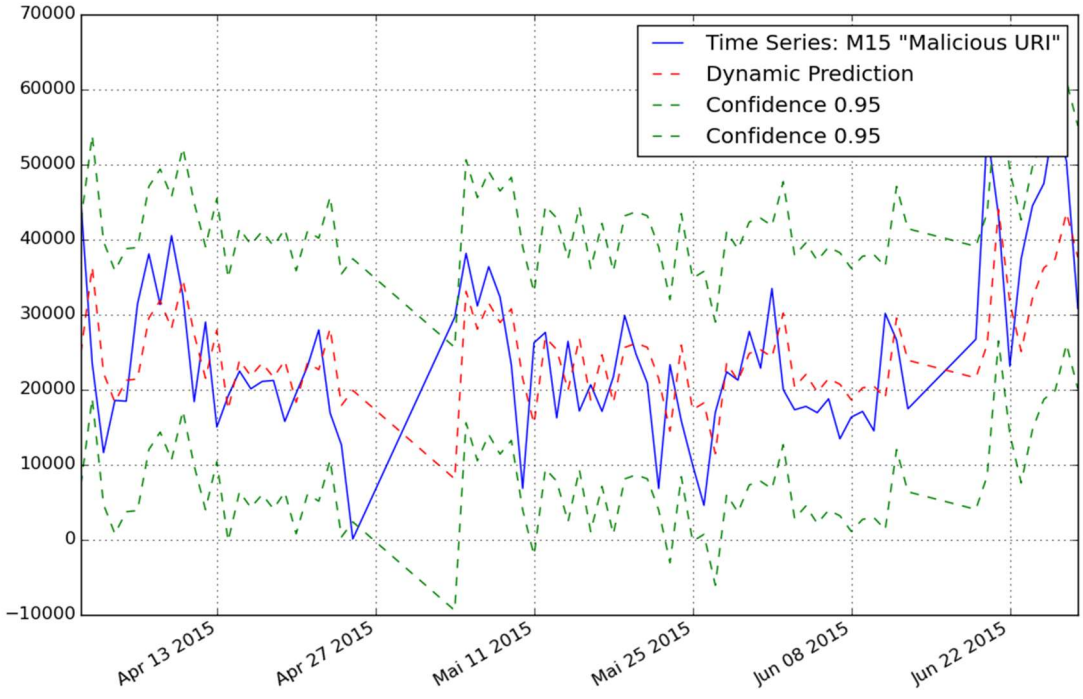

Figure 15: Number of reports per category "malicious URI"

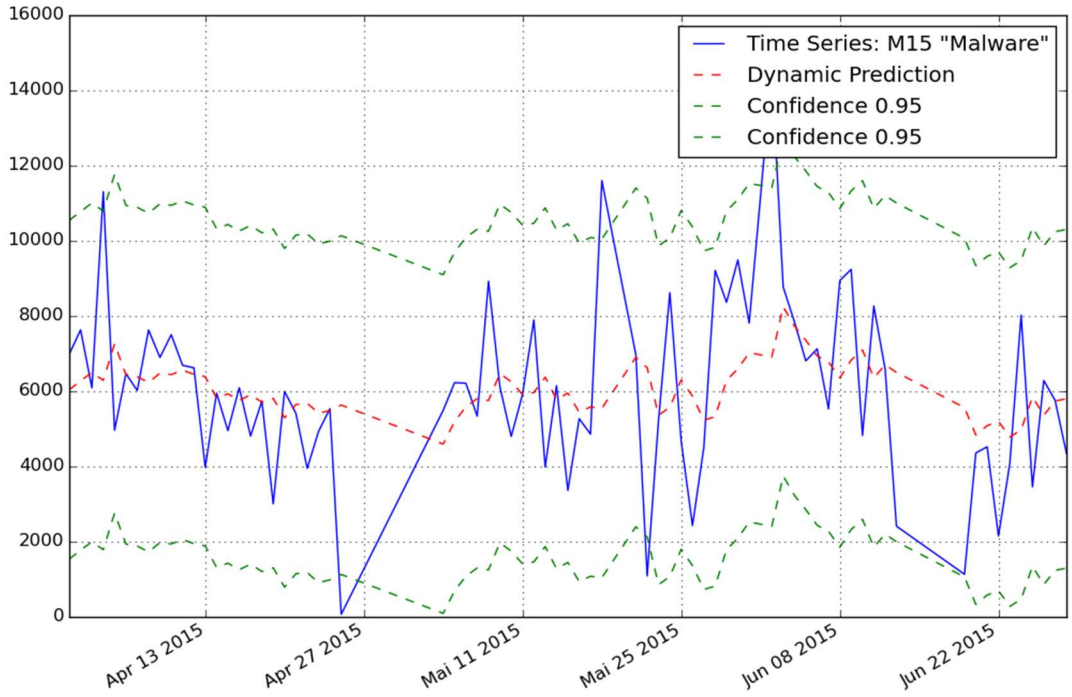

Figure 16: Number of reports per subcategory "malware" omitting outliers

3.4.3 Unique IP Addresses and Reports per Country

This section presents statistics pertaining the number of unique IP addresses and reports per country focussing on Germany and China. Since the results for the other countries do not exhibit any qualitative differences these graphs are omitted in this section.

Figure 17 and Figure 18 present the statistics of the unique IP addresses related to attacks originating from China and Germany. Both graphs have in common that the fluctuations of the IP addresses per day are very high. Furthermore, the day-to-day correlations are not significant.

The statistics detailing the number of reports affecting systems from China and Germany are shown in Figure 19 and Figure 20. It is important to note, that the report statistic for China exhibits a better statistical stability compared to the number of unique IP addresses. The graph in Figure 20 is dominated by a very high peak around the 7th May 2015. Since this peak does not occur in the IP statistic in Figure 18, it is likely that a massive scan originating from a single or a few systems caused that peak.

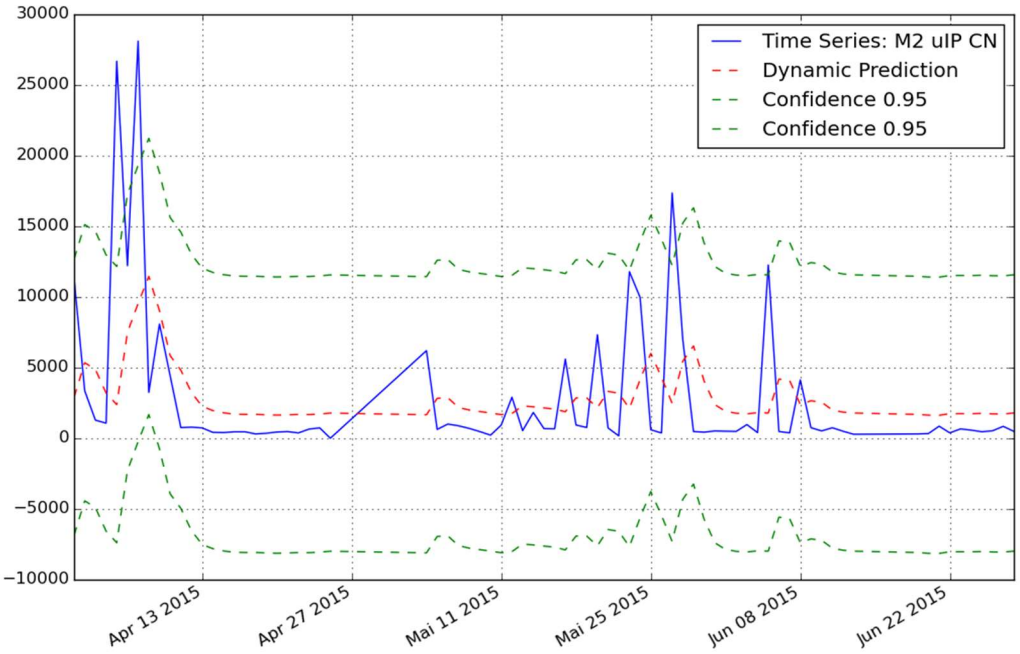

Figure 17: Unique IP addresses per country: China

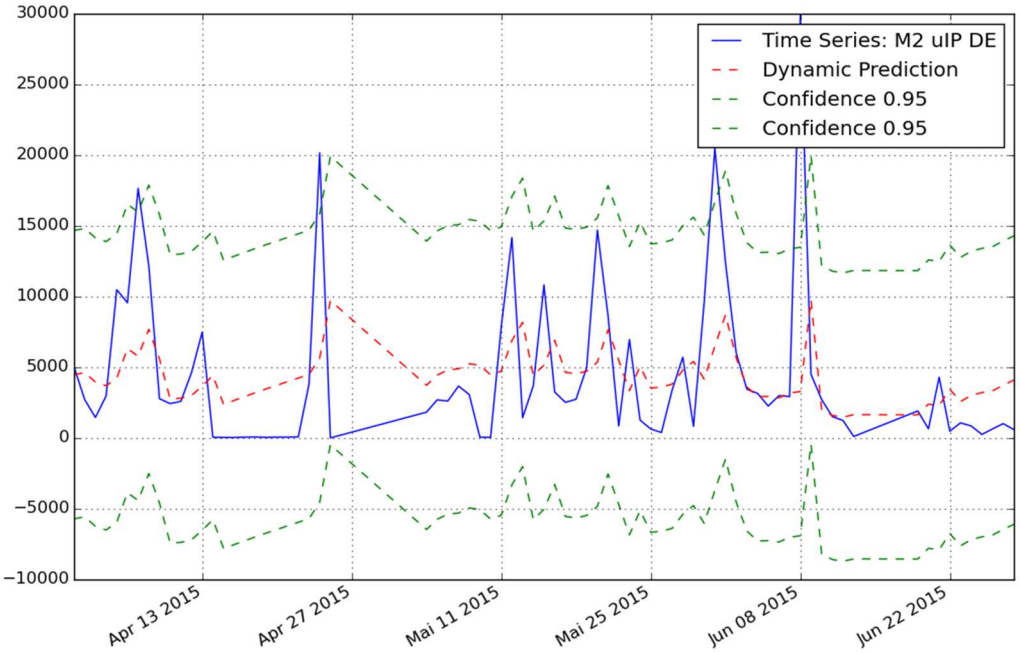

Figure 18: Unique IP addresses per country: Germany

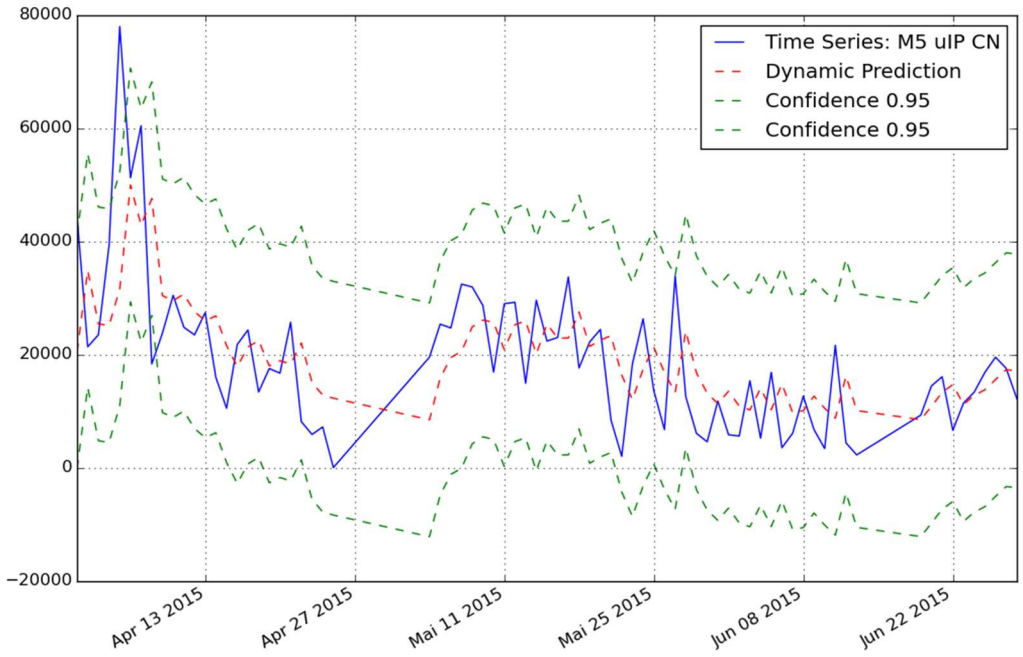

Figure 19: Number of reports per country: China

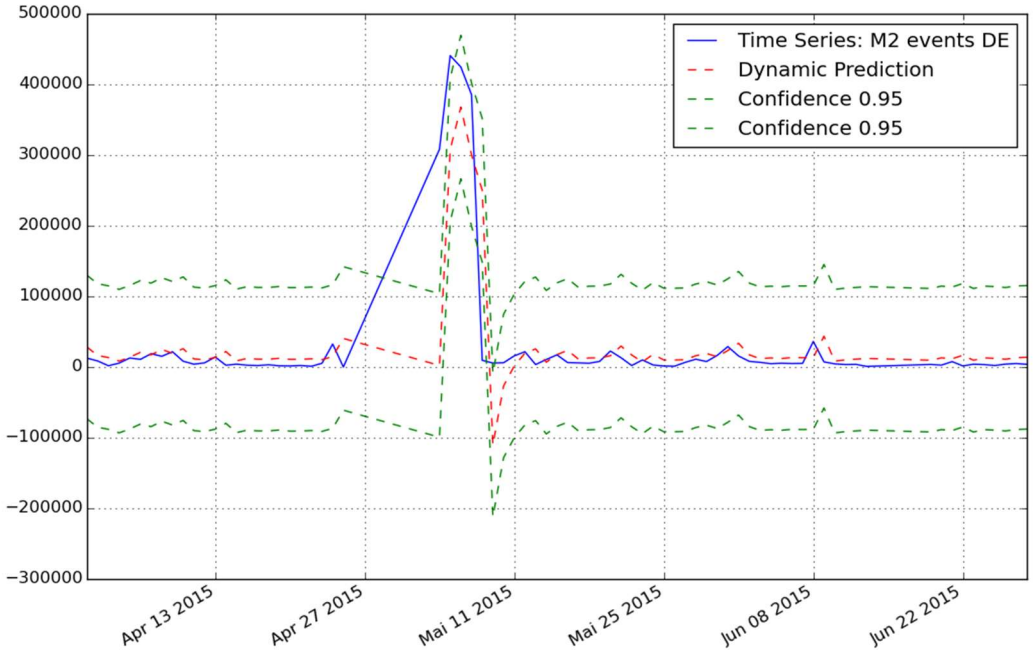

Figure 20: Number of reports per country: Germany

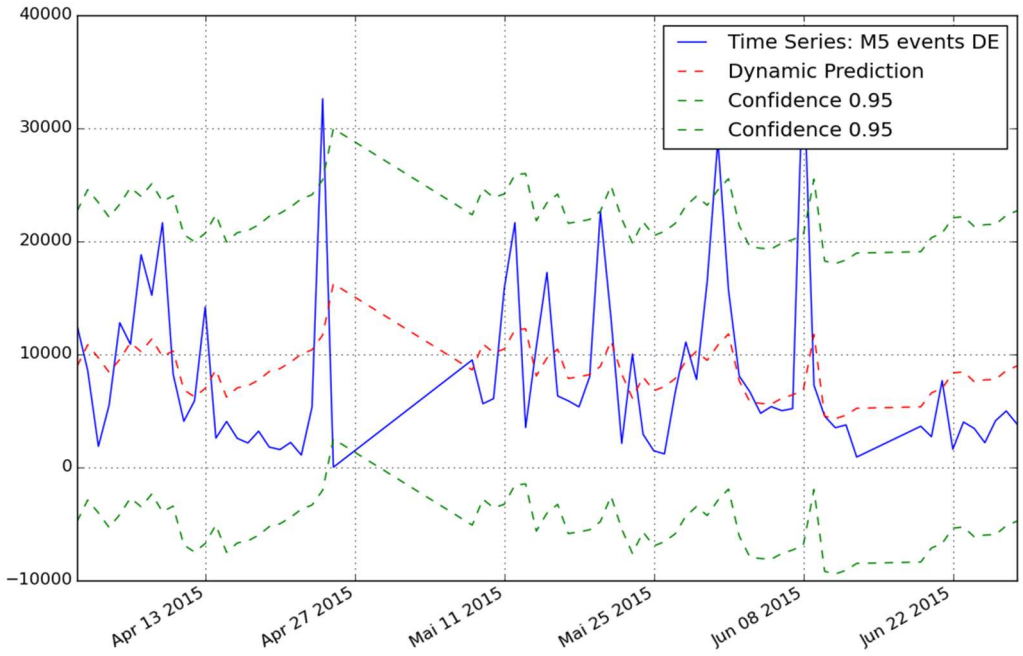

Figure 21: Reports per country omitting outliers: Germany

Figure 21 presents the number of reports related to Germany after removing the large peak described above. It can be seen that this graph is highly correlated to the graph in Figure 18. Thus, the number of reports is quite similar to the number of unique IP addresses.

In summary, the ARMA models provide a good means to understand the properties of the aggregated ACDC data. An important result of the analysis is that the models perform better for highly aggregated data such as the overall number of reports per category or subcategory. Their applicability also depends on the stability of the data. The highly aggregated data exhibits a higher stability compared to other data that, for example, includes the number of reports or IP addresses per ASN. The latter data exhibits a larger non-predictable fluctuation. In particular, the model reveals that the highly aggregated data exhibit temporal correlations between two adjacent lags. Thus, the value of a given day depends significantly of the value of the previous day. This correlation is crucial to accurately predict future values. The prediction capabilities of the ARMA model have been used to detect anomalies. We assume the presence of an outlier if the observed value is outside of the confidence interval of the forecast value. Here, we used a confidence level of 95%. To address large fluctuations in the data, it might be advantageous to address the problem of outliers in the model and to choose a more strict confidence level.

3.5 *Assessing data quality*

After descriptive analysis and statistical validation, we return to the central question in evaluating data quality: to what extent does the data in the CCH improve mitigation by the participating networks and countries? A serious limitation in answering this question is that we cannot look at the detailed event reports themselves – due to privacy restrictions and the way data sharing has been set up around the CCH and Community Portal. Ideally, we would like to ascertain whether the detailed reports that would be sent to partners, such as ISPs, would contain enough information for them to act. The detailed ACDC data format documents suggest this to be the case, but the reports are not present in the research feed for us to validate or refute.

