Individual Assignment

Economics of Security (WM0824)
Member of Group 1

Lisa de Wilde (s1091514)

November 8, 2015

Abstract

ZeuS is a very effective Trojan with low detection rates which is part of the security issue for law enforcement agencies. Within this research it is investigated that there is no correlation between the ZeuS detection rate of antivirus programs and the time a ZeuS Command&Control was online before it was removed. However, there were multiple limitations within this research, so it may be better to improve the datasets and then perform the research again. Still there is a relation between the detection rate and the time the MD5 hashes are known. So, it is recommended that AV programs make sure their databases are up-to-date and that different security parties or law enforcement agencies keep taking down ZeuS C&Cs as soon a possible to limit the damage by the ZeuS Trojan.

1 Introduction

"The state of financial Trojans" report of Symantec (Wueest [2015]) shows that Trojans targeting financial institutions are one of the most prevalent threats nowadays. Attackers profit when an online back account has been compromised successfully and to let it stay profitable, criminals continuously update their Trojans to bypass new protection methods.

One of those threats is ZeuS, a Trojan that steals (especially) financial data. This Trojan and its variants compromised around 4.000.000 computers in 2014 and while the number of financial Trojans has decreased by 53% in 2014, the number of infections by ZeuS (and its variants) grew by ten times from 2012 to 2014 (Wueest [2015]).

ZeuS is an effective Trojan that is mostly spread via email as well by drive-by infections (ZeuSTracker [2015]). This effects individuals, but also companies and even governments and has already lead to a worldwide loss of millions (Wyke [2011]).

To track ZeuS Command&Control servers (C&C) and malicious hosts which are hosting ZeuS files, ZeuSTracker [2015] can be used. This tracker provides a block list containing well-known ZeuS hosts, in order to avoid and detect ZeuS infections in your network.

As discovered during previous assignments, law enforcement agencies all over the world, can try to track the C&C's and take down ZeuS C&Cs and eventually trace criminals. Besides that, they can protect civilians and organizations against the ever adapting financial malware and methods of spreading financial malware, which includes creating awareness about financial malware. This results in the following security issue: *Law enforcement agencies have to fight against ever adapting financial malware and methods of spreading financial malware.* The focus within this security issue is specifically on the ZeuS Trojan.

Law enforcement agencies can perform multiple strategies with the purpose of positively influencing the security issue. One of them is cooperating with security companies and make sure that security companies develop new or better AV (antivirus) programs, which detect and remove the ZeuS Trojan. Currently, the detection rates of ZeuS are very low, which result in multiple infections and stolen information. When civilians and companies use improved AV programs the number of infections may decrease and less credentials will be stolen. Eventually, it might also positively influence the security issue for law enforcement agencies, since better AV programs might even detect the ever adapting malware.

This research consists of two parts, first a literature review is done to find out what information already exists about the security issue and the detection of the ZeuS Trojan by AV programs (see section 2). The second part of this research is a statistical analysis of data-sets to support the information found during the literature review. In order to perform the analysis a research question, an objective and two hypotheses are defined (see section 3). To answer the research question a design of the research, including a description of the used data-sets, is given in section 4. The results of the performed study can be found in section 5. Finally, the limitations and conclusions are given in section 6 and 7.

2 Literature Review

Within this section the literature review about ZeuS and its detection is discussed. Papers were found using search terms as ZeuS, MD5 hashes, ZeuS signature, malicious software, ZeuS crimeware toolkit etc. In order to find relevant papers, different academic search engines like Web of Science, Google Scholar, Scopus and ScienceDirect were used. As the papers should be highly cited, the used papers were selected on the number of references. The last section contains an outline of the appearance of Trojans and especially ZeuS and its detection rate. For this multiple reports of different organizations are used. Note that those reports are not highly cited, but it is assumed that they are reliable, since they are created by leading antivirus companies.

2.1 Malicious Software

Malicious software can be classified by how the software infects its target and might replicate after infection. This results in three categories: viruses, Trojans and worms. Trojans, such as ZeuS, pretend to be something they are not, in this way they can bypass the computer defenses. So, it looks like a normal piece of software, however they can execute on the computer of the target resulting in major issues (Cass [2001]).

Trojans have different ways of spreading: they are presented as legitimate software, thus users will download them or it is downloaded and installed together with legitimate software without the users knowing it. The Trojan can also be sent to computer users as an e-mail attachment by someone posing as a bank employee for example. Trojans created to perform financial fraud are defined as banking Trojans (Stahlberg [2007]). ZeuS is a banking Trojan, since its purpose is to steal financial credentials.