We would also like to check to what extent the ACDC sensor data captures events that are already covered by feeds from entities like ShadowServer and Spamhaus, which are already received by many ISPs. If the ACDC feeds add a lot of new events to existing feeds, that would be of great benefit for mitigation. While we cannot check this directly ourselves, we know from other research that it is quite likely that a lot of the events picked up by the ACDC sensors are unique. A comprehensive analysis by the CERT Coordination Center (CERT/CC) revealed that existing data feeds and blacklists actually had remarkably little overlap, typically less than one percent [3]. In other words, most events were unique to one feed. This suggest that a feed based on different sensors, as is the case for ACDC, would also generate mostly unique events. Unfortunately this cannot be verified with the research feeds. An internal check was conducted by DFN-CERT during 16th April 2014 to the 27th July 2015 regarding compromised machines in the German research networks. They found, positively, that ACDC data had no duplicate with external data sources. This confirms our conjecture that the feeds add unique events to the existing data sources that ISPs receive.

The volume of events submitted to the CCH, around 500,000 every day, and the number of unique IP addresses associated with these events, around 114,000 every day, suggest that the sensors potentially contribute thousands of previously unidentified bots to the data that ISPs are receiving. To illustrate: this number of IP addresses per day is larger than the amount seen in the global sinkhole for GameOver Zeus, but smaller than the amount seen in the Conficker sinkhole. In other words, the volume is non-trivial and, in light of the typical lack of overlap among sources, potentially contributes new events and infections to the existing feeds. This implies the ACDC feed can have substantial benefits for mitigation and can contribute to reducing infection rates.

One quality aspect we can assess more directly with the available metrics, is the “confidence” level reported by the CCH for each of the feeds. In other words, how confident is the data submitter that reported incidents or attacks are correct and not false positives or otherwise problematic.

We have plotted the confidence levels for the various categories in Figure 22. It shows that bot feeds have a confidence level of around 80% and malware feeds of around 50%. Confidence information is important for acting partners, for instance when they need to decide whether to notify or quarantine a customer. This is even truer when the notifications are handled in a (semi-) automated manner. We assume that the presence of the metric means the reliability information is captured by sensors and provided to acting partners for individual reports. On an aggregate level and looking at the timeline, we see a relatively stable patterns, albeit with some fluctuations, suggesting no major change in sensor configurations.

A final aspect and one that we can test is regarding the bias of the data feeds towards certain Autonomous Systems (ASes), which might be the result the sensors geography or collection methodology. We can check for this by looking at the percentage of IP addresses (or events) falling into each network and country for each of the data feeds. This is shown in Figure 23 for the top four feeds. One can see that although particular ASNs or countries rank higher in the feeds, they are not the same in all the feeds, nor do they present the partners themselves. Most experiments or sensors will be dominated by a particular type of attack originating from a source, at some point in time. However, this reflects the actual dynamics of network attacks, i.e., constitutes accurate measurements rather than a bias. For instance, spam originates from certain networks and countries more at various points in time. What’s important is that all charts exhibit a long tail – as indicated by the black section of the charts. Those cover the many ASes or countries with a very low number of IP addresses in the data. The presences of these tails implies that the feeds are not limited to only a select number of sources. Figure 24 corroborates this in another way: it shows the number of ASNs and countries seen in all the feeds combined per day – more than 100 countries and two thousand ASNs are visible at any point in time.

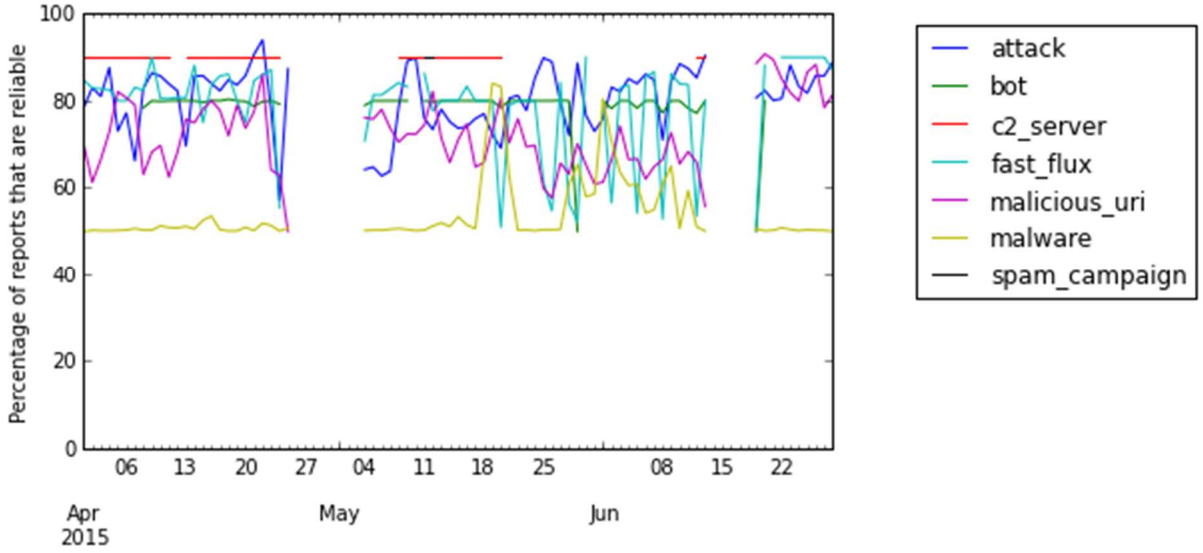

Figure 22 : Confidence levels of data feeds submitted to the CCH

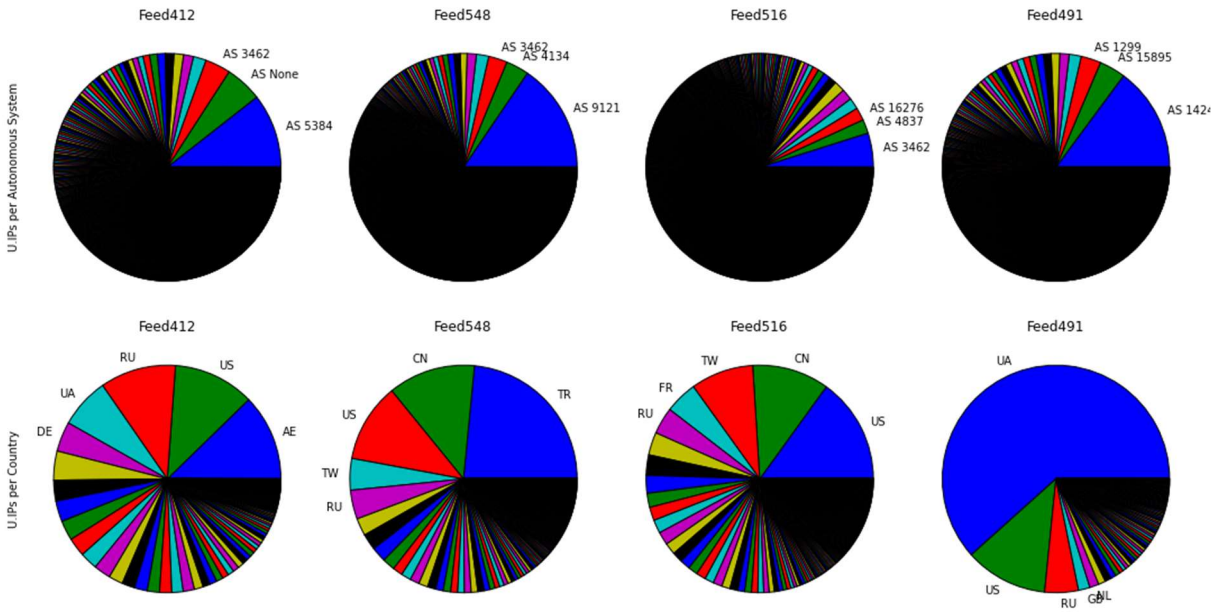

Figure 23: Distribution across ASes and countries of CCH data

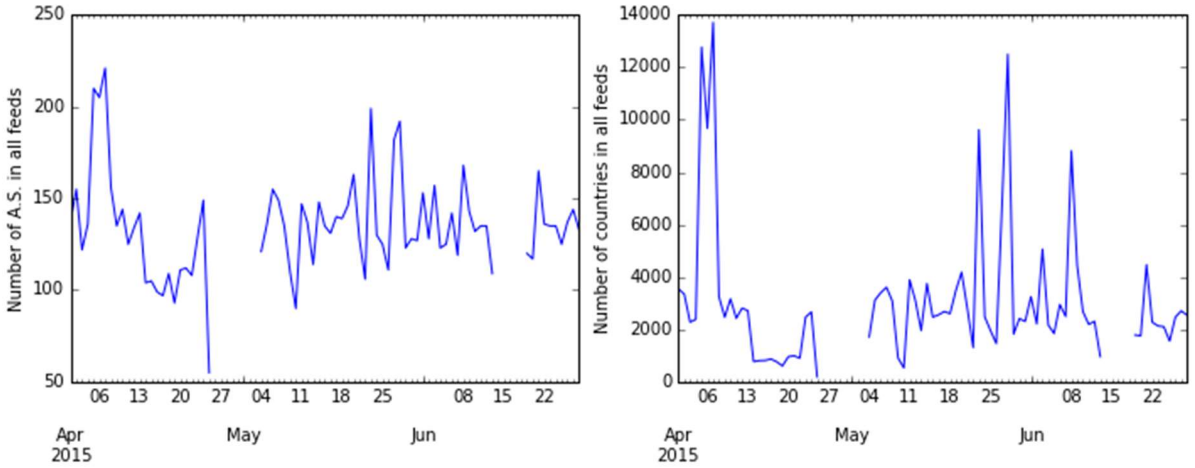

Figure 24: Volume of ASes and countries in CCH data

3.6 *Conclusion*

In this chapter, we assessed the quality of data being submitted to the CCH. Does it enable better mitigation? There are several limitations we have to keep in mind when answering this question. First of all, because of privacy considerations and sharing policies, we could not analyze the full event reports. Related to this, some of the categories are not completely clear in their definition, which makes it difficult to fully assess the aggregate metrics that are now being produced by the CCH.

Keeping these issues in mind, we can conclude that the volume of events being submitted each day is consistently around 500,000. This is a non-trivial amount, in the same range as major botnet sinkholes. Furthermore, there are good reasons to assume that a significant portion of these events are unique to the ACDC data. In other words, the infected machines in this data are not already included in the existing feeds of other organizations like ShadowServer and Spamhaus. If these assumptions hold, the CCH can make an important contribution to enhanced mitigation by ISPs, CERTS and others.

Data submitters give most categories relatively high confidence levels, suggesting they contain a tolerable amount of false positives. Coverage of the data does not seem biased in terms of geography or AS. This further underlines the potential for mitigation.

The second use of the CCH data, to provide metrics than can benchmark and incentivize ISPs and countries in terms of mitigation (see section 2.1), is also feasible, though it is undermined by the fact that the CCH can currently not produce normalized metrics which are more suited for direct comparisons across ISPs and countries.

4 Measuring the Impact of Mitigation

4.1 *Approach*

As has been discussed in Chapter 2, we are going to use a variety of external data sets to generate the metrics on relative infection rates in different countries and ISP networks. The main reason for this is the fact that the CCH has only recently been receiving significant amounts of botnet incident data (i.e., data on infected machines). The external data sets have been shared with TU Delft and cover a longer time period, allowing us to track trends in infection rates. The data has been developed and used in peer-reviewed studies on botnet mitigation [4]

This chapter discusses the data sources we will employ, and the methodological process of deriving comparative infection metrics from this data. In the next Chapter, we are going to use these metrics and explore the impact of ACDC by looking at the trends in infections rates in the countries in which ACDC has helped set up a National Support Center.

4.2 *Observing botnets*

The basic methodology employed in this report consists of collecting and analyzing Internet measurement data on infected machines. We then interpret these measurements by connecting them to other variables related to the operators of the networks containing infected machines, such as the country in which it is located and the number of subscribers. This way we can develop comparative metrics to determine the performance of ISPs and countries with regards to botnet mitigation.

As covered in [5], there is currently no authoritative data source to identify the overall population of infected machines around the world. Commercial security providers typically use proprietary data and shield their measurement methods from public scrutiny. This makes it all but impossible to correctly interpret the figures they report and to assess their validity. The publicly accessible research in this area relies on two types of data sources:

- Data collected external to botnets. This data identifies infected machines by their telltale behavior, such as sending spam or participating in distributed denial of service attacks;
- Data collected internal to botnets. Here, infected machines are identified by intercepting communications within the botnet itself, for example by infiltrating the command and control infrastructure through which the infected machines get their instructions.

Each type of source has its own strengths and weaknesses. The first type typically uses techniques such as honeypots, intrusion detection systems and spam traps. Their advantage is that they are not limited to machines in a single botnet, but can identify machines across a wide range of botnets that all participate in the same behavior, such as the distribution of spam. The drawback is that false positives may occur. The second type typically intercepts botnet communications by techniques such as redirecting traffic, engaging in peer-to-peer communication, or infiltrating IRC channel communication. The advantage of this approach is accuracy: bots connecting to the command and control server are really infected with the specific type of malware that underlies that specific botnet. The downside is that measurement only captures infected machines within a single botnet. Given the fact that the number of botnets is estimated to be in the hundreds [6], such data is probably not representative of the overall population of infected machines.

Neither type of data sources sees all infected machines, they only see certain subsets, depending on the specific data source. In general, one could summarize the difference between the first and the second source as a trade-off between representativeness versus accuracy. The first type captures a more representative slice of the problem, but will also include false positives. The second type

accurately identifies infected machines, but only for a specific botnet, which implies that it cannot provide a representative picture.

Taking these criteria into account, we have obtained the following data sources, in which we group into two categories: global sources and Dutch sources (only NL IP addresses). From these sources, only spam is categorized as "external" to the botnet; all the other sources are internal obtained either via sinkholes or sandboxes.

4.3 *Data sources for the evaluation*

4.3.1 Spam trap (Spam)

Spam data are obtained from a spamtrap located on a domain in the West of the United States. The spam-trap is an email server that hosts unused email addresses and hence receives only spam. The spam-trap used in this study has been running for over a decade. It logs the IP addresses of machines sending the spam. Since botnets have been the dominant distribution mechanism for spam during recent years, many of these machines are bots. We reduce the false positive rate by including only sources located in the networks of retail ISPs. In 2010, the spam trap recorded an average of 888,000 unique IP addresses and 162 million spam messages every day. Triangulation with spam reports of security vendors shows that the sample is representative.

There is no guarantee that the listed spam sources are indeed originating from botnets, though so far that is still the main platform for distribution. The more important limitation is that the spam has become a less important part of the botnet economy, as indicated by the substantial drop in overall spam levels. The reports of security firms confirm these overall trends. Symantec reported a significant decrease in the volume of spam messages, "from highs of 6 trillion messages sent per month to just below 1 trillion" [7] until 2012. Cisco, TrendMicro and Kaspersky show that the spam volume has been fluctuating since that period, but stayed at more or less the same level (see [8] and [9]). All of this means that the source is becoming less representative of overall infection levels.

4.3.2 Conficker Sinkhole (Conficker)

Established in 2004, the ShadowServer Foundation comprises volunteer security professionals that "gathers intelligence on the darker side of the Internet". They have created the Conficker working group, which provides reports and data on "the widespread infection and propagation of Conficker bots" [10].

Several members of the working group run sinkholes that continuously log the IP addresses of Conficker bots. The sinkholes work in this fashion: computers infected with Conficker frequently attempt to connect to command and control servers to receive new payloads (i.e., instructions). In order to protect the botnet from being shut down, Conficker attempts to connect to different C&C domains every day. The working group has succeeded in registering some of these domain names and logging all connections made to them. Since these domains do not host any content, all these connections are initiated by bots. Therefore, we can reliably identify the IP addresses of the Conficker bots.

The Conficker dataset is unique in several ways. First of all, unlike the other two datasets, it is not a small sample of a much larger population, but rather captures the universe of its kin. This is because of the way the bot works – most of them will eventually contact one of the sinkholes. Second, this dataset is basically free from false positives, as, apart from bots, no other machine contacts the sinkholes. These features make the dataset more reliable than external datasets like spam or firewall logs. The difference, however, is that the dataset is only indicative of the patterns applicable to one specific botnet, namely Conficker. Although Conficker has managed to replicate very successfully,

with around several million active bots at any given moment, it has not been used for any large-scale malicious purposes – or at least no such uses have been detected yet. This means ISPs and other market players may have less powerful incentives to mitigate these infections, different from spam bots, for example. These differences make the Conficker dataset complementary to the two other sets. Overall, the Conficker dataset adds a fresh, robust and complimentary perspective to our other two datasets and brings more insight into the population of infected machines worldwide.

4.3.3 GameOver Zeus sinkhole

Zeus botnet started making headlines in 2007, as a credential stealing botnet. The first version of Zeus was based on centralized command and control (C&C) servers. The botnet was studied by various security researchers and multiple versions were also tracked [11]–[14] . In recent years Zeus has transformed, into more robust and fault tolerant peer-to-peer (P2P) botnet, known as P2P Zeus or Gameover Zeus. The botnet supports several features including RC4 encryption, exfiltrating stolen information, anti-poisoning and auto blacklisting. It also can be divided into sub-botnet, based on BotIDs , where each sub-botnet can be used to carryout diverse task controlled by different botmasters.

The botnet is divided into three sub-layers, which provide following functionality.

- Zeus P2P Layer (Peer): This is the bottom most layer and contains information of infected machines. Bots in P2P layer exchange peer list with each other in order to maintain updated information about compromised machines.
- Zeus Proxy Layer (Proxy) : A subset of bots from P2P layer are assigned the status of proxy bots. This is done manually by the botmaster by sending proxy announcement message. Proxy bots are used by Peer-to-peer layer bots to fetch new commands and drop stolen information.
- Domain Generation Algorithm Layer: DGA layer provides a fallback mechanism, if a bot cannot reach any of its peers, or the bot cannot fetch updates for a week. The Zeus algorithm generates 1000 unique domain names per week. Bots which lose connection with all connected peers search through these domains until they connect to live domain.

More details about the architecture and functioning of the botnet can be found in literature.[15] . This dataset is sub-divided into three feeds, GameOver Peer, GameOver Proxy and GameOver DGA. The botnet is spread in around 212 countries with on tens of thousands unique IP addresses per day. Hence it is gives us insight of botnet infection level at global level, and compare various countries and ISPs.

4.3.4 ZeroAccess sinkhole

ZeroAccess is a Trojan horse, which uses a rootkit to hide itself on Microsoft Windows operating systems. The botnet is used to download more malware and open a backdoor for the botmaster to carry out various attacks including click fraud and bitcoin mining.

The botnet is propagated and updated through various channels, including but not limited to drive-by downloads, redirecting traffic and dropping rootkits at potential host or updating the already compromised host through a P2P network. The ZeroAccess sinkhole data comes from several machines that impersonate nodes and supernodes in the network, which allows them to intercept P2P communication, including data on other infected machines in the network. These nodes see bots in around 220 countries with an average of about 12,000 unique IP addresses per day. It was generously shared with us by the operators of a ZeroAccess sinkhole, namely the team of Katsunari Yoshioka at Yokohama National University in Japan.

4.3.5 *Morto sinkhole*

Morto is a worm that exploits the Remote Desktop Protocol (RDP) on Windows machines to compromise its victims. It uses a dictionary attack for passwords to connect as Windows Administrator over RDP to vulnerable machines in the network. After successfully finding a vulnerable machine, it executes a dropper and installs the payload.

We have a time series data of Morto for the past 4 years with an average of 5,000 daily unique IP addresses distributed globally. This is relatively small, but it complements our other data sources by providing a longitudinal perspective. This data was also kindly shared with us by the team of Katsunari Yoshioka at Yokohama National University in Japan.

4.4 *Mapping infected machines to ISPs and countries*

For each unique IP address that was logged in one of our data sources, we looked up the Autonomous System Number (ASN) and the country where it was located. The ASN is relevant, because it allows us to identify what entity connects the IP address to the wider Internet -- and whether that entity is an ISP or not.

However, there are some ISPs, that operate in various countries across Europe. We employ IP-geolocation databases [16] from Maxmind [17] to single out IP addresses used in each of the European countries

As both ASN and geoIP information change over time, we used historical records to establish the origin for the specific moment in time when an IP address was logged in one of our data sources (e.g., the moment when a spam message was received or network attack was detected). This effort resulted in time series for all the variables in the datasets, both at an ASN level and at a country level. The different variables are useful to balance some of the shortcomings of each – a point to which we will return in a moment.

We then set out to identify which of the ASNs from which botnet IP data belonged to ISPs – that is, the actual companies who manage these networks. To the best of our knowledge, we developed the first database that reliably maps ASNs onto ISPs. This is not surprising. Estimates of the number of ISPs vary from around 4,000 – based on the number of ASNs that provide transit services – to as many as 100,000 companies that self-identify as ISPs – many of whom are virtual ISPs or resellers of other ISPs' capacity.

So we adopted a variety of strategies to connect ASNs to ISPs. First, we used historical market data on ISPs – wired, wireless and broadband – from TeleGeography's GlobalComms database 2013[18]. We extracted the data on all ISPs in the database listed as operating in a set of 60 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. To this list we added 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 below.

The process of mapping ASNs to ISPs was done manually. First, using the GeoIP data, we could identify which ASNs were located in each of the 60 countries. We rank these ASNs by a number of criteria and listed 2260 ASNs for the next steps; these ASNs cumulatively cover around 80 percent of bots IPs, and approximately 75 percent of the IP addresses in their respective countries. We used historical WHOIS records to lookup the name of the entity that administers each ASN in a country.

We then consulted a variety of sources – such as industry reports, market analyses and news media – to see which, if any, of the ISPs in the country it matches. In many cases, the mapping was straightforward. In other cases, additional information was needed – for example, in case of ASNs named after an ISP that had since been acquired by another ISP. In those cases, we mapped the ASN to its current parent company.

It is important to notice that ISP change their AS size over time (mergers, selling/buying blocks). To cope with this, we use historical BGP data and produce our metrics matching the timestamp of the botnet data with the BGP tables of the respective period.

4.5 *Compensating for known limitations in Internet measurements*

4.5.1 Dealing with churn

Our approach allows us to robustly estimate the relative degree in which ISP networks harbor infected machines. It has certain limitations, however, that need to be compensated for. The effects of two technical issues need to be taken into account when interpreting the data: the use of Network Address Translation (NAT), the use of dynamic IP addresses with short lease times. The key issue is to understand how these technical practices affect the number of machines that are represented by a single unique IP address.

NAT means sharing a single IP address among a number of machines. Home broadband routers often use NAT, as do certain other networks. This potentially underrepresents the number of infected machines, as multiple machines show up as a single address. Dynamic IP addresses with short lease times imply that a single machine will be assigned multiple IP addresses over time. This means a single infected machine can show up under multiple IP addresses. As such, it over-represents the number of infected machines. Both of these practices counteract each other, to some extent. This limits the bias each of them introduces in the data, but this does not happen in a consistent way across different networks.

This is a classic problem in the field of Internet measurement: how many machines are represented by a single IP address? Ideally, one IP address would indicate one machine. But reality is more complicated. Over an extended time period, a single address sometimes indicates less than one machine, sometimes more than one. This varies across ISPs and countries. Earlier research by Stone-Gross *et al* .[19] has demonstrated that in different countries, there are different ratios of unique IP addresses to infected machines – referred to as ``DHCP churn rates'' .

In this report, to control for the bias caused by churn rates, we use shorter time scales when counting the number of unique IP addresses in a network, for all the datasets.

On shorter time scales, the potential impact of churn is very limited. Earlier research found that churn starts to affect the accuracy of IP addresses as a proxy for machines on timescales longer than 24 hours. 12 We therefore worked with a time period of 24 hours. All our comparative analyses are based on the daily average number of IP addresses from an ISP network. This compensates for churn, but has a downside: in these estimates, the number of infected machines may be now grossly undercounted, depending on the prefix, AS and/or ISP evaluated.

While the number of bots measured in a 24 hour period is the most reliable for comparisons across networks, it cannot indicate the actual infection rate of a network in absolute terms. For absolute estimates -- in other words, of the actual number of infected machines – we use larger time periods, depending on the situation: months, quarters or even the whole 18-month measurement period. For sources we have checked both the daily average number of unique IP addresses, and the total number of unique IP addresses for that particular metric. That way, we can compensate for the various measurement issues. Patterns that hold across these different measurements can be said to

be robust and valid. These measurement issues are revisited in more detail in each section of the findings chapter.

The result of this approach is time series data on the number and the location of infected machines across countries and ISPs. We have paid special attention to whether these machines are located in the networks of the main ISPs in the wider OECD.

4.5.2 Actively measuring DHCP churn

As part of its work within ACDC, the team at TU Delft has been developing and testing an active-measurement based approach that are able to capture, for each ISP/AS and network block, their respective DHCP churn rates. This approach sets out to estimate ISP and Internet-wide DHCP churn rates, in order to better understand the relation between IP addresses and hosts, as well as allow us to correct data relying on IP addresses as a surrogate metric.

The first results passed peer review and were published in [20]. We proposed a scalable active measurement methodology and then validated it using ground truth data from a medium-sized ISP. Next, we built a statistical model to estimate DHCP churn rates and validate against the ground truth data of the same ISP, estimating correctly 72% of DHCP churn rates. Finally, we apply our measurement methodology to four major ISPs, triangulate the results to another Internet census, and discuss the next steps to more precisely estimate DHCP churn rates.

Since then, the work has been maturing and the more recent advances are currently undergoing a second round of peer review. We hope to be able to publish DHCP-churn normalized metrics as an outcome of ACDC in the second half of 2015.

4.6 *Generating infection metrics*

Using the data sources and infection counts outlined above, we were able to generates the metrics as defined in D4.1. These basically include the average number of unique IP addresses (IPs) observed daily in each of the infection datasets per country, per ASN, and per ISP (operator). The period can differ – unless specified otherwise, it is based on data from July 1st 2014 till June 30th 2015. The metrics include both the absolute counts, and in some instances, normalized counts – that is divided by the number of broadband subscribers in the particular country or ISP.

In the next Chapter, we will use these metrics to evaluate infection trends in different countries and ISPs, most notably in those countries that have set up a NSC.

5 Evaluating the Impact of National Support Centers

5.1 *Approach*

As was discussed in Chapter 2, it is not possible to undertake a straightforward statistical analysis of the impact of ACDC in terms of enhanced mitigation and reduced infection rates. Mitigation specifically based on ACDC data has only recently begun, and this provides too little time for the impact to be statistically discernable in time series data. Also, ACDC is not the only project disseminating data to support cleanup. CERTs and ISPs are already receiving data from other entities, like ShadowServer or Spamhaus. The volume of the ACDC data is non-trivial (see Chapter 3), but not so large that its impact can be distinguished from those other sources over the short term.

To overcome these important limitations, our approach is to look at the improvements in mitigation in the countries that have set up National Support Centers (NSCs). The NSCs are at the heart of the end to end approach of ACDC: botnet data identifying infected machines is combined with tools to support cleanup by ISPs, end users and others.

The question this chapter sets out to answer is: are NSCs associated with enhanced mitigation and a reduction of the botnet problem? We use the term 'associated' because the setup of an NSC itself does not come out of the blue, nor does it occur in a vacuum. It is typically a part of a broader development in that country: increased awareness and willingness among public and private actors to act against botnets. For this reason, we will look at the current infection rates in the countries with an NSC compared to other countries. We also want to explore the improvements these countries have gone through over time. This is very difficult, because longitudinal botnet data is very scarce. We have data from three sources that have been operating for several years, but we need to keep the limitations in mind when using them, as infection levels that can be seen from the sinkhole data are influenced by other factors as well.

We selected 60 countries to be included in the analysis – including 8 countries with an NSC. All countries were ranked based on their number of bots and those that cumulatively make up the top 80 percent were selected. Missing remaining countries from the OECD and the European Union were added for comparative purposes.

We will study infection rates at both the level of countries and the level of ISPs. There are several reasons for this. The country-level metrics are the most inclusive in terms of botnet incidents. But the bulk of mitigation has to be achieved by ISPs, because the bulk of the infections are there [5] . Furthermore, the ISP-level metrics are based on more reliable data in terms of the number of users present in each network, as they are based on Telegeography's GlobalComms database of market intelligence on broadband markets (see Chapter 4 for more details).

The core methodological challenges of developing comparative botnet metrics have been extensively discussed in D4.1. Several metrics were designed to deal with these challenges (D4.1, Chapter 5). In this evaluation, we will focus on metrics based on IP addresses:

- Daily absolute number of unique IP addresses per country and per ISP
- Daily relative number of unique IP addresses per country and ISP (which is the absolute number normalized by the number of Internet users or ISP subscribers)
- Rank position of country or ISP compared to other ISPs or countries, based on the daily absolute number of unique IP addresses
- Rank position of country or ISP compared to other ISPs or countries, based on the daily relative number of unique IP addresses

In sum: we will first present the current situation, a static snapshot of current infections levels. In other words, have the ongoing mitigation strategies in those countries led to better performance

than in other countries, as measured in lower infection levels per user/subscriber. Then, on a subset of the data, we will explore whether there is evidence for improvements over time. We will present the metrics for each of the data sources separately, in order to give readers the most unbiased and transparent view into the data.

5.2 *Background on the centers*

In the course of the project, ACDC was involved in the setup of NSCs, inspired by the German center that was launched in 2010. Subsequently NSCs were started in Spain (ES), Italy (IT), Croatia (HR), Belgium (BE), France (FR), Portugal (PT) and Romania (RO). There are also associated initiatives in Luxemburg and Bulgaria, but these partnerships are still gearing up, so we are leaving them out of the group of countries to evaluate.

The NCSs are not all the same. There are certain shared requirements, but also room for diversity. We won't go into them in detail (see D1.3.2 for an elaborate description of the different NSCs). Since the evaluation focuses on impact in terms of infection levels, the key element that we focus on is whether they have an operational role and contribution in mitigation. The first four we mentioned (DE, ES, IT, HR) have such a direct operational contribution. They have an abuse notification process that directly contacts ISPs and CERTs with relevant incidents in their networks. The other centers focus more on providing information and support to consumers and small and medium enterprises via a website, without an operational abuse handling process with the ISPs.

For our analysis, this means we will assess the performance of this group of countries with an NSCs: DE, ES, IT, HR, BE, FR, PT, RO. We will compare them to a wider population of 52 countries and to a special reference group: a set of countries that already had an national anti-botnet center or initiative: Australia, Finland, Ireland, Japan, Korea.

5.3 *Country-level analysis*

Table 4 shows the number of infections per country for each dataset (as counted by the daily average number of unique IP addresses). It lists the top 10 countries with the highest number of infections and the rates. The rates for the ACDC member countries are in Table 5. Next to each country is the rank number: the relative position within the group of 60 countries. Position 1, the lowest rank number, means that this is the country with the most infections. The higher the rank number, the more clean a country is – or reverse: the more countries are worse off than this country. Rank number 60 is the least infected country, according to that specific metric.

Analyzing this table, we see that larger countries tend to dominate the rankings on the absolute size of botnets. That has been a consistent finding in earlier research as well: the number of infections is a function of the number of users. In this respect, we should not be too surprised to find ACDC-member countries France, Italy and Germany in the top 10 for some botnet data sources. That being said, it is not just a matter of population size, because then none of these countries would be in the top 10. Wealth and the quality of infrastructure have also been shown to increase infection rates [1].

To create a more informative comparison of mitigation effectiveness, we need to control for the size of the user base in each country. This gives us a metric that calculates the number of infected sources *per user* in the country. In and Table 7 we present the rankings based on the relative number of infections per million broadband subscribers, as reported by the TeleGeography's GlobalComms database [18].

This table shows that rankings differ across botnets when we normalize for the number of subscribers in those countries. This is expected. We know that botnets have different geographical patterns. What this means is that any evaluation of infection rates needs to be based on an array of different sources, in order to compensate for these botnet-specific patterns. These difference are not

extreme. There are still significant correlations among the different normalized rankings (see Table 8 for the correlations). In other words, the rankings are not just representing some stochastic processes, but do contain relevant signals as to the performance of countries.

We can see that some ACDC member countries are still in the top 10 (IT, PT, RO). Italy, in fact, is ranked as the third most infected country for GameOver Zeus Peer. Italian users have been especially vulnerable to GameOver Zeus, and the country is ranked higher than most other ACDC countries for other sources as well. The other ACDC members are distributed along different positions on the ranking, though most are above rank 30 – that is, they are better than average.

Table 4: Top 10 countries (daily average of unique IP addresses)

#	**GameOver Zeus**		**Conficker**		**Morto**		**ZeroAccess**		**Spam**	
	CC	IPs	CC	IPs	CC	IPs	CC	IPs	CC	IPs
1	JP	4969.09	CN	124620.9	CN	235.2	US	2374.33	RU	10490.21
2	IT	4617.21	RU	47342.7	BR	209.4	ES	1761.28	US	4228.35
3	UA	2363.82	BR	46870.92	US	184.86	TR	1518.86	CN	3583.23
4	US	2032.83	VN	46028.21	RU	95.03	IT	1404.04	KR	1787.86
5	GB	1919.87	ID	35180.84	TR	93.74	JP	1143.62	UA	1741.82
6	BY	1540.63	IT	25481.11	TH	80.88	BR	919.55	VN	1456.63
7	ID	1338.53	KR	24714.88	VN	66.83	FR	800.51	BR	1041.16
8	VN	965.2	US	24017.43	IT	66.12	VN	767.17	JP	1039.63
9	FR	911.74	PK	23018.77	DE	50.84	AR	760.22	KZ	901.36
10	TR	856.73	AR	22683.05	AR	34.46	TH	744.47	AR	888.05

Table 5: Ranking of ACDC countries (daily average of unique IP addresses)

	GameOver Zeus		**Conficker**		**Morto**		**ZeroAccess**		**Spam**	
CC	Rank	IPs	Rank	IPs	Rank	IPs	Rank	IPs	Rank	IPs
BE	41	75.13	42	1348.54	26	12.63	35	66.74	45	42.99
DE	12	690.32	21	8653.01	9	50.84	11	600.43	15	433.22
ES	26	268.49	12	20258.65	11	34.35	2	1761.28	25	250.06
FR	8	911.74	24	7104.27	28	11.32	7	800.51	24	259.13
HR	45	55.77	33	2222.13	46	3.05	32	73.2	53	25.27
IT	2	4617.21	6	25481.11	8	66.12	4	1404.04	18	349.96
PT	51	32.26	36	1889.34	25	13.64	43	41.33	39	71.98
RO	30	219.86	13	19480.61	34	8.1	16	320.59	28	197.81

Table 6: Top 10 countries (daily average of unique IP addresses per million subscribers)

#	**GameOver Zeus**		**Conficker**		**Morto**		**ZeroAccess**		**Spam**	
	CC	IPs	CC	IPs	CC	IPs	CC	IPs	CC	IPs
1	BY	718.88	VN	8049.81	TH	14.86	TR	176.3	KZ	444.67
2	UA	367.05	EG	7588.22	CY	13.52	ES	139.45	RU	410.58
3	IT	324.02	ID	6701.18	IL	12.19	BG	139.01	UA	270.47
4	ID	254.96	PK	5614.33	VN	11.69	TH	136.8	VN	254.75
5	ZA	237.27	BG	5351.89	ZA	10.98	VN	134.17	BY	248.23
6	KZ	190.06	RO	5079.69	TR	10.88	AR	126.7	MA	199.36
7	PE	187.94	MA	4331.4	SA	10.11	MY	122.24	CL	172.22
8	MY	179.05	RS	4312.42	BR	8.85	RS	104.54	AR	148.01
9	IL	178.08	AR	3780.51	EG	8.28	IT	98.53	PL	121.88
10	VN	168.8	MY	3713.63	TH	14.86	LT	90.58	ID	118.62

Table 7: Ranking of ACDC countries (daily average of unique IP addresses per million subscribers)

	GameOver Zeus		Conficker		Morto		ZeroAccess		Spam	
CC	Rank	IPs per	Rank	IPs per	Rank	IPs per	Rank	IPs per	Rank	IPs per
BE	48	19.46	47	349.36	25	3.27	43	17.29	58	11.14
DE	44	23.66	48	296.56	39	1.74	41	20.58	54	14.85
ES	46	21.26	23	1604.01	30	2.72	2	139.45	50	19.8
FR	41	35.52	49	276.81	56	0.44	31	31.19	59	10.1
HR	31	59.08	15	2354.03	26	3.23	16	77.54	39	26.77
IT	3	324.02	21	1788.15	22	4.64	9	98.53	42	24.56
PT	53	11.83	37	693.08	17	5.01	45	15.16	40	26.4
RO	32	57.33	6	5079.69	35	2.11	13	83.59	23	51.58

What we see is that ACDC member countries are starting out from very different positions when improving their botnet mitigation. For some countries, it is part of keeping infection levels at a low level – a level that was achieved through earlier efforts. For others it is part of reducing relatively high infection levels. Germany is a good example of the first, Italy of the second type.