2.2 ZeuS

ZeuS is a crimeware kit which can easily be used and is powerful in stealing data from various online sources. Since the kit is based on the do-it-yourself model, (low-level) cyber criminals can create their own version of the ZeuS Trojan (Dahbur et al. [2011]). This may result in multiple versions of the ZeuS Trojan. In addition, the software is designed to be profitable also by its development (Riccardi et al. [2013]).

Before explaining more about the ZeuS crimeware kit, the terms “bot” and “botnet” need to be defined first. According to Binsalleeh et al. [2010] a bot is “a software robot or a malware instance that runs autonomously and automatically on a compromised machine without being noticed by the victim user” and a botnet is “a network of bots that are controlled by an attacker (botmaster)”.

The purpose of the Zeus botnet is to make machines behave as spying agents with the aim of getting financial benefits by stealing data such as passwords and online banking accounts. This data then can be sold on the black market.

To set up a botnet over a high-scaled networked infrastructure criminals can use the ZeuS crimeware toolkit, which consists of the following components (Binsalleeh et al. [2010]):

1. *Control panel*
   A set of PHP scripts to monitor the botnet, collect the stolen information into a database and display it to the botmaster. Botmasters can also monitor, control and manage bots that are registered within the botnet using this panel.
2. *Configuration files*
   Two types of files to customize the parameters of the botnet:
   - configuration file: lists the basic information;
   - web injects file: identifies the targeted websites and defines the content injection rules.
3. *Generated encrypted configuration file*
   An encrypted version of the configuration parameters of the botnet.
4. *Generated malware binary file*
   The bots binary file that infects the victims machines.
5. *Builder program*
   Generate two files: the encrypted configuration file and the malware (actual bot) binary file.

First, the C&C server is set up using an installation script that configures the database and control panel. Within this database information about the botnet is stored together with information stolen by the bots from the infected machines. Each ZeuS instance has a set of targeted URLs that are fed by the web injects file.

At the same time, these URLs are targeted by ZeuS to steal information and to modify the content of specific web pages before they get displayed on the user's computer screen. As mentioned before the web inject file contains rules, which can be used to harvest web form data. When a computer users enters his credentials to one of the targeted sites the bot steals them. Then the bot posts the encrypted information to a drop location that is meant to store the bot update reports containing the stolen data and finally the C&C decrypts the stolen data and stores it into the database (Binsalleeh et al. [2010]).

2.3 Signature detection

According to Sanok Jr [2005] antivirus software companies develop different methodologies which can be used to defend against Trojans and other malicious software. One of the methods to find malicious software is file signature detection.

AV programs can scan the whole computer and look for a specific string of code that could be the virus (signature). This specific piece of code will be compared to the signatures in the database for verification. When the signature is found in the database, some removal steps should be performed by the AV program.

Nowadays, scanning can be performed on-demand and on-the-fly. When the user wishes to check the computer an on-demand scan will be performed. Using this method the user can choose when to scan his computer and he can set different configurations. Scanning on-the-fly is a background process in which files will be scanned at the moment they are accessed, for example when files are downloaded from the Internet or when a new message enters an email inbox.

2.4 MD5 Hash

The MD5 message-digest function can be used for signature detection of ZeuS. MD5 is an algorithm that creates a 128-bit message digest, when taking a message of arbitrary length as input. It should be infeasible to create the same message digest for two messages and to create any message having a given prespecified target message digest (Rivest [1992]).

When applying it to the ZeuS Trojan, the MD5 hash of the binary file is compared to the MD5 hashes in the database of the AV program. If it appears in the database some removal steps should be taken (Pedersen et al.).

2.5 One step ahead

Once the signature of a virus is stored in a database of an AV program, the virus will be detected. However, when criminals change their malware, the AV programs have to update their signature database to detect the new malicious malware. ZeuS, for example, bypasses signature detection by re-encrypting itself with every infection, then a new signature will be produced. So, criminals frequently create new techniques to develop malware, such that it cannot be detected by the AV program and to stay one step ahead of AV programs (Alazab et al. [2012]).

2.6 Financial Trojans

For this section reports of different organizations are used to outline the appearance of Trojans and more specifically ZeuS.

First of all, the report from Pandalabs [2014] shows that 68,84% of the newly created malware in 2014 were Trojans. The reason for this is the emergence of Potentially Unwanted Programs (PUPs). PUPs are applications (not malicious per se) that install unwanted software without properly informing the user about this. Besides that, the report shows that 65,02% of the malware infections in 2014 were caused by Trojans.