This pattern is not exclusive to ACDC. If we look at a number of more mature anti-botnet initiatives, we see the same variation. Australia, Finland, Ireland, Japan and Korea are all known for their national initiatives [21] . In Table 9, we can see the ranking of these countries. Finland has long been the gold standard in mitigation – and this is confirmed in the best possible rank (60) in all sources, except for Morto (59). They started as early as 2006 and through this ongoing effort have managed to achieve consistently low infection levels. On the other side we find Japan and Korea, which started out with high infection rates and have been slowly but consistently improving.

These differences raise an underlying question: is there is an improvement over time associated with the NSCs? We have three data sources that cover multiple years. Table 10 presents the ranking of countries with an NSC over the period 2011-2014. For Germany, the country with the longest running NSC, we can see an improvement in Conficker and Morto and a fluctuating ranking in spam. The other countries show a mixed picture. Some, like Belgium, are by and large static in the ranking. Others improve in some area, and worsen in another.

Table 8 : Spearman correlation of country rankings for different botnets

	Morto	Conficker	GameOver Zeus
Conficker	0.57	-	-
GameOver Zeus	0.46	0.50	-
Spam	0.49	0.59	0.58

Table 9: Ranking of countries with an anti-botnet initiative (daily average of unique IP addresses per million subscribers)

	GameOver Zeus		Conficker		Morto		ZeroAccess		Spam	
CC	Rank	IPs per	Rank	IPs per	Rank	IPs per	Rank	IPs per	Rank	IPs per
AU	39	42.77	52	239.54	36	1.79	32	30.36	26	43.3
FI	60	2.26	60	13.83	59	0.00	59	0.97	60	4.99
IE	19	103.14	36	710.92	31	2.68	23	42.85	31	35.78
JP	17	110.66	51	249.13	58	0.14	33	25.47	43	23.15
KR	42	31.94	28	1235.09	41	1.64	49	13.22	14	89.35

We have to keep the limitations of this approach in mind. Only three of our sources allow for this longer-term view, unfortunately. Their peculiarities may interfere with the signal we are trying to distill from them, namely the effectiveness country-level mitigation. Still, it seems plausible that if the mitigation would make a drastic difference, it should be visible here as well.

In short, there is no clear trend in the ranking at the country level. We should emphasize that this does not mean there is no improvement. When we look at the number of infected machines seen in the data, then there are actually visible improvements.

In Figure 25 to Figure 29, we have plotted the daily average number of unique IP addresses seen in the data sources. In all our sources, we see the number of infected machines decrease. In Figure 30 *to* Figure 34, we present the same time trends, but then normalized for the size of the user population. There is a clear decay pattern across all countries – except for Morto, but this botnet has already been decimated, so the remaining infections are more or less static. In this sense, there is definitely progress.

Given the of our data sources, most of these downward trends are more or less inherent to the source: except for spam, they are all sinkholes – in other words, the capture data from botnets that have been wrestled out of the control of the attackers. It is normal for those botnets to follow a process of natural decay, as the bots slowly get cleaned up.

The reason that this data is still relevant to study the impact of mitigation, is that this decay should be faster in countries with more effective mitigation practices. If mitigation is in fact more effective in a country, this would be revealed via an improved ranking over time compared to the other countries. After all, the ranking expresses how well countries perform compared to each other. As we have seen, the rankings do not present evidence for clear improvements that can be associated with the NSCs. The general level of the rankings of countries with an NSC is not leaning one way other the other. Some of these countries are better than average (meaning: they have a rank number between 30 and 60), others are worse (rank number between 1 and 30). In fact, the average ranking of all of the ACDC partner countries is consistently around 30.

Table 10: Country ranking over time (daily average of unique IP addreses per million subscribers)

CC	**Conficker**				**Morto**				**Spam**			
	2011	2012	2013	2014	2011	2012	2013	2014	2011	2012	2013	2014
BE	48	49	47	47	17	24	31	25	53	59	59	57
DE	45	45	46	48	16	34	35	41	50	41	38	47
ES	29	27	26	24	31	32	33	30	42	23	16	30
FR	50	52	52	50	55	56	58	56	54	57	54	55
HR	24	21	15	15	14	35	13	27	24	29	31	38
IT	27	25	21	20	20	27	18	20	38	38	21	31
PT	37	36	36	37	15	14	9	9	30	30	24	29
RO	4	4	3	5	52	53	48	37	11	10	12	17
Avg.	*33*	*32*	*31*	*31*	*28*	*34*	*31*	*31*	*38*	*36*	*32*	*38*

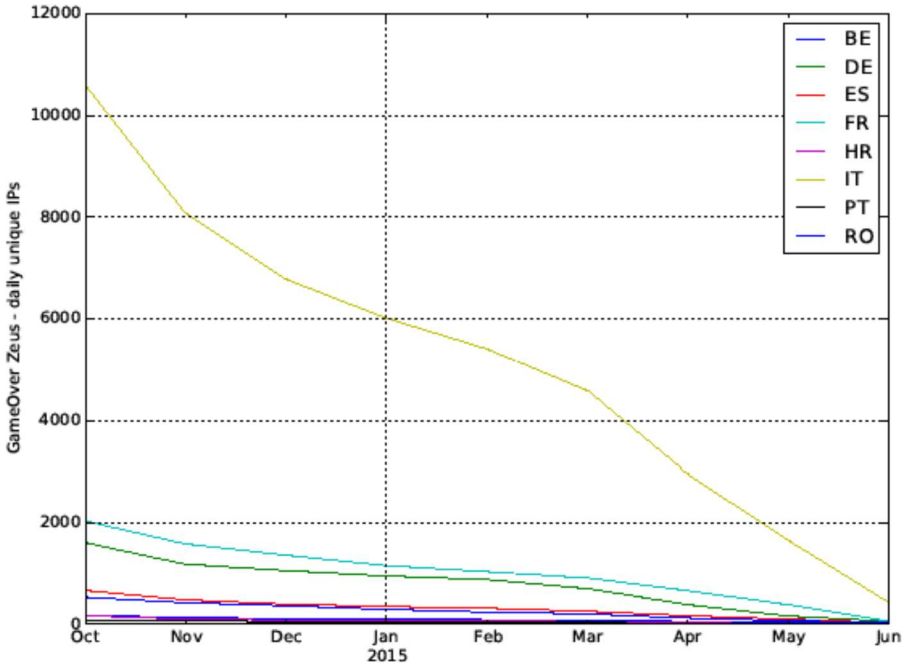

Figure 25: Daily average unique IP addresses for GameOver Zeus

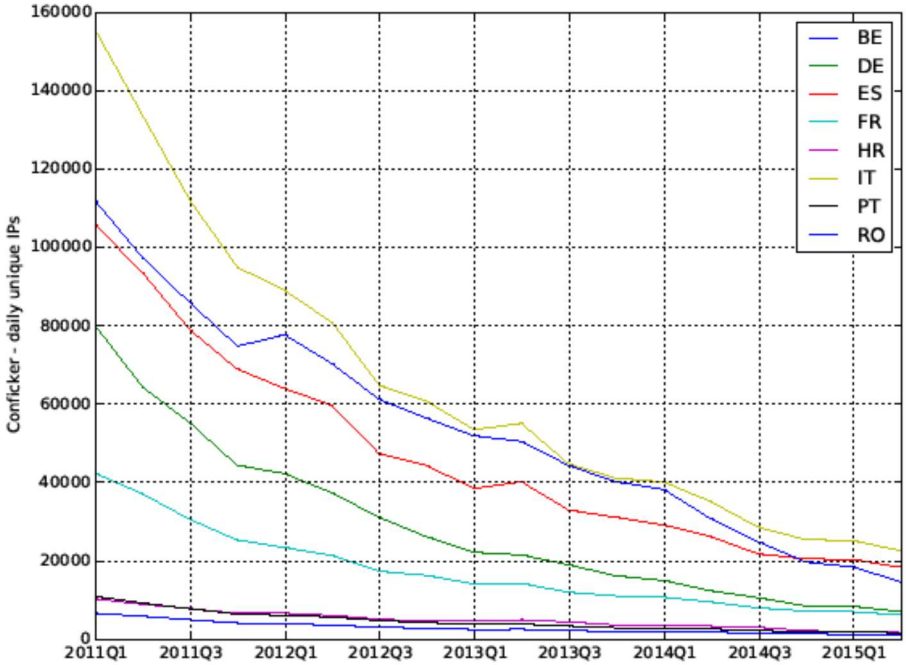

Figure 26: Daily average unique IP addresses for Conficker

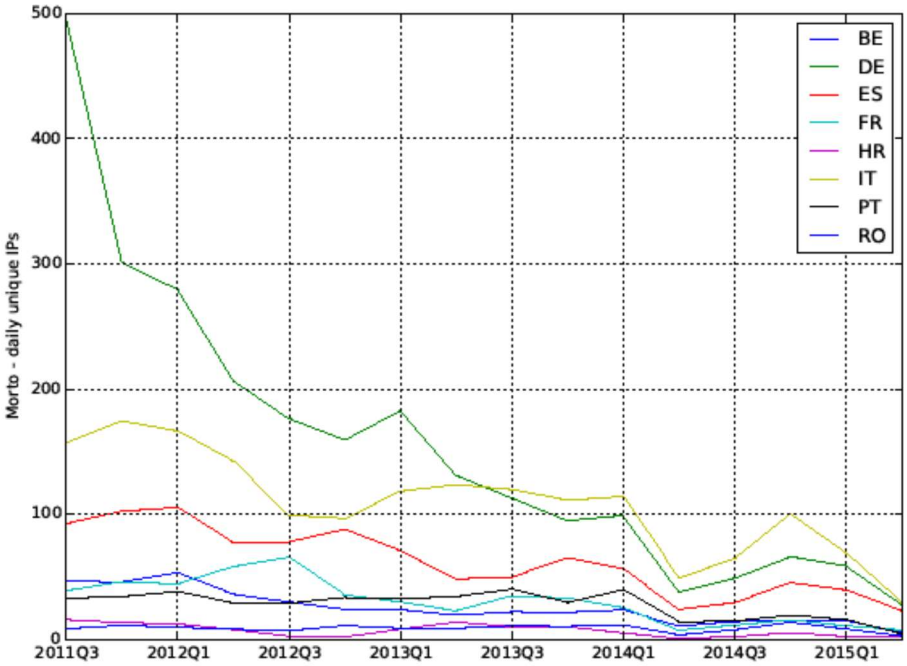

Figure 27 : Daily average unique IP addresses for Morto

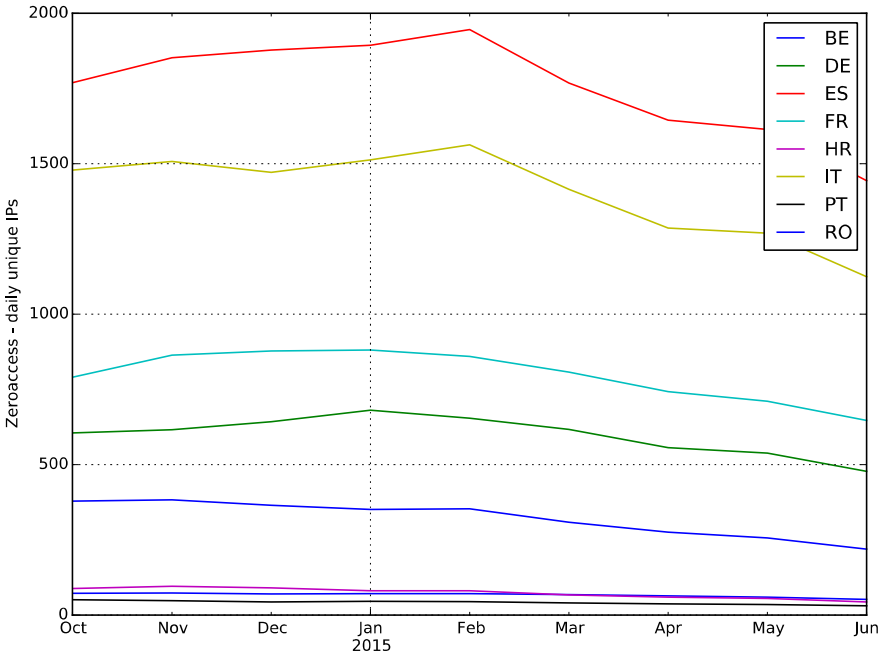

Figure 28: Daily average unique IP addresses for ZeroAccess

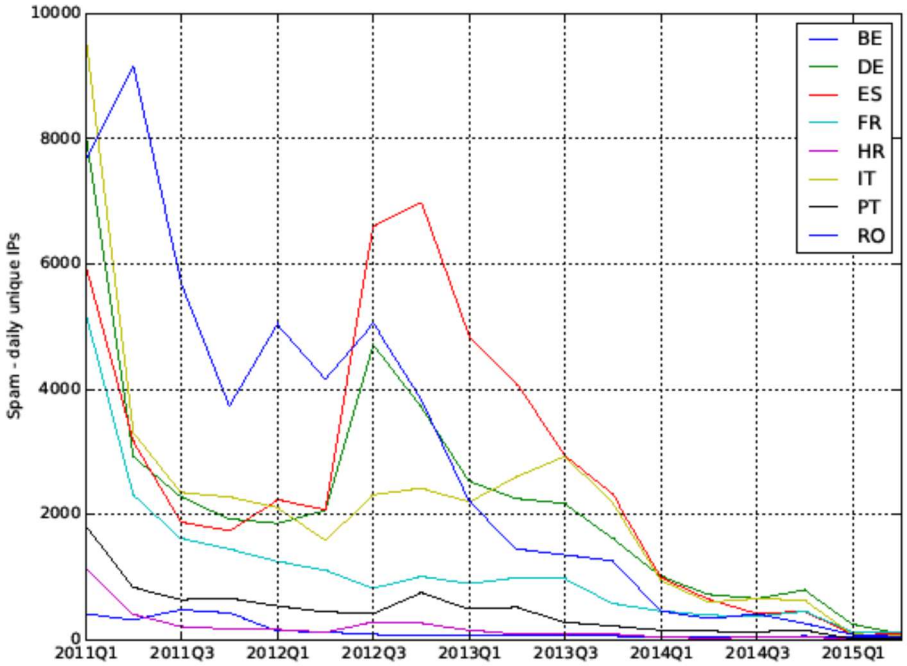

Figure 29: Daily average unique IP addresses for spam

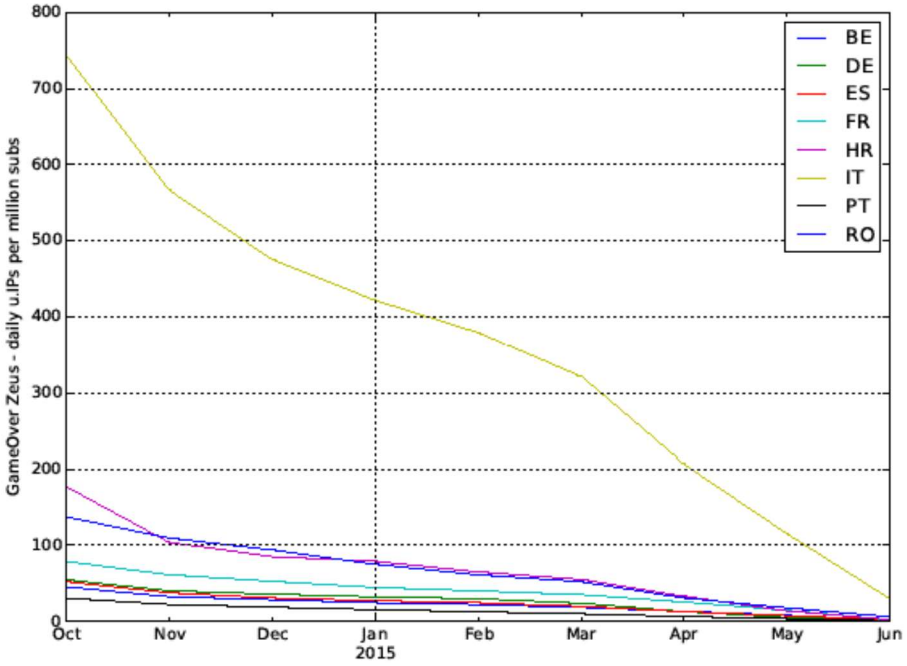

Figure 30: Daily average unique IP addresses per million subscribers for GameOver Zeus

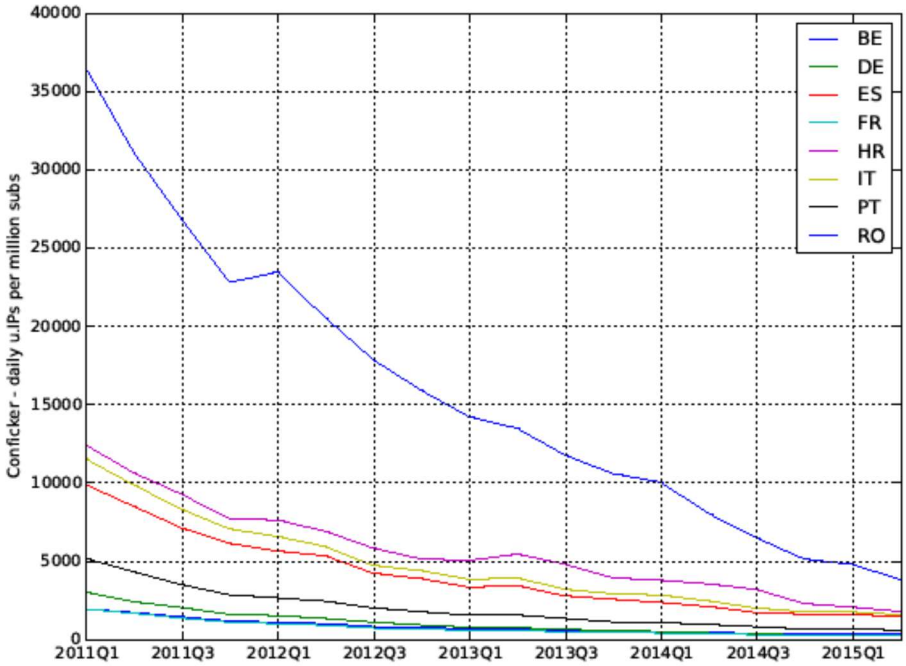

Figure 31: Daily average unique IP addresses per million subscribers for Conficker

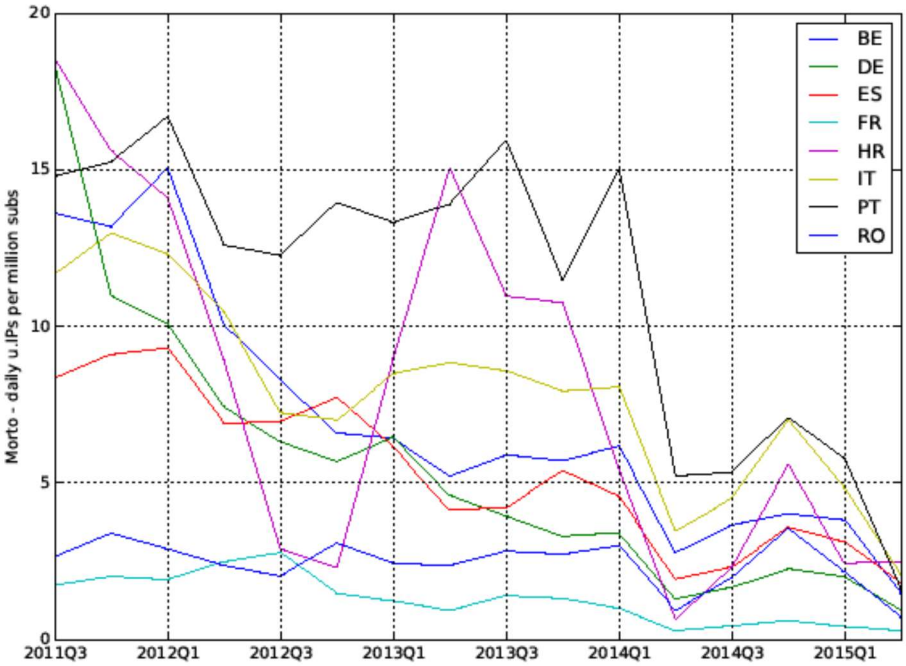

Figure 32: Daily average unique IP addresses per million subscribers for Morto

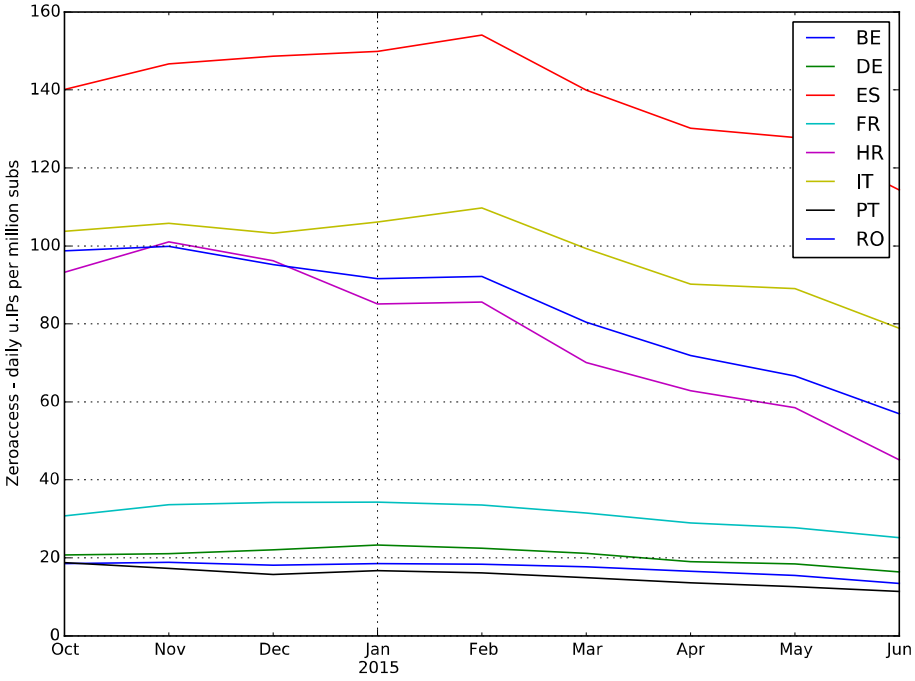

Figure 33: Daily average unique IP addresses per million subscribers for ZeroAccess

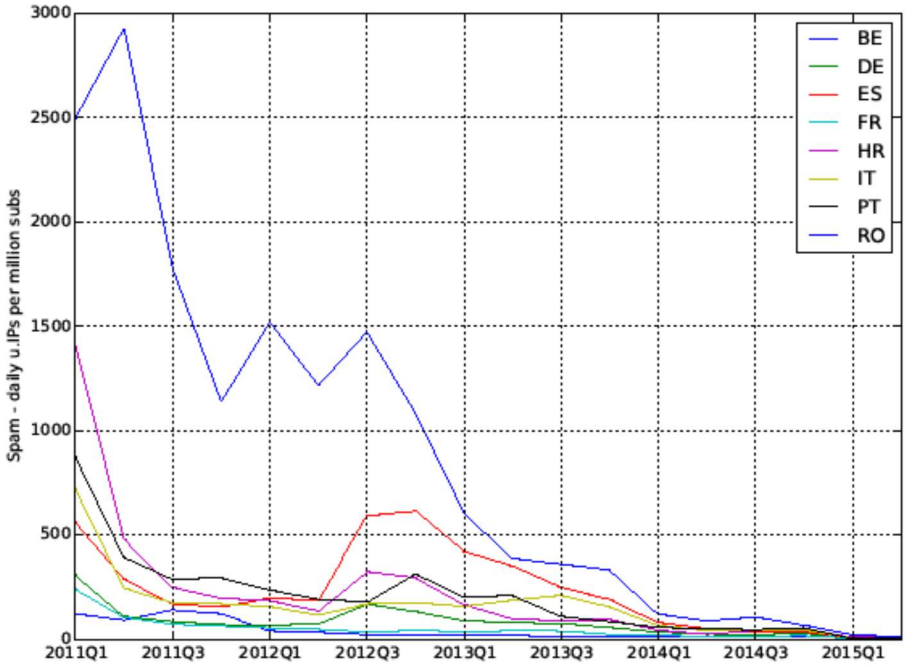

Figure 34: Daily average unique IP addresses per million subscribers for spam

5.4 *ISP-level analysis*

We now turn to the ISP-level analysis. As was explained earlier, there are several reasons for studying mitigation specifically at the level of ISPs. The country-level metrics are the most inclusive in terms of botnet incidents – it includes all incidents that are geo-located in that country. Prior research has shown, however, that the bulk of all infections reside in the networks of retail ISPs – typically around 80 percent. For this reason, improving mitigation by ISPs is a core component of any anti-botnet solution, including the NSCs. This has, in fact, been an important part of the services offered by the more developed NSCs: sending notifications to ISPs about infected customers.

An additional reason to study mitigation at this level is that the ISP-level metrics are based on more reliable data in terms of the number of users present in each network (from TeleGeography's GlobalComms database). In other words, the normalization of the infection rate can be done much more reliably (see Chapter 4 for more details). In short: an ISP level analysis provides us with a clearer, more focused analysis of mitigation, without scoping out too much of the overall problem.

We will look at individual ISPs later in this section, but first we compare all ISPs in a country to those the other countries. We calculate a weighted average of the ISP infection rates for each country by adding up all infections in the networks of the ISPs in that country and then dividing them by the total sum of subscribers of those ISPs. This weighted average for the ISP infection rate allows us to generate a ranking similar to those discussed in the previous section.
Table 11 and
Table 12 present the daily average of the number of IP addresses seen in the data for the top 10 most infected countries in each of the sources. There are no surprises here. The ranking for the ISP networks is similar as those for the countries as a whole. This confirms our earlier assumption that the bulk of these infections, typically around 60 to 80 percent, are in ISP networks.

Turning to the normalized rankings (Table 13 and Table 14), we see the same distribution across the ranking for the countries with an ACDC NSC. Notice that there is no relation with the type of NSC. Whereas the NSCs in DE, ES, HR and IT are all operationally involved in notifying ISPs about infections in their networks, their rankings are as diverse as those of the countries with NSCs that have a more limited set of services (BE, FR, PT, RO).

There is no clear time trend for most countries either (see Table 15). For the three sources that we can track over multiple years, the rankings are quite stable. Germany improves its ranking for Conficker and Morto, while spam stays the same up until 2014, but then also improves in 2015 – as illustrated by rank (58) in Table 14. Other countries, like Belgium, are static in terms of ranking or have a mix of modest improvements and deteriorations.

In sum: we do not see evidence in the rankings for an improvement path being associated with the NSCs. We have to re-iterate, however, that the rankings only capture the performance relative to other countries. As can be seen in the time trends (Figure 35 to Figure 39), the size of the infected population as seen in our sources is decreasing. This is a combination of cleanup and the fact that we are using mostly sinkhole data – see the previous section for more details.

Table 11: Top 10 countries (daily average of unique IP addresses in ISP networks)

#	GameOver Zeus	Conficker	Morto	ZeroAccess	Spam

	CC	IPs	CC	IPs	CC	IPs	CC	IPs	CC	IPs
1	IT	3766.13	CN	111080.5	CN	191.6	ES	1624.25	RU	4051.69
2	JP	3220.71	VN	38090.71	BR	178.36	US	1584.32	CN	3036.82
3	GB	1492.42	BR	31673.21	US	79.01	TR	1413.09	US	1688.07
4	BY	1370.09	ID	26432.73	TR	77.17	IT	1215.22	KR	1515.57
5	US	1254.28	KR	20689.91	TH	67.55	BR	779.44	VN	1220.08
6	VN	828.01	IT	20453.48	IT	56.56	JP	759.66	KZ	708.41
7	ID	822.79	EG	20299.25	VN	48.26	FR	753.37	JP	699.43
8	TR	782.91	ES	17589.31	DE	38.57	VN	664.38	AR	656.23
9	UA	779.86	AR	16989.59	ES	30.68	TH	645.5	BR	603.1
10	FR	773.11	PK	16611.08	AR	27.62	AR	621.91	BY	380.03

Table 12: Ranking of ACDC countries (daily average of unique IP addresses in ISP networks)

	GameOver Zeus		**Conficker**		**Morto**		**ZeroAccess**		**Spam**	
CC	#	IPs	#	IPs	#	IPs	#	IPs	#	IPs
BE	41	58.57	38	1153.64	20	12.49	31	56.94	43	25.64
DE	12	590.44	19	7209.2	8	38.57	11	518.46	28	137.91
ES	22	237.95	8	17589.31	9	30.68	1	1624.25	24	169.8
FR	10	773.11	21	6021.55	25	10.1	7	753.37	31	94.67
HR	43	35.92	33	1662.54	46	1.73	33	47.26	50	14.08
IT	1	3766.13	6	20453.48	6	56.56	4	1215.22	18	235.23
PT	48	28.28	32	1704.35	19	13.56	38	39.39	34	62.91
RO	29	176.48	12	13090.86	33	6.11	16	278.96	30	121.89

Table 13: Top 10 countries (daily average of unique IP addresses per million subscribers in ISP networks)

#	**GameOver Zeus**		**Conficker**		**Morto**		**ZeroAccess**		**Spam**	
	CC	IPs	CC	IPs	CC	IPs	CC	IPs	CC	#
1	BY	639.3	EG	6869.46	CY	13.19	TR	164.03	KZ	349.49
2	IT	264.29	VN	6661.63	TH	12.41	ES	128.6	VN	213.38
3	PE	181.12	ID	5034.85	IL	10.82	TH	118.61	BY	177.33
4	ID	156.72	PK	4051.48	TR	8.96	VN	116.19	RU	158.58
5	VN	144.81	MA	3793.5	VN	8.44	MY	106.55	MA	158.51
6	KZ	141.86	RO	3413.52	SA	7.99	AR	103.65	CL	137.15
7	GR	141.24	MY	3125.3	BR	7.54	GR	86.88	AR	109.37
8	MY	140.92	AR	2831.6	EG	7.47	IT	85.28	IL	91.24
9	CY	133.2	CY	2756.27	GR	5.84	RS	81.63	MY	84.19
10	UA	121.1	RS	2717.97	MY	5.7	PE	81.24	PE	77.54

Table 14: Ranking of ACDC countries (daily average of unique IP addresses per million subscribers in ISP networks)

	GameOver Zeus		**Conficker**		**Morto**		**ZeroAccess**		**Spam**	
CC	#	IPs	#	IPs	#	IPs	#	IPs	#	IPs
BE	46	15.17	46	298.87	18	3.24	40	14.75	53	6.64
DE	43	20.24	48	247.08	36	1.32	36	17.77	58	4.73
ES	44	18.84	20	1392.66	22	2.43	2	128.6	47	13.44
FR	40	30.12	49	234.62	51	0.39	28	29.35	59	3.69
HR	34	38.05	14	1761.23	25	1.83	17	50.07	42	14.92

IT	2	264.29	18	1435.33	17	3.97	8	85.28	37	16.51
PT	51	10.37	29	625.22	13	4.97	41	14.45	32	23.08
RO	29	46.02	6	3413.52	27	1.59	12	72.74	22	31.78

Table 15: Country ranking over time (daily average of unique IP addreses per million subscribers in ISP networks)

CC	**Conficker**				**Morto**				**Spam**			
	2011	2012	2013	2014	2011	2012	2013	2014	2011	2012	2013	2014
BE	47	45	47	46	16	19	22	17	53	59	59	58
DE	41	42	45	47	17	28	30	33	49	38	31	50
ES	26	21	22	21	23	24	24	22	32	16	12	21
FR	46	47	48	48	55	57	53	54	51	58	56	54
HR	23	19	17	14	13	30	13	25	22	23	30	39
IT	24	22	20	20	18	21	16	15	30	33	18	26
PT	36	33	33	30	15	13	7	8	23	26	22	19
RO	6	3	4	5	52	50	39	31	10	11	10	16
Avg.	*31*	*29*	*30*	*29*	*26*	*30*	*26*	*26*	*34*	*33*	*30*	*35*

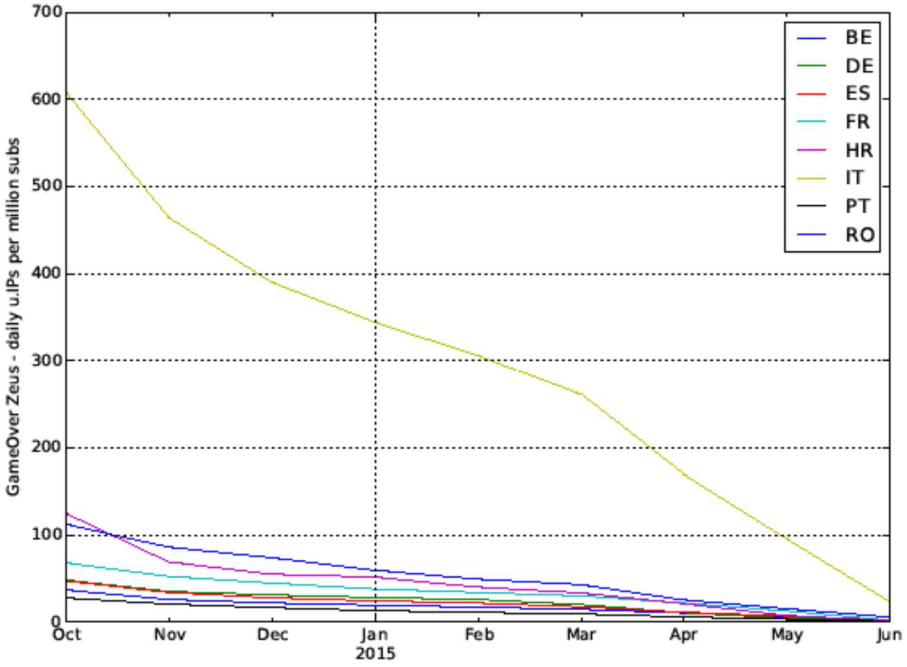

Figure 35: ISP infection rates for GameOver Zeus (daily average of unique IP addresses per million subscribers in ISP networks)

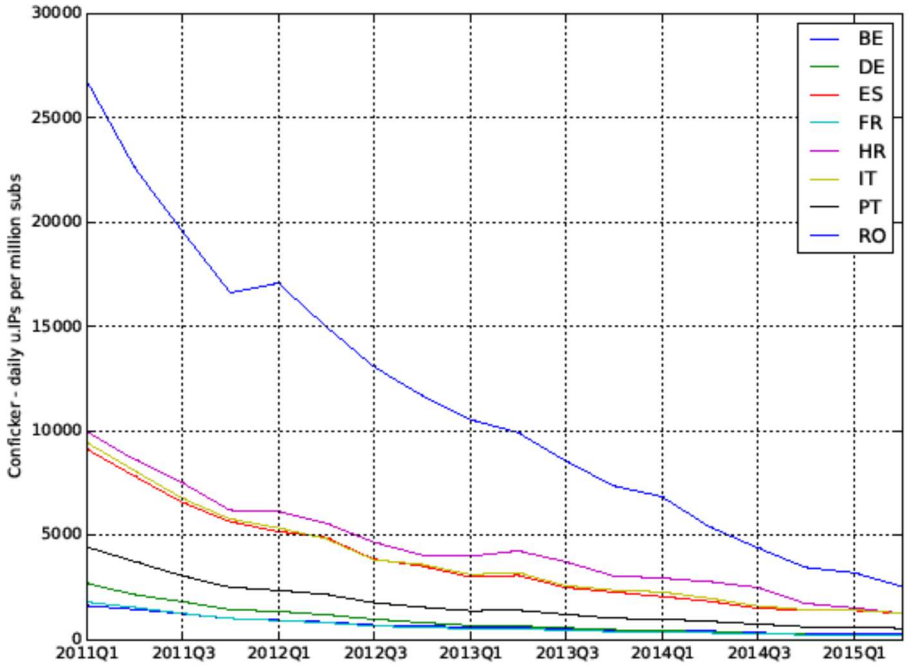

Figure 36: ISP infection rates for Conficker (daily average of unique IP addresses per million subscribers in ISP networks)