Trojans can be categorized in different types, one of those types are financial Trojans. According to Kaspersky [2014] 72% of the attacks that target user money are banking malware. They also state that ZeuS is still in the top 10 banking malware families.

According to Wueest [2015], financial Trojans targeted around 1.467 financial institutions in 86 countries in 2014. The ZeuS Trojan targeted circa 200 financial institutions and compromised around 4.000.000 computers in 2014.

Besides the mentioned method of signature detection, antivirus programs use other techniques as well to detect all kind of viruses. Nevertheless, the detection rate for the ZeuS trojan is very low (Riccardi et al. [2013]). For example: the real botnet test of Security [2015] shows that only 10/21 AV programs prevented the malware from capturing login data within the same session.

3 Research Question, Objective and Hypothesis

It is hard to find specific detection rates for ZeuS in literature or in reports. Besides that it is known that ZeuS is one of the most well-known and widespread Trojans. A lot of research is done about ZeuS and how it works, but the detection rates of AV programs are still low.

Next to that, there is no direct information within the literature about the time each C&C is up before it is removed by the criminals or any other party. Since the correlation between the detection rate and the up-time of the C&Cs is not known yet, this will be explored within this research. This results in the following research question:

- What is the correlation between the detection rate and the period the ZeuS C&C was online before it was removed?

As mentioned in section 2 it takes some time before AV programs have updated their database with new MD5 hashes. Since the detection rate of AV programs depends on the MD5 hashes within their database, it must be investigated how the detection rate depends on the date of the detection rate analysis. Next to that, it is expected that there is a positive linear correlation between the detection rate and the period the ZeuS C&C was online before it was removed, because the AV programs then have more time to update their database. This results in the following two hypotheses:

1. The detection rate will be higher when the MD5 hash is known for a longer period of time;
2. The detection rate will be higher when the C&C is online for a longer period of time;

The objective of this paper is to provide more insight into ZeuS Trojans regarding the detection rate of AV programs. When knowing the correlation between the detection rate and the period the C&C was online before it was removed, law enforcement agencies can decide how they want to improve their strategies regarding the security issue.

4 Research Design

Within this section the research design is discussed. First, it is described how the data-sets are gathered and were they consist of (see section 4.1). After that, the research methods that will be used to analyze the data-sets will be explained (see section 4.2).

4.1 Data Collection

To answer the research question and test the hypotheses four data-sets from ZeuSTracker [2015] and VirusTotal [2015] will be analyzed. Zeus Tracker has been discussed before and VirusTotal is a free service that analyzes suspicious files and URLs. For this research MD5 hashes obtained from Zeus Tracker will be entered in the VirusTotal tool. The tool then returns a report containing the following:

- SHA256;
- File name;
- Detection rate;
- Date of the analysis;
- Overview of which AV program detect the Trojan and which do not.

VirusTotal calculates the detection rate by testing which AV programs detect the MD5 hash and which do not. For example: when 40 programs detect the hash and 10 do not, the detection rate will be $40/50 = 80\%$.

4.1.1 Data-set 1

For the first data-set information was extracted from the ZeuS Tracker statistics page. This data consists of 9224 ZeuS binaries divided among 11 detection rate categories between 0% and 100%. It is not explained how ZeuS Tracker calculated those detection rates, but it is assumed that the detection rates from VirusTotal are used for this.

4.1.2 Data-set 2

The second data-set is extracted from Zeus Tracker as well. The overview of BinaryURLs is extracted, which contains 86 entries with the following information:

- Date added;
- ZeuS binaryURL;
- MD5 hash;
- Detection rate (%).

The detection rate within this data-set was obtained by ZeuS Tracker from VirusTotal. Since this data-set does not contain the date that the detection rate analyses were performed, this was added manually using VirusTotal [2015]:

- Date detection rate analysis.

4.1.3 Data-set 3

Using VirusTotal the detection rate of the last detection rate analyses could also be found and these were added to data-set 2 including the date the detection rate analyses were performed. This results in data-set 3.

4.1.4 Data-set 4

The fourth data-set contains the following information:

- Date added;
- Removal date;
- Host;
- Detection rate (%).

This information was extracted from Zeus Tracker and contains 4908 entries. As there is no link added to the detection rate is it not known at which date the detection rate analyses were performed.

4.2 Research methods

To analyze the data-sets a quantitative research will be performed. Data will be transformed into statistics, which then can be used to quantify the correlation between the up-time of ZeuS C&Cs and the detection rates of AV programs.