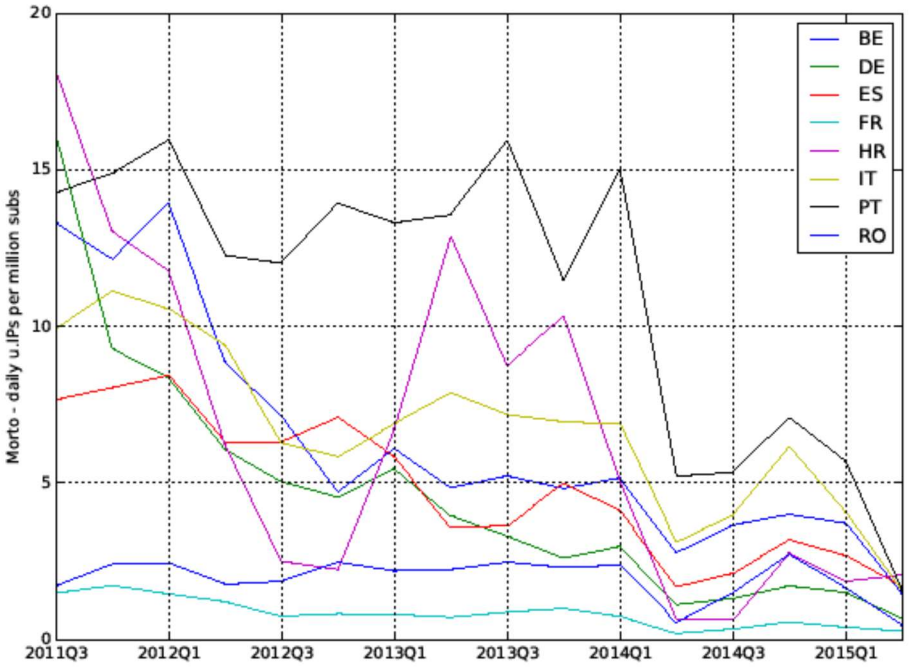

Figure 37: ISP infection rates for Morto (daily average of unique IP addresses per million subscribers in ISP networks)

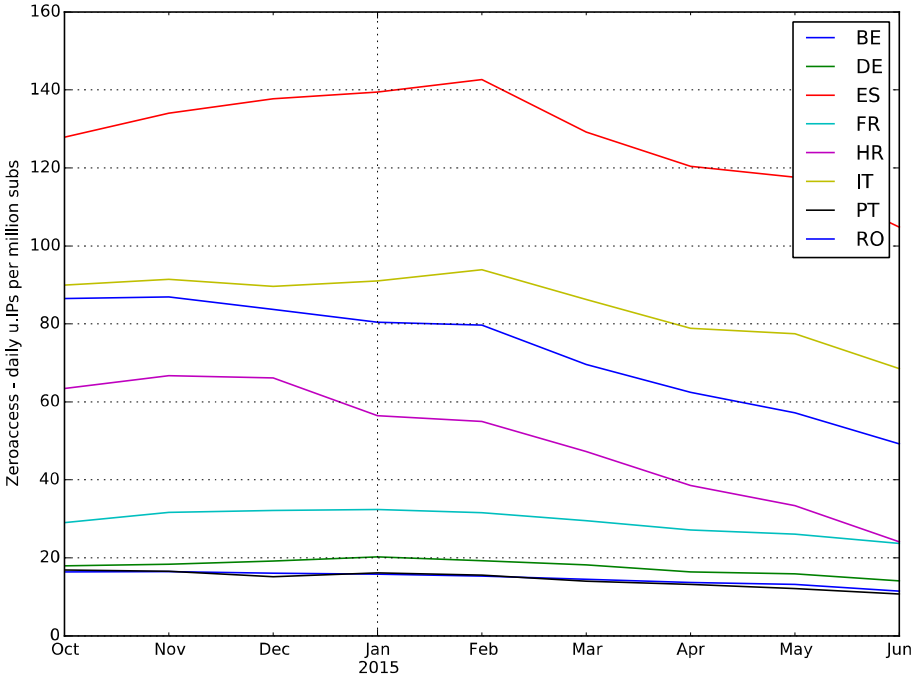

Figure 38: ISP infection rates for ZeroAccess (daily average of unique IP addresses per million subscribers in ISP networks)

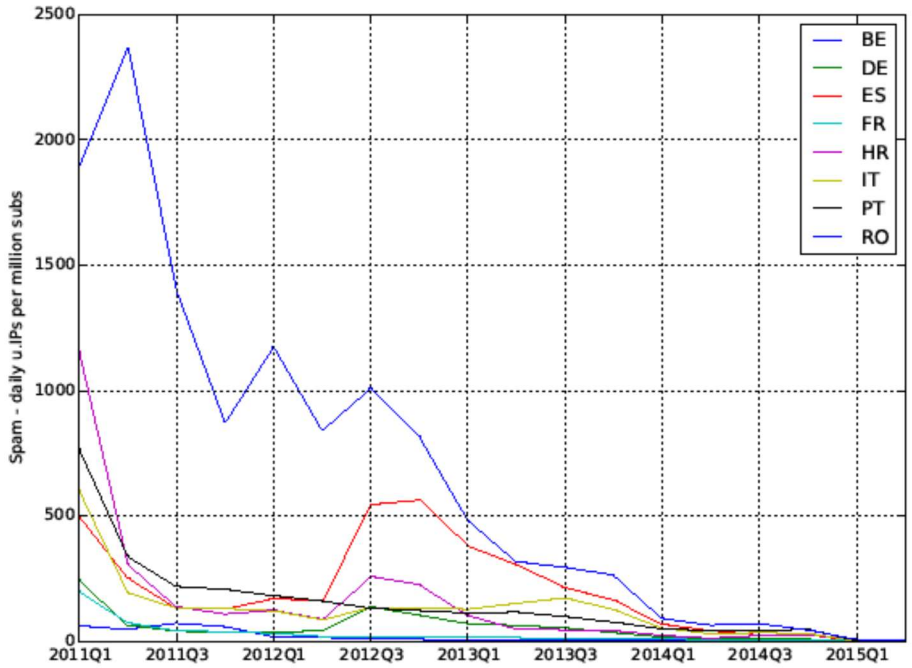

Figure 39: ISP infection rates for spam (daily average of unique IP addresses per million subscribers in ISP networks)

5.5 *Performance of individual ISPs*

The analysis in the previous section looked at the average infection rates for the ISPs in a country. Such averages can obscure the underlying variance among the ISPs. The average might be driven by one or two ISPs with the most subscribers and the other ISPs in the country might perform much better or worse. In earlier work, we found that even within one country, the infection rates of ISPs can differ by one, sometimes two, orders of magnitude [1]. Metrics based on averages or medians cannot accurately convey such variance.

In this section, we will go beyond country averages and look instead at the performance of individual ISPs. The ISPs will remain unnamed, as the goal is not to evaluate company practices, but rather to understand better the overall landscape of ISPs and the national markets in particular.

In Figure 40 to Figure 49, we have plotted the infection rates of 265 ISPs in the 60 countries included in the analysis. Together, these ISPs make up over 80 percent of the broadband markets in these countries. Each dot represents an ISP. The red dots are ISPs in a country with an ACDC NSC. The horizontal axis is the size of each ISP as measured by the number of subscribers, in log scale. The vertical axis shows the average number of unique IP addresses seen in the data source, also in log scale. In other words, the distance between any two grid lines represents a difference of a factor of 10 – horizontally as well as vertically.

The diagonal line is a regression line that captures the linear relationship between the size of the ISP and the number of infected machines. As was mentioned above, the infection rate is to some extent simply a function of the number of subscribers in a network. More computers means more infections. The regression line can be interpreted as the average infection rate of ISPs across all sizes. Some ISPs are vertically above the line, which means have more infections than an average ISP of that size. Some are below, which means fewer infections than average.

We have made two plots for each data source: one with all ISPs (in blue and red) and one with only the ISPs in a country with an ACDC NSC, color coded per country. The latter allows us to see the variance within each country.

The first thing that we can see across all figures is that the overall variance among ISPs of similar sizes spans several orders of magnitude – often around 3 orders, in fact. In other words, even though the ISPs are of a similar size, one has a thousand times more infections than the other. The ISPs in the countries with an ACDC NSC are distributed across almost this whole variance in all data sources except spam. So there are large differences among those ISPs as well.

In the ACDC-only plots, we can see the distribution of ISPs within each country. We expected very large variances here, and there is indeed some variance here and there, but remarkably often, the ISPs in a country are often clustered together. Take a look at the plots for GameOver Zeus and Conficker, for example (Figure 41 and Figure 43, respectively). We can see that all Italian ISPs struggle with Zeus and they all have a similar distance above the regression line, which suggests that they perform similarly poorer than the average: one order of magnitude poorer, in fact. This means that they have 10 times more infections than the average ISP of their size.

A same type of clustering holds most other countries. Take France and Spain. French ISPs are all more or less on top of the line, except for one, indicating average infection rates. Spanish ISPs are all just below the line, which indicates slightly better than average performance. For Conficker, where Romanian ISPs perform worse than average, they all perform similarly worse. Only German ISPs have a consistently large variance, with some ISPs showing up above the line and some far below it. Belgian ISPs also have considerable variance.

We see in these distribution why the ACDC countries, taken together, have an average performance in terms of their rankings. Some are mostly above the line, some close to it and some mostly below it. Only for spam do nearly all ISPs in countries with an ACDC NSC perform better than average (Figure 49). This was also clear from the rankings we discussed in the previous section.

The surprising conclusion is that ISP infection rates within a country are closer together than expected. Our country rankings were quite representative after all. The performance of ISPs in a country is remarkably similar in many countries. We think this points to the fact that there are institutional forces at work. Perhaps the awareness of ISPs to undertake mitigation is something that is being fostered at the national level. This fits with our finding that active regulators are associated with lower infection rates. Left to their own devices, ISPs might orient themselves on what they peers and competitors are doing. If most do not invest in mitigation, then none of them do. If some of them do, then more might do it. The national anti-botnet initiatives might mobilize the ISPs and raise their awareness and willingness to act. Setting up a center can support this process and it can also make it easier for ISPs to start taking in infection reports, the first step towards enhanced mitigation. But this is not a given. The existence of an NSCs is not enough. They can only nudge ISPs in the right direction. The ISPs have to decide themselves that they want to increase mitigation. This suggests that the NSCs might want to work with the national regulator and CERT community to mobilize their ISPs. That way, the market as a whole can start to move in the direction of improvement.

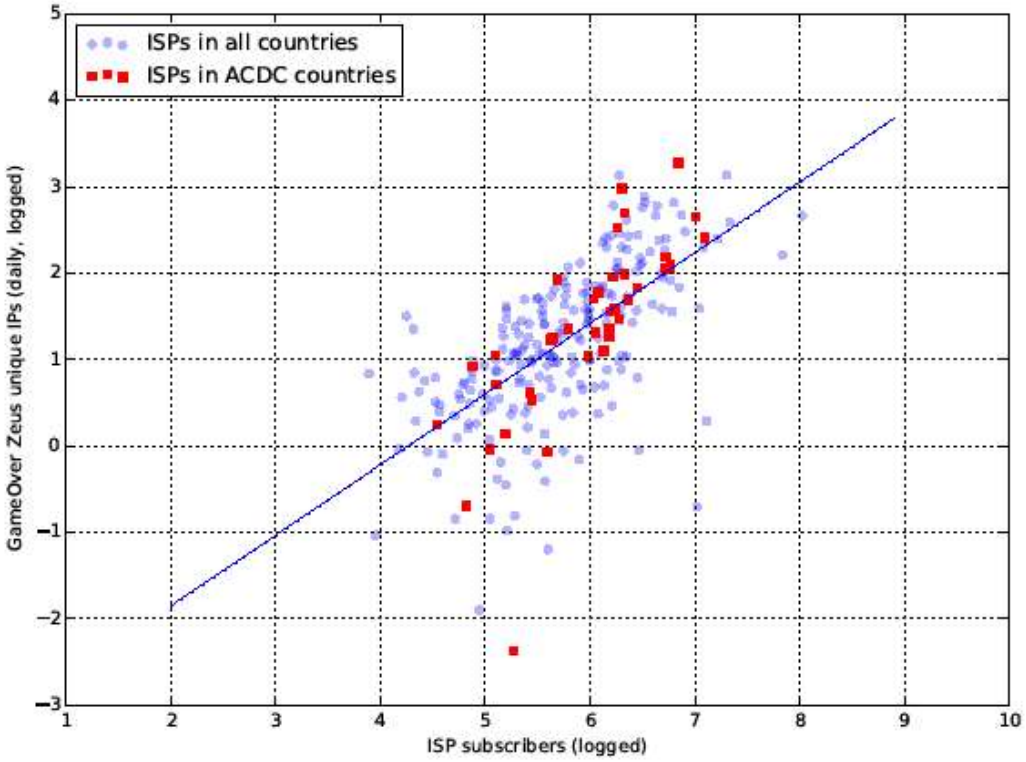

Figure 40: GameOver Zeus infections in ISP networks against number of subscribers

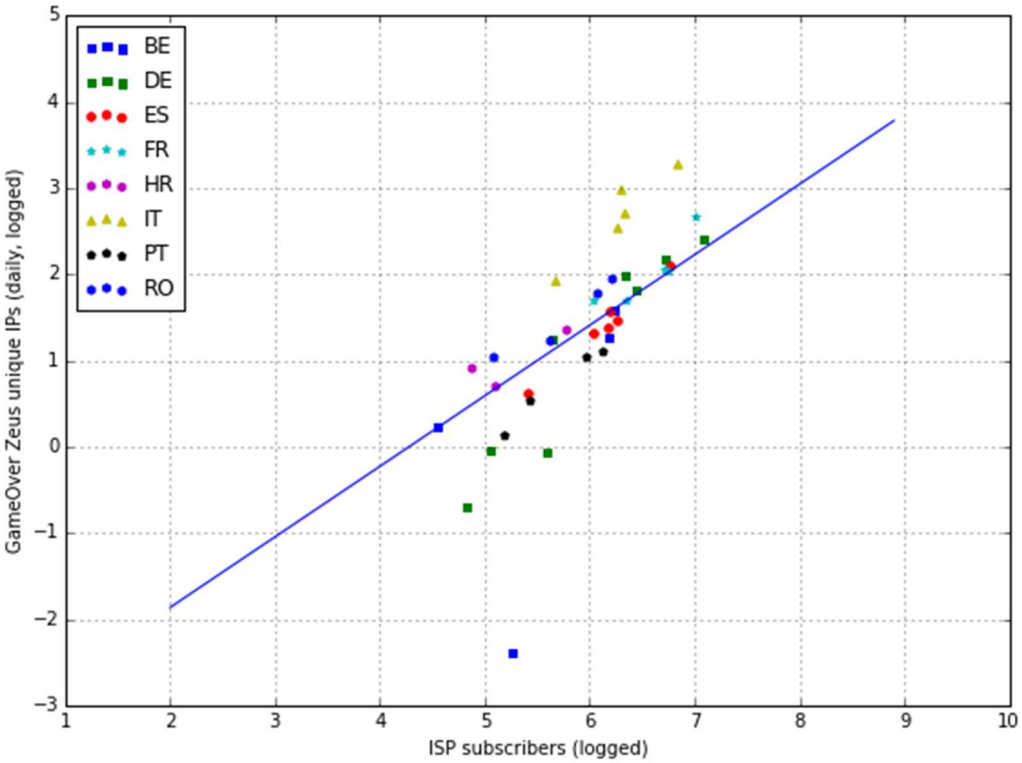

Figure 41: GameOver Zeus infections in ISP networks against number of subscribers (ACDC only)

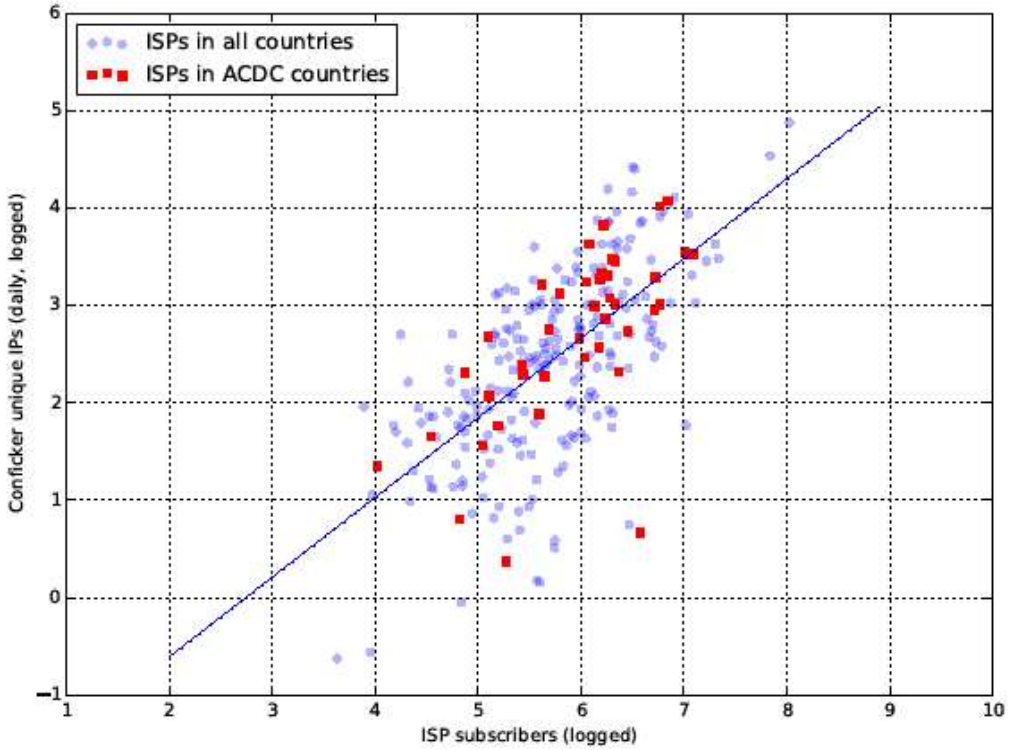

Figure 42: Conficker infections in ISP networks against number of subscribers

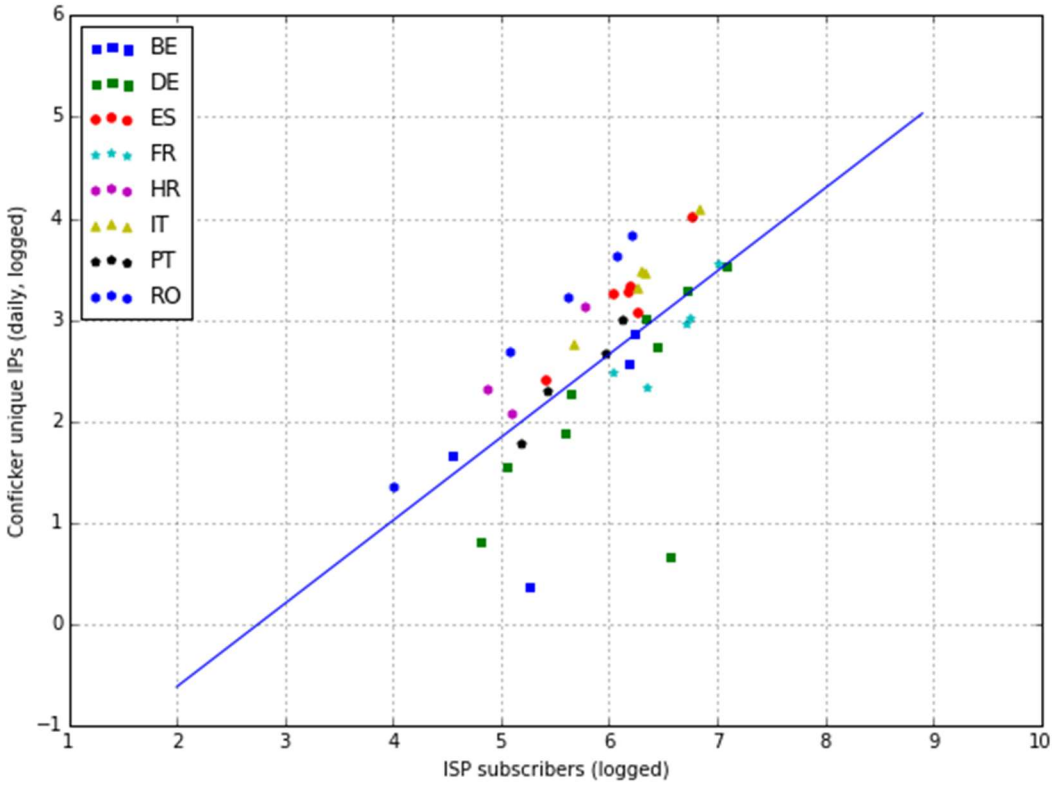

Figure 43: Conficker infections in ISP networks against number of subscribers (ACDC only)

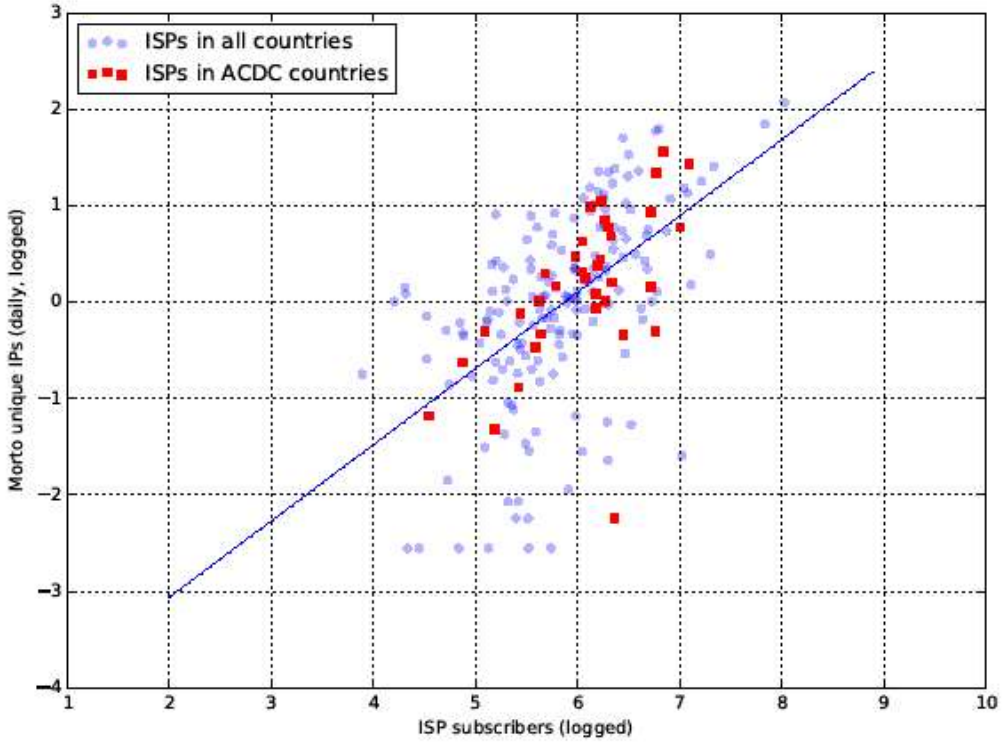

Figure 44: Morto infections in ISP networks against number of subscribers

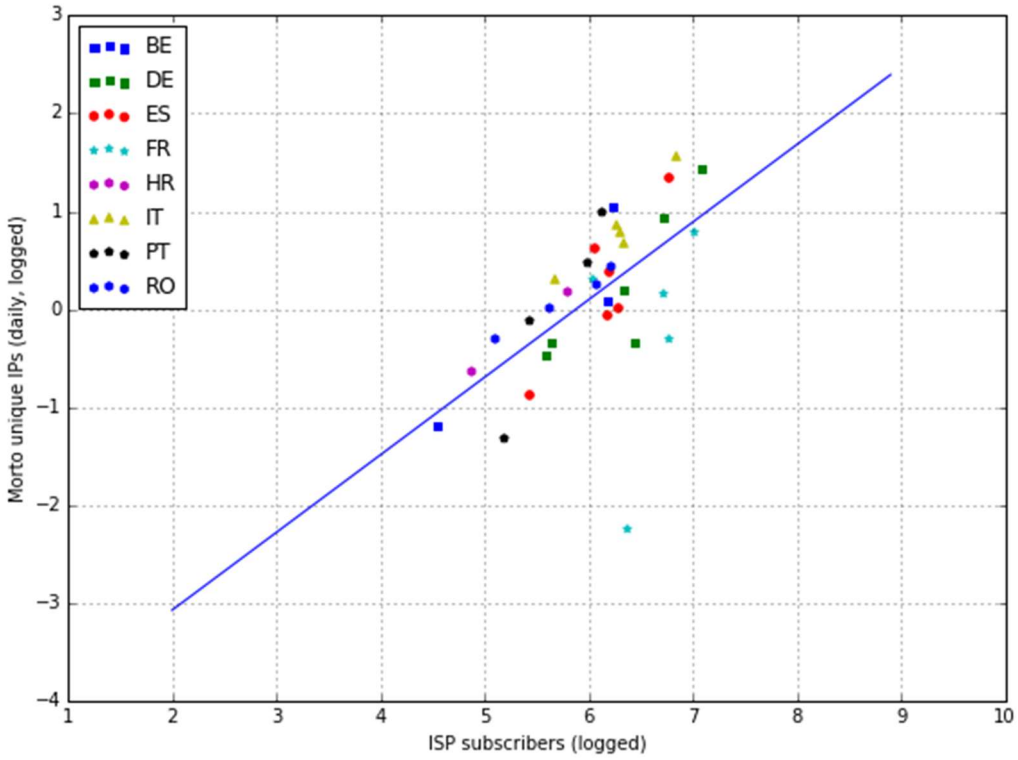

Figure 45: Morto infections in ISP networks against number of subscribers (ACDC only)

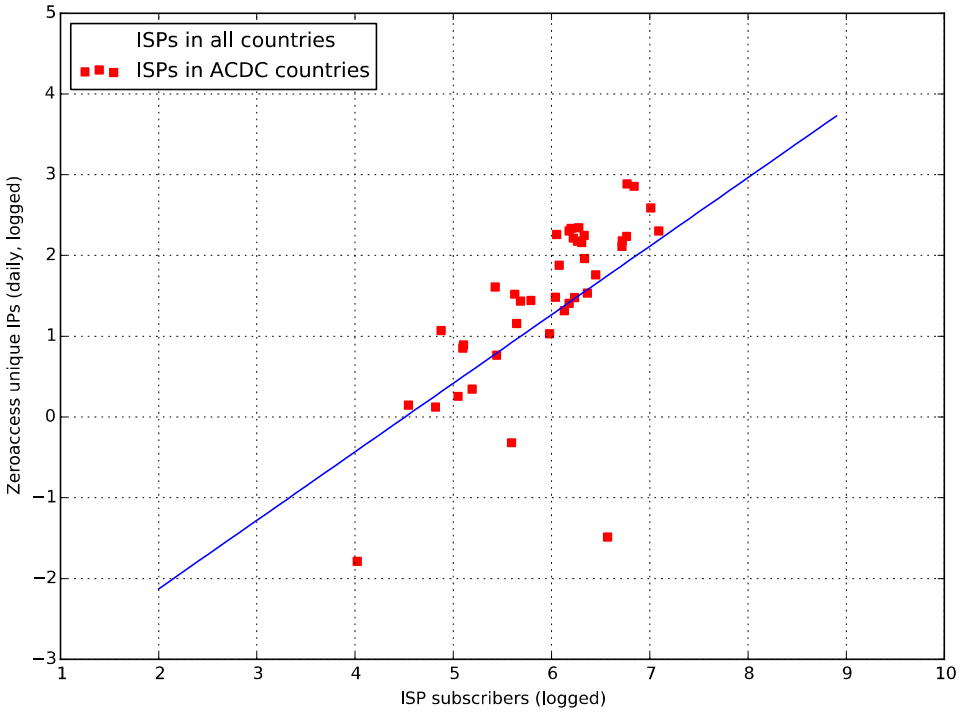

Figure 46: ZeroAccess infections in ISP networks against number of subscribers

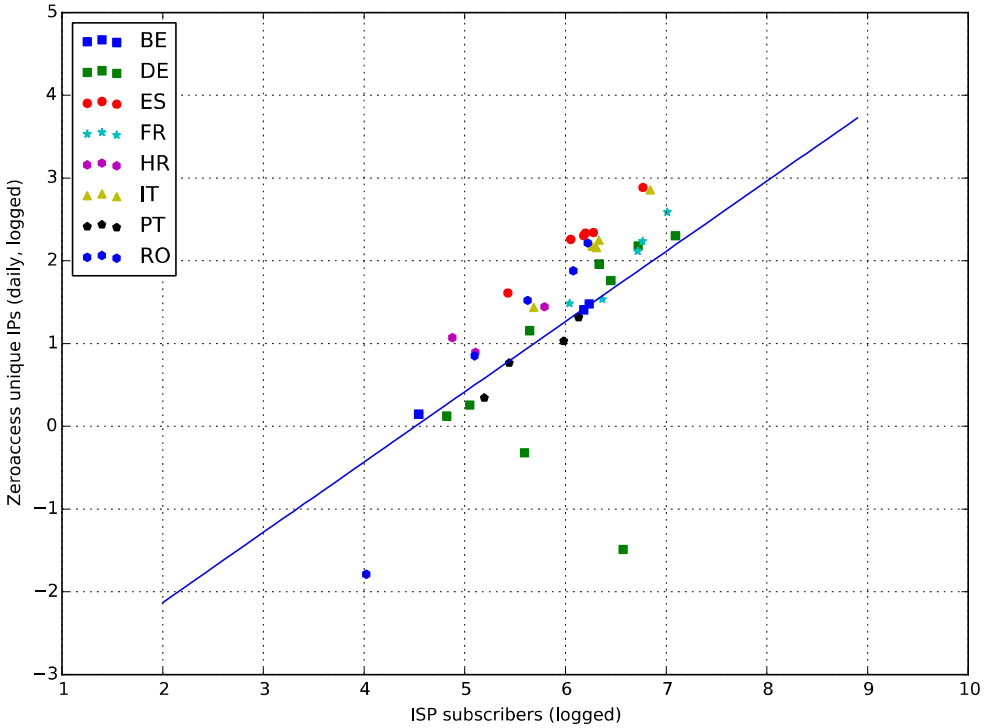

Figure 47: ZeroAccess infections in ISP networks against number of subscribers (ACDC only)

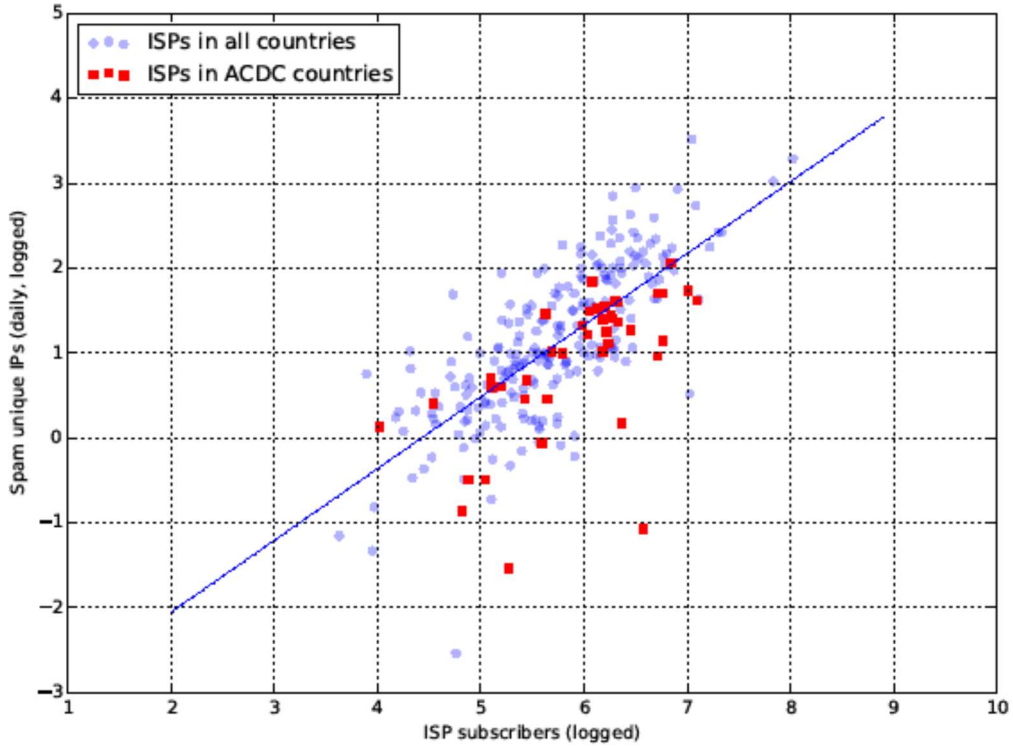

Figure 48: Spam infections in ISP networks against number of subscribers

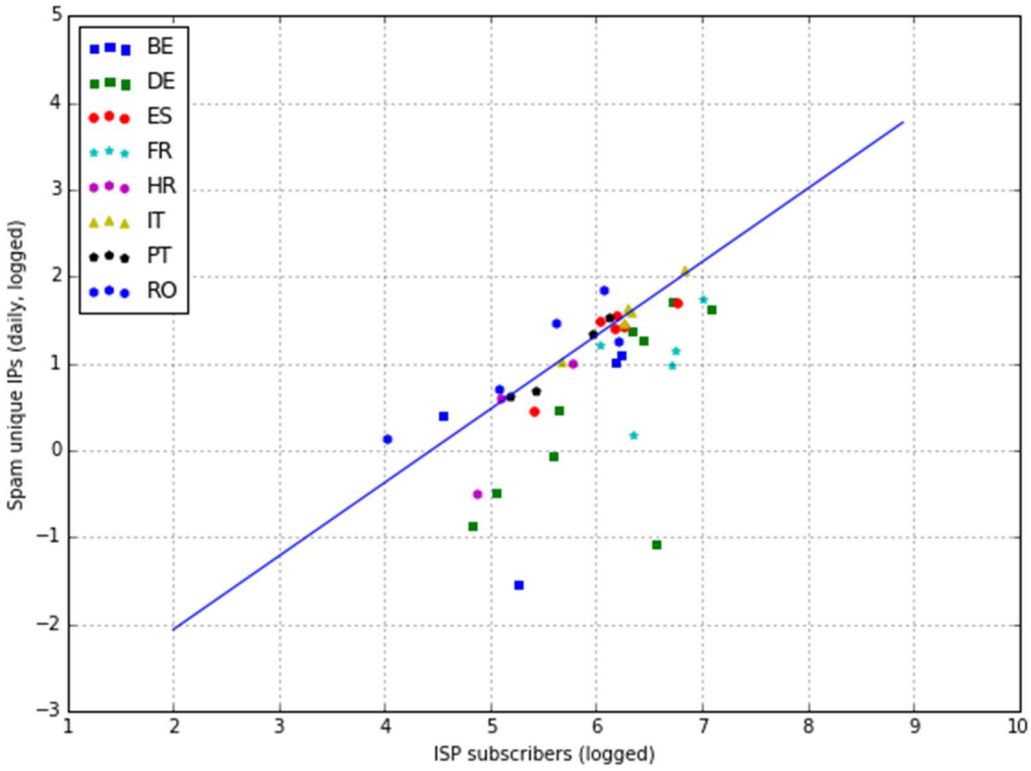

Figure 49: Spam infections in ISP networks against number of subscribers (ACDC only)

5.6 *Interpreting the main results*

Summarizing the main results from this analysis, we can state that the countries with an ACDC NSC perform very different, as do the ISPs in those countries. They are similarly distributed as the other ISPs and countries. Some have infection rates that are substantially better (i.e., lower) than average. Others are average or worse than average. Neither did we find clear evidence for substantial improvements over time. The countries and ISPs that did well maintained their position. The ones who struggled with higher infection rates, continue to struggle with them. There does not seem to be a relationship with the presence of an NSC, nor with whether the NSC is operationally involved in mitigation.