Within this research multiple statistical methods will be conducted. First, one of the basic statistical methods will be performed, namely calculating the average. Using this method the average detection rate can be calculated for all data-sets. Since the data-sets are containing different information the different averages can be compared regarding the key characteristics of the data-sets.

For data-set 1 multiple groups with different detection rates will be compared against the number of binaries within the groups. This statistical analysis will be performed to investigate which detection rate groups are most common.

Data-set 2 and 3 will be analyzed by dividing the data-sets in groups regarding their characteristics. Those groups will be compared to evaluate the characteristics of the groups.

Finally, the correlation between the detection rates and up-time of C&Cs will be calculated using the correlation-coefficient. This coefficient determines whether or not there is a correlation within data-set 4. When there is a correlation, the actual relationship will be determined by conducting a regression analysis.

5 Results

In this section the results of statistical analyses are discussed. First each data-set will be analyzed separately and at the end of this section the results of the different analyses will be compared.

5.1 Analysis Data-set 1

The first data-set is used to create a graph that shows the division of ZeuS binaries against detection rate categories (see figure 1). The hatched surface shows that 70% of the 9224 binaries have a detection rate lower than 50%. Besides that only 66 binaries have a detection rate between 90% and 100%. Finally, the detection rates equals 40,1%, which is expected as stated by the literature, but is still alarming.

This data-set has a large number of entries (9224), nevertheless this data-set does not give any information about other factors, such as date of the detection rate analysis and the up-time of the ZeuS C&C. The influence of those factors is explored hereafter.

Finding 1: ZeuS binaries have an average detection rate of 40,1%.

5.2 Analysis Data-set 2

As mentioned before the detection rates that were used in data-set 1 did not give any information about the date that the analyses were performed. To be accurate the detection rate should be obtained the first day the C&C was discovered. Since data-set 2 contains the date of the detection rate analyses and the date the C&C appeared on the tracker, the difference between those dates could be easily calculated. It turned out that not every analysis was performed on the same day as the C&C was discovered, which resulted in three cases:

1. the analysis was performed before the C&C appeared on the tracker;
2. the analysis was performed on the same day that the C&C appeared on the tracker;
3. the analysis was performed after the C&C appeared on the tracker.

5.2.1 Before

22/86 analyses were performed before the C&C appeared on the tracker. The detection rates for this case can be found in figure 2a. As can been seen the detection rates are quite high, which result in an average detection rate of 76,45%.

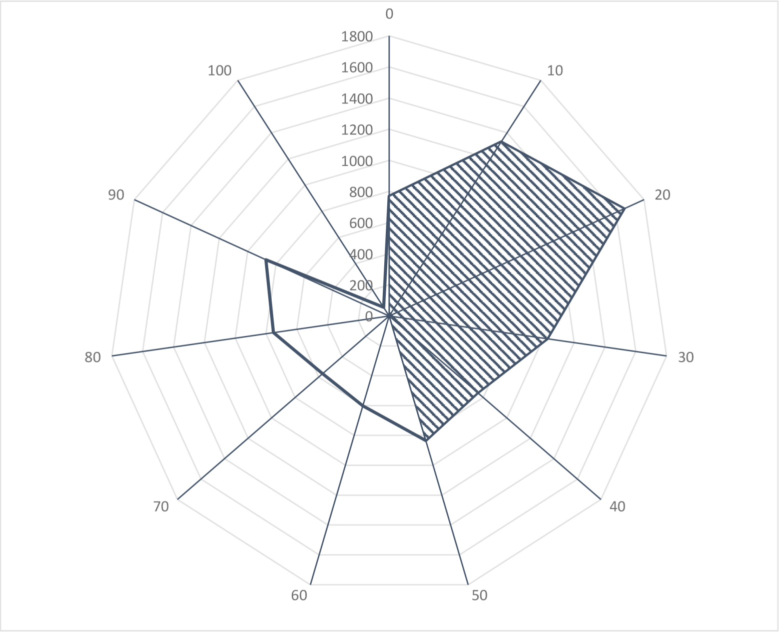


Figure 1: Number of ZeuS binaries per detection rate category

5.2.2 Same day

Only 43/86 of the analyses were performed on the same day the C&C appeared on the tracker. As figure 2b shows the detection rates for those 43 analyses deviate at lot, with even quite low detection rates. Resulting in an average detection rate of 57,66% for the second case.

5.2.3 After

The last case are the analyses (21/86) that were performed after the C&C appeared on Zeus Tracker, so the difference is > 0 days (see figure 2c). As not can been seen in the figure, there were also 9/21 detection rates of 0%. It may be that non of the AV programs knew about the MD5 hash, or the binary was not malicious at all. Together with the zero rates, the average detection rate is 31,45%.

5.2.4 Overview

The average detection rates for the three cases can be found in table 1, together with the average of all cases together.

As can been seen in the table the detection rate is the highest for case that the analysis was performed before the C&C was discovered by the tracker. This means that the MD5 hash of the binary file was already known by VirusTotal before Zeus Tracker found out about the new C&C. It is not known why those rates are much higher than when the analysis was performed on the same they the C&C appeared on the tracker. A reason may be that most AV programs already knew about the Trojan before the tracker and VirusTotal did.

Table 1: Average detection rate

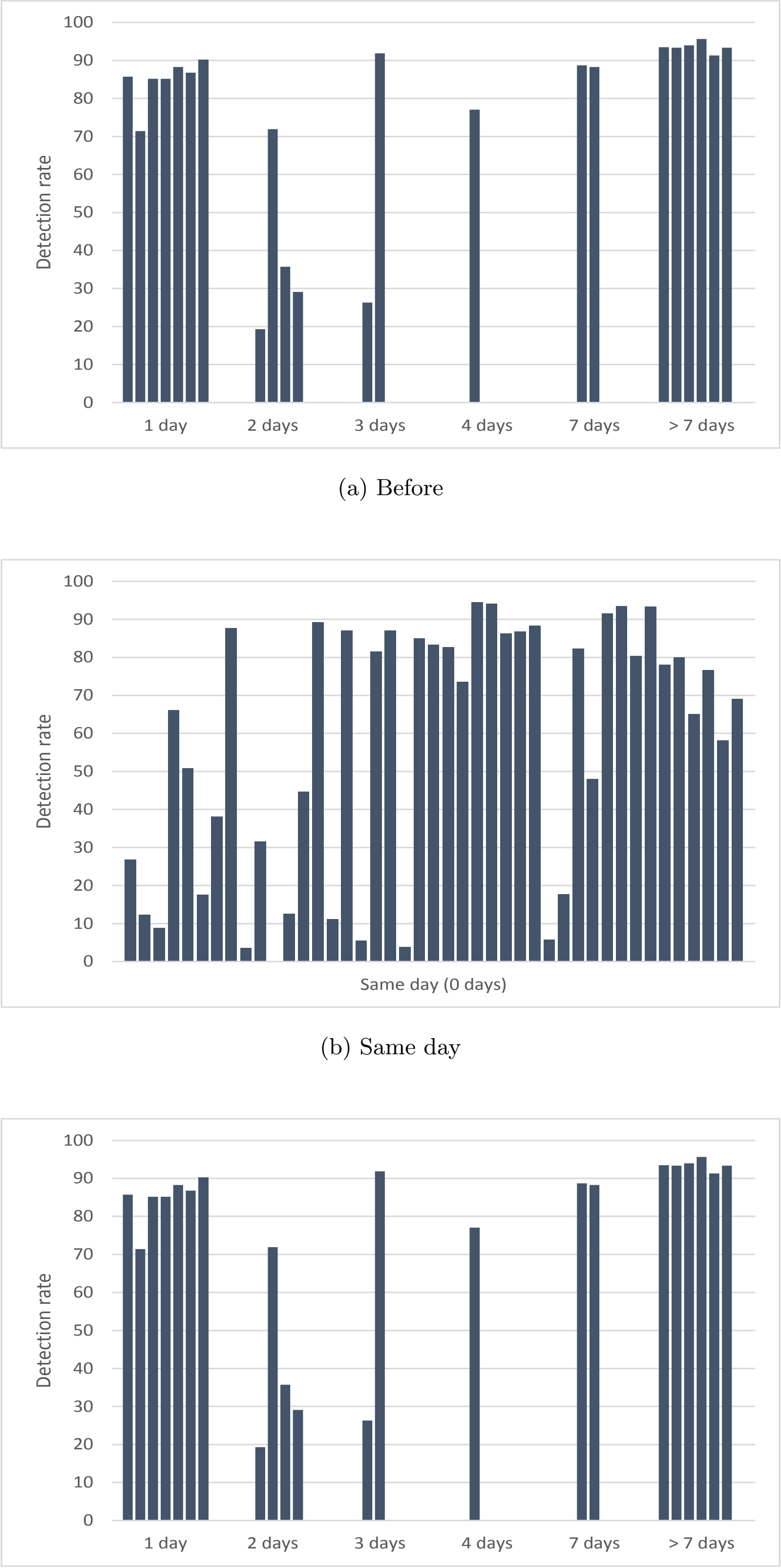
Measured	**Average detection rate**
Before	76,46%
Same day	57,66%
After	31,45%
Average	**56,07%**

The detection rates for the same day and after) differs a lot. According to those statistics it may be that VirusTotal and the AV programs did not know about the Trojan when an analysis was performed later in time or the AV programs did not update their (signature) database yet.

According to this statistical analysis the detection rate will be higher when the detection rate analysis is performed earlier by VirusTotal, even if the C&C is not known by ZeuS Tracker.

Please note that those results may be biased by the fact that the data-set is limited and the fact that there were some zero detection rates. For more about this see section 6.

Finding 2: ZeuS binaries have an average detection rate of 56,07% and the earlier the detection rate analysis is performed, the higher the detection rate.



(a) Before

(b) Same day

(c) After

Figure 2: Detection rates for the three different cases

5.3 Data-set 3

As mentioned in section 2 and 5.2.4 it is suggested that the detection rate will be higher when the first analysis is performed earlier in time regarding the first appearance of the C&C. It is also expected that it takes an amount of time before the majority of the AV programs can detect the MD5 hashes. To confirm this a comparison with a analysis later in time should be done too. Unfortunately, there are no detection rate analyses available for multiple periods of time. However, the last analysis can be found using VirusTotal, so this is used for this research.

First, the difference between the first analysis and last analysis for every binaryURL was calculated. This resulted in a range between 1 and 1452 days. After that, two groups were created: the first group contains the detection rates of detection rate analyses that were performed within a year after the first analyses and the second group the rates of analyses that were performed after a year.

5.3.1 Within a year

Figure 3 shows how the detection rates of the first analyses (see section 5.2) increased or decreased when the last analyses were performed within a year. As can been seen most of the detection rates increased and most of the low detection rates became higher detection rates. The average grow of the infection rates is 26,83%.

5.3.2 After a year

The other group contains the detection rates of the last analyses that were performed after more than one year. This results in an average grow of 12,5%. As can been seen in figure 4 most of the low detection rates grow to high detection rates after a year. Just as described in section 5.2 the zero rates are not visible in this graph (first and last analysis 0%), but are taken into account when calculating the average detection rate.

5.3.3 Overview

For both groups the average increase is added to the detection rates of the first analysis, which has to be calculated first. This results in the average detection rate of the last analyses. Finally, the average increase of both groups together is added to the average detection rate of data-set 2 (see section 5.2). Those results can be found in table 2.

As the table shows, the increase for the first group is much higher than the second group. Since their are no intermediate detection rates available, it is impossible to give the reason for that difference. But, the average detection rate of 74,89% is high. However within the time it took to update the AV programs, multiple computer users could have been infected and their data already may be stolen.

Using those statistics it could be concluded that the detection rate will be higher when the MD5 hash is known for a longer period of time. To verify this the correlation coefficient is calculated using the detection rates of the first and last analyses. This leads to a correlation coefficient of 0,71 using the correlation function of Microsoft Excel. So, there is a strong positive linear relationship between the first and last analyses. Thus, the first hypothesis “the detection rate will be higher when the MD5 hash is known for a longer period of time” can be confirmed.

Finding 3: there is a strong positive linear relationship between the detection rates of the first and last analyses. The detection rate at the last analysis is on average 18,83% higher than during the first analysis.

5.4 Analysis Data-set 4

So, there is a relation between the detection rate and the period the MD5 hash is known. Now, the fourth data-set will be used to determine if the number of days before the C&Cs were removed influences the detection rate.

First the difference between the date the C&Cs were added and removed was calculated. After that, the detection rates were grouped per number of days the C&C was online. Those groups were added to a graph, that shows how the detection rate change as the number of days changes (see figure 5).

As can been seen all detection rates are below 60%, which is quite low, especially compared to the detection rates seen before. Besides that, the average detection rate is 50,44%. This result could be influenced by the fact that it was not known when the detection rate for each C&C was determined.

It can also been seen that the linear trend line slowly decreases, which means that the detection is lower the longer the C&C is online. To confirm this, there should be a correlation between those two variables. To see if there is a correlation, the correlation coefficient is calcu-

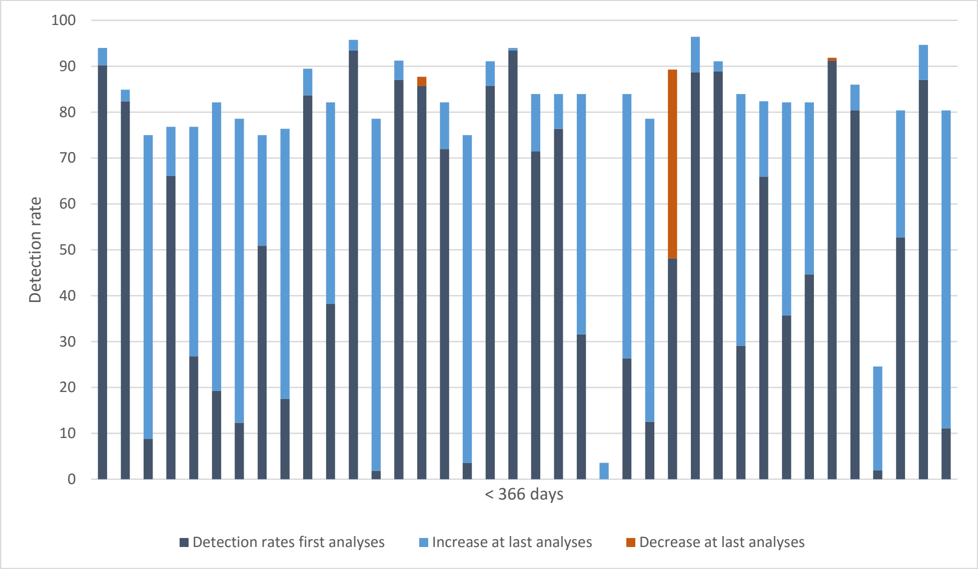


Figure 3: Detection rates within a year

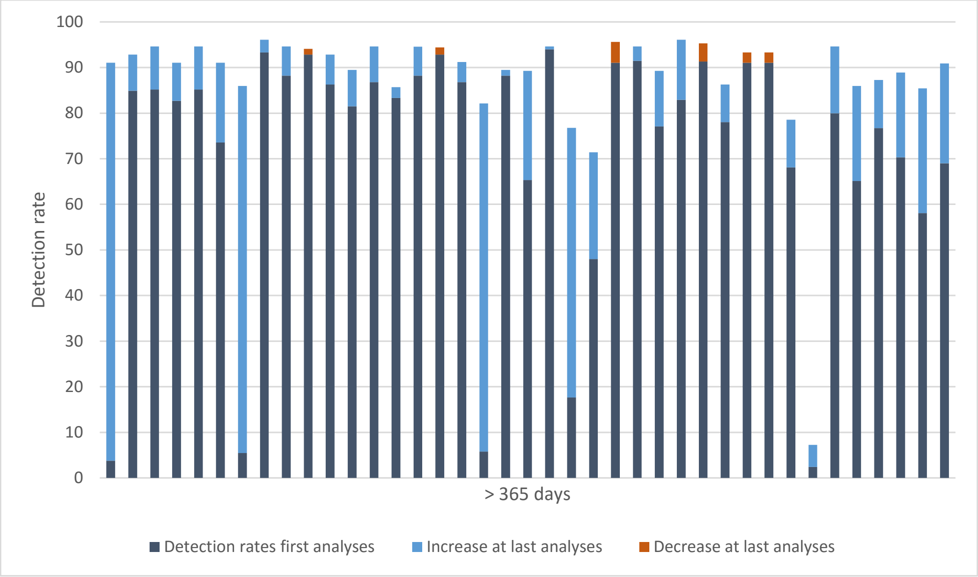


Figure 4: Detection rates after a year

Table 2: Increase of detection rates within and after a year

Group	First analyses	Average increase	Last analyses
Within a year	52,80%	26,83%	79,63%
After a year	58,66%	12,50%	71,16%
Average	**56,06%**	**18,83%**	**74,89**

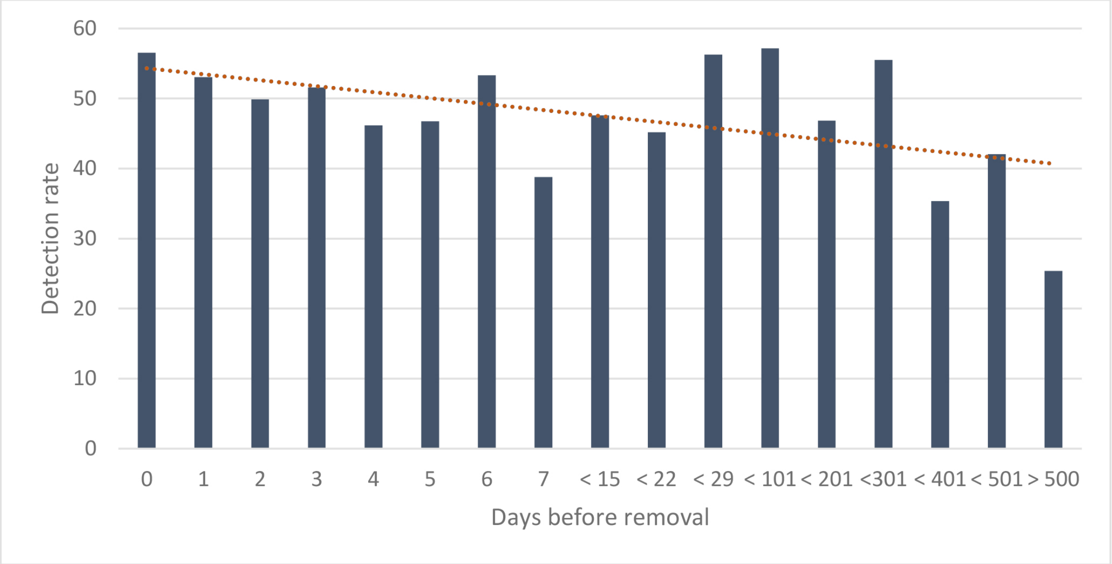


Figure 5: Detection rate per number of days before removal

lated using the correlation function of Microsoft Excel. This resulted in a correlation coefficient of -0,17, which is a very weak negative linear relationship between the detection rate and the up-time of the C&C. So, the second hypothesis “the detection rate will be higher when the C&C is online for a longer period of time” will be rejected.

Finding 4: there is a very weak negative correlation between the detection rate and up-time of the ZeuS C&C and ZeuS binaries have an average detection rate of 50,44%.

5.5 Overview of the Statistical Analyses

Multiple analyses have been performed to test the hypotheses and answer the research question. The conclusions will be discussed in section 7, but here an short overview is given. The key findings of the statistical analyses are:

1. ZeuS binaries have an average detection rate of 40,1%;
2. ZeuS binaries have an average detection rate of 56,07% and the earlier the detection rate analysis is performed, the higher the detection rate;
3. There is a strong positive linear relationship between the detection rates of the first and last analyses. The detection rate at the last analysis is on average 18,83% higher than during the first analysis;
4. There is a very weak negative correlation between the detection rate and up-time of the ZeuS C&C and ZeuS binaries have an average detection rate of 50,44%.

An overview of the detection rates mentioned in the key findings of every data-set can be found in table 3.

6 Limitations

Within this section the limitations of this research will be discussed. For each of the limitations a recommendation or improvement for other researchers is given as well.

6.1 Sources

This research was limited by the availability of ZeuS Tracker. During this research ZeuS Tracker was offline for a few days. A wayback machine (a digital archive of the information

Table 3: Overview detection rates

Data-set	Analysis #	Average detection rate
1	Unknown	40,10%
2	First	56,07%
3	Last	74,89%
4	Unknown	50,44%

on the internet) was used to get access to the site again. So, the site could be accessed again, however it was a version of 5 September 2015, so the information was out-dated.

Using VirusTotal only the first and last analyses were found, since there are no periodically analyses, it is hard to draw correct conclusions without unrealistic generalizations.

When cooperating with ZeuS Tracker and VirusTotal the limitations of the data sources could be avoided, since they can directly sent the data to the researcher.

6.2 Detection Rate

There is also a limitation in the calculation of the detection rate by VirusTotal. The program checks for a certain number of AV programs if they detect the MD5 hash of a binary file. However, not for every analysis the same number of AV programs is used, which may give a twisted view. Assume, for example, that the first analysis is performed for 40 AV programs, which all detect the trojan. After that the analysis is performed using 50 AV programs, from which 40 detect the trojan. This results in a detection rate of 100% and 80%, but it might be that those 10 extra AV programs, did not detect the trojan when they were added in the first analysis.

When multiple hashes will be compared with each other, but the first hash has been tested against 40 programs and the second against 60 programs, the detection rates may also give a twisted view.

To avoid this, all analyses should be performed using the same number of AV programs. This may be hard since new AV programs will be created as well and it is desirable to test their performance as well. Besides that, some AV programs may disappear, then they cannot be tested anymore.

Another limitation regarding the detection rate is that there were binaries with a detection rate of 0%. That is not a problem per se, however it could have two meanings. In the first place the detection rate is actually 0% and there is no AV program that detects the Trojan. The other case appears