The crude lesson from these findings would be that the mere presence of an NSC does not improve mitigation. This would be a rather uninformative conclusion, however. Given the short time span in which NSC have been active, it is unrealistic to expect a measureable impact on infection levels. This is why we explored whether the setup of an NSC is part of an ongoing process of improvement in mitigation. That did not turn out to be the case.

Even in Germany and Spain, countries that performed well, and where the NSC is doing active notifications to the ISPs, we saw no acceleration in cleanup. Perhaps it is too early to tell. That could very well be. However, Germany has had an active center for around five years now. Perhaps it has accelerated cleanup in ways that our measurements did not capture, but judging by our data this acceleration did not take place.

How can this result be understood? There is another explanation we did not yet look at: scale. It is good to notify infected end users, but the scale at which this takes place matters greatly if it really has to have real impact on infection levels. In D1.3.2 ('Specification of Tool Group "Support Centre"'), we can find more details about the scale of notifications. The German NSC has processed over 250.000 customer tickets and almost 10.000 support calls. The different malware cleaners have been downloaded over 2.500.000 times since the launch of the service. This is in the course of around 54 months of operation. In other words: this means around 4600 tickets per month. How to interpret this number? Well, according to Microsoft, and our own analysis, at any moment in time around one percent of all Windows PCs are infected [MSIR [22] and our 2011 botnet study]. Germany has around 29,000,000 broadband subscribers (see Annex I). If one percent of those have an infection at any point in time, this translates to 290,000 infections. Suddenly, the number of tickets (4600 per month) does not seem so substantial anymore.

For the Spanish NSC, we can make a similar analysis. It has reported receiving reliable data on more than 5,000,000 incidents per day, which affect more than 350,000 unique IP addresses per day in Spain. Since November 2014, the NSC runs a pilot with Telefónica, the main ISP. In that period, the NSC has notified Telefónica about 957.470 unique IPs. Telefónica has sent 82.400 notifications to 16.627 different end-users. Each month 1.450 new customers are notified. These efforts are substantial and we do not want to undervalue their importance. It is also important, however, to keep in mind that one percent of Spain's broadband subscribers amounts to 126,300. That is an order of magnitude higher than the total number of subscribers that have been contacted in around eight months.

In sum: when interpreting the findings of our evaluation, it seems that setting up the NSCs are important and necessary steps, but real impact requires scaling up these efforts well beyond their current levels.

We want to re-iterate that these findings do not mean that cleanup is not happening. In fact, in most of our data sources we see a decreasing number of bots. This is partially due to the nature of the data source. Most of them are sinkholes. These are botnets who have been taken over by defenders and

are no longer under control of the attackers. They typically follow a slow process of decay, where more and more bots get cleaned up over time. So cleanup is happening in all countries. We looked mostly at the speed of cleanup. The rankings we developed and presented throughout this chapter capture whether a country or ISP is faster in cleanup than its peers. That would be visible through an improved ranking over time.

The key to improved mitigation is in the hands of the ISPs. We expected to find large variance among ISPs and we did, in fact, find such variance. ISPs of similar size can differ by two or even three order of magnitudes in the number of infected machines in their networks. Given such variance, perhaps we suspected that the country averages we used in the rankings are not that informative. It is not the country that drives the performance, but individual companies.

We tested this idea by looking at the distribution of ISPs in each of the countries with an NSC. Surprisingly, in most countries the ISPs perform at similar levels – as evidenced by the fact that they are clustered at similar distances from the regression lines in Figure 41, Figure 43, Figure 45, and Figure 49. There was not much less variance than we expected. This seems to suggest that ISPs' mitigation policies are less idiosyncratic and are in fact guided by what the peers/competitors in their own country are doing, as well as by the attention and pressures of public actors (see also [23])

To put it differently: at the national level, ISP seem to move as a pack. From the perspective of national and European policies, this is a relevant finding. A NSC might not be able to change anything by its mere presence, but it can get real traction if some of the leading ISPs commit to acting on infection notifications sent by the NSC. Our measurements suggest that the rest of the market will follow. This way a country could move to a consistently lower infection rates – i.e., higher rank – and reduce the societal damage cost by botnets.

6 Conclusions

The overarching question for the evaluation of ACDC is to what extent the end-to-end approach has been realized and has been effective. To answer this question, we have conceptualized the end-to-end approach as a flow of inputs to outputs to outcomes (see Figure 2, p. 8). We focused the evaluation of ACDC on a core output and outcome:

- Output: data that the CCH is receiving from the sensors and experiments and that can be disseminated to various stakeholders to improve mitigation
- Outcome: the impact of the National Support Centers on infection rates in their countries.

We will briefly summarize the main conclusions for each evaluation.

6.1 *Evaluation of data quality*

First, we evaluated the quality of the data (output) that ACDC is contributing to the fight against botnets. Does it enable better mitigation? The short answer is: yes.

There were a few limitations we need to re-iterate, most notably the fact that we worked on aggregated quality metrics, as privacy considerations did not allow full visibility into the data feeds. Keeping these issues in mind, we can say that the volume of events being submitted each data is consistently around 500,000. This is a non-trivial amount, in the same range as major botnet sinkholes run by, for example, ShadowServer. Furthermore, the data that the ACDC sensors and experiments submit to the CCH seems to be relatively unique. There is likely little overlap with the other data feeds that ISPs and CERTS are receiving from organizations like ShadowServer and Spamhaus. This means he CCH can make an important contribution to enhanced mitigation by ISPs, CERTS and others.

The statistical quality of the data is reasonable. It contains some gaps and anomalies, but auto-regressive analysis suggests consistency over time for many of the feeds going into the CCH. The submitters give most categories relatively high confidence levels, suggestion they contain a tolerable amount of false positives. Coverage of the data does not seem biased in terms of geography or AS. This further underlines the usefulness of the data for improving mitigation in the countries with ACDC partners.

The data quality also allows for another function to be realized, namely to provide metrics than can benchmark and incentivize ISPs and countries in terms of mitigation (see section 2.1). This is being implemented in D4.4. We did note, however, that the benchmarking effect is undermined by the fact that the CCH can currently not produce normalized metrics which are more suited for direct comparisons across ISPs and countries. This is because of the proprietary nature of the subscriber data needed for such normalization.

6.2 *Evaluation of impact*

Second, we performed a qualitative evaluation of the changes in the infection levels in the countries that have set up national support centers. (A quantitative statistical evaluation was not possible for reasons discussed in Chapter 2.)

We generated rankings and time trends for the countries with ACDC National Support Centers and compared those to a population of 54 other countries. The countries with an ACDC NSC perform very different, as do the ISPs in those countries. They are similarly distributed as the other ISPs and countries. Some have infection rates that are substantially better (i.e., lower) than average. Others are average or worse than average.

This findings should not be too surprising. The NSCs themselves have only been operational for a brief period and it is unlikely that they could have a measureable impact on infection levels. This is why we explored whether the setup of an NSC is part of an ongoing process of improvement in mitigation over time. In other words, we look at whether the countries with an NSC did cleanup faster than other countries. It turned out that they did not. The countries and ISPs that did well maintained their position. The ones who struggled with higher infection rates, continue to struggle with them. There does not seem to be a relationship with the presence of an NSC, nor with whether the NSC is operationally involved in mitigation.

None of this is to say that there is no improvement in cleanup as such. In most of our data sources, the number of infections are going down. This is partially related to the type of data sources we use (mostly sinkholes), and partially actual ongoing cleanup.

It is too early to declare a substantial verdict on the impact associated with a NSC. It is a well-known problem in the field of evaluation research that evaluating projects on the basis of outcomes is a harsh test under the best of circumstances, as there are so many factors are at play that can undermine the impact of the intervention. Furthermore, the impact may be there but not strong enough to be detectable by those indicators that can be measured. In other words, the “no effect” conclusion might be more related to the measurement limitations than to the actual lack of an effect.

That being said, from a policy perspective, it makes sense to consider these findings seriously. There are two main implications we want to highlight. First is the issue of scale. According to Microsoft, and our own analysis, around one percent of all Windows PCs are infected at any moment in time. Germany has had an active center for around five years now. In the course of around 54 months of operation, the German NSC has processed over 250.000 customer tickets and almost 10.000 support calls. This means around 4600 tickets per month. How to interpret this number? Well, according to Microsoft, and our own analysis, at any moment in time around one percent of all Windows PCs are infected. Germany has around 29,000,000 broadband subscribers (see Annex I). If one percent of those have an infection at any point in time, this translates to 290,000 infections. Suddenly, the number of tickets (4600 per month) does not seem so substantial anymore.

A similar simple calculation can be done for the Spanish NSC. Since November 2014, the NSC runs a pilot with Telefónica, the main ISP. In that period, the NSC has notified Telefónica about 957.470 unique IPs. Telefónica has sent 82.400 notifications to 16.627 different end-users. Each month 1.450 new customers are notified. Compare this to a one percent infection level. Spain’s has around 126,000 broadband subscribers. That is an order of magnitude higher than the total number of subscribers that have been contacted in around eight months, and two orders of magnitudes higher than the number of customers contacted each month.

In sum: when interpreting the findings of our evaluation, it seems that setting up the NSCs are important and necessary steps, but real impact requires scaling up these efforts well beyond their current levels.

The second policy-relevant finding is that we found less variance among ISPs in the same country than we expected. In several countries, we saw surprisingly similar infection levels. This seems to suggest that ISPs’ mitigation policies are less idiosyncratic and are in fact guided by what the peers/competitors in their own country are doing, as well as by the attention and pressures of public actors. In other words, ISPs seem to move together as they orient their mitigation efforts on what is considered ‘normal’ in their market and what their peers/competitors are doing.

From the perspective of national and European policies, this suggests that an NSC might get real traction if some of the leading ISPs commit to acting on infection notifications sent by the NSC. Our measurements suggest that the rest of the market will follow. This way a country could move to a

consistently lower infection rates – i.e., higher rank – and reduce the societal damage cost by botnets. ACDC can help push this process forward by contributing data from the CCH on which the ISPs can act.

In sum: ACDC has been able to develop valuable output in the form of useful and new data that can drive mitigation. The current levels at which this mitigation takes place does not have a visible impact on infection levels. This might situation might improve if the NSCs start operating at larger scales.

7 Annex I: List of countries included in the analysis

Country Code	Country Name	Region	Broadband subscribers (2014Q3)
AR	Argentina	Latin America	6,000,000
AU	Australia	Asia-Pacific	6,325,400
AT	Austria	EU	2,635,000
BY	Belarus	Europe, Other	2,143,100
BE	Belgium	EU	3,860,000
BR	Brazil	Latin America	23,650,000
BG	Bulgaria	EU	1,489,000
CA	Canada	North America	12,480,000
CL	Chile	Latin America	2,450,000
CN	China	Asia-Pacific	199,767,000
CO	Colombia	Latin America	4,775,000
HR	Croatia	EU	943,966
CY	Cyprus	EU	241,000
CZ	Czech Republic	EU	2,801,000
DK	Denmark	EU	2,360,000
EG	Egypt	Africa	2,955,000
EE	Estonia	EU	395,000
FI	Finland	EU	1,712,790
FR	France	EU	25,665,000
DE	Germany	EU	29,178,000
GR	Greece	EU	3,065,000
HU	Hungary	EU	2,390,000
IS	Iceland	Europe, Other	117,000
ID	Indonesia	Asia-Pacific	5,249,950
IE	Ireland	EU	1,248,675
IL	Israel	Asia	2,062,000
IT	Italy	EU	14,250,000
JP	Japan	Asia-Pacific	44,903,366
KZ	Kazakhstan	Asia	2,027,000
LV	Latvia	EU	510,000
LT	Lithuania	EU	742,200
LU	Luxembourg	EU	180,200
MY	Malaysia	Asia-Pacific	3,000,000
MT	Malta	EU	150,000
MA	Morocco	Africa	939,653
NL	Netherlands	EU	6,898,950
NZ	New Zealand	Asia-Pacific	1,390,000
NO	Norway	Europe, Other	1,955,050
PK	Pakistan	Asia	4,100,000
PE	Peru	Latin America	1,720,000
PH	Philippines	Asia-Pacific	6,550,000
PL	Poland	EU	5,680,000
PT	Portugal	EU	2,726,000
RO	Romania	EU	3,835,000

RU	Russia	Europe, Other	25,550,000
SA	Saudi Arabia	Asia	3,240,000
RS	Serbia	Europe, Other	1,234,900
SK	Slovakia	EU	1,150,000
SI	Slovenia	EU	548,000
ZA	South Africa	Africa	1,720,000
KR	South Korea	Asia-Pacific	20,010,643
ES	Spain	EU	12,630,000
SE	Sweden	EU	3,273,000
CH	Switzerland	Europe, Other	3,475,000
TH	Thailand	Asia-Pacific	5,442,000
TR	Turkey	Europe, Other	8,615,000
UA	Ukraine	Europe, Other	6,440,000
GB	United Kingdom	EU	22,989,000
US	United States	North America	111,350,000
VN	Vietnam	Asia-Pacific	5,717,923

8 References

[1] Van Eeten, M., J. Bauer, H. Asghari and S. Tabatabaie (2010). The Role of Internet Service Providers in Botnet Mitigation: An Empirical Analysis Based on Spam Data. STI Working Paper 2010/5. OECD.

[2] R. H. Shumway and D. S. Stoffer, *Time series analysis and its applications*. Springer Science & Business Media, 2013.

[3] L. Metcalf and J. M. Spring, "Everything you wanted to know about blacklists but were afraid to ask," *Softw. Eng. Inst. CERT Coord. Cent. Pittsburgh PA Tech Rep CERTCC-2013-39*, 2013.

[4] H. Asghari, M. Ciere, and M. J. G. van Eeten, "Post-Mortem of a Zombie: Conficker Cleanup After Six Years," in *24th USENIX Security Symposium (USENIX Security 15)*, Washington, D.C., 2015.

[5] M. van Eeten, H. Asghari, J. Bauer, and S. Tabatabie, "ISPs and Botnet Mitigation: A Fact-Finding Study on the Dutch Market," Dutch Ministry of Economic Affairs, The Hague, The Netherlands, 2011.

[6] L. Zhuang, J. Dunagan, D. R. Simon, H. J. Wang, and J. D. Tygar, "Characterizing Botnets from Email Spam Records," in *Proceedings of the 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats*, Berkeley, CA, USA, 2008.

[7] B. Krebs, "Spam Volumes: Past & Present, Global & Local," Jan-2013. [Online]. Available: http://krebsonsecurity.com/2013/01/spam-volumes-past-present-global-local/.

[8] Cisco Systems, "Spam overview - SenderBase," 2014. [Online]. Available: http://www.senderbase.org/static/spam/#tab=1.

[9] TrendMicro USA, "Global Spam Map," 2015. [Online]. Available: http://www.trendmicro.com/us/security-intelligence/current-threat-activity/global-spam-map/.

[10] ShadowServer, "Conficker Working Group," Dec-2013. [Online]. Available: http://www.confickerworkinggroup.org/.

[11] "Zeus Tracker." [Online]. Available: https://zeustracker.abuse.ch/.

[12] H. Binsalleeh, T. Ormerod, A. Boukhtouta, P. Sinha, A. Youssef, M. Debbabi, and L. Wang, "On the analysis of the zeus botnet crimeware toolkit," in *Privacy Security and Trust (PST), 2010 Eighth Annual International Conference on*, 2010, pp. 31–38.

[13] M. Riccardi, D. Oro, J. Luna, M. Cremonini, and M. Vilanova, "A framework for financial botnet analysis," in *eCrime Researchers Summit (eCrime), 2010*, 2010, pp. 1–7.

[14] N. Falliere and E. Chien, "Zeus: King of the bots," *Symantec Security Response*, 2009. [Online]. Available: http://bit.ly/3VyFV1.

[15] D. Andriesse and H. Bos, "An analysis of the zeus peer-to-peer protocol," Technical report, Free University of Amsterdam, 2013.

[16] I. Poese, S. Uhlig, M. A. Kaafar, B. Donnet, and B. Gueye, “IP Geolocation Databases: Unreliable?,” *SIGCOMM Comput Commun Rev*, vol. 41, no. 2, pp. 53–56, Apr. 2011.

[17] Maxmind, “Maxmind,” 2012. [Online]. Available: http://www.maxmind.com/.

[18] TeleGeography, “GlobalComms Database Service,” Mar-2014. [Online]. Available: http://www.telegeography.com/research-services/globalcomms-database-service/.

[19] B. Stone-Gross, M. Cova, L. Cavallaro, B. Gilbert, M. Szydlowski, R. Kemmerer, C. Kruegel, and G. Vigna, “Your botnet is my botnet: analysis of a botnet takeover,” in *Proceedings of the 16th ACM conference on Computer and communications security*, 2009, pp. 635–647.

[20] G. C. M. Moura, C. Gañán, Q. Lone, P. Poursaied, H. Asghari, and M. van Eeten, “How Dynamic is the ISPs Address Space? Towards Internet-Wide DHCP Churn Estimation,” in *Networking Conference, 2015 IFIP*, Toulouse, France, 2015.

[21] OECD, “Proactive Policy Measures by Internet Service Providers against Botnets,” July 2012.

[22] “Microsoft Security Intelligence Report, Volume 18,” Jul-2014. [Online]. Available: http://download.microsoft.com/download/7/1/A/71ABB4EC-E255-4DAF-9496-A46D67D875CD/Microsoft_Security_Intelligence_Report_Volume_18_English.pdf.

[23] H. Asghari, M. van Eeten, and J. Bauer (forthcoming), “Economics of Fighting Botnets: Lessons From a Decade of Mitigation,” *IEEE Security & Privacy magazine*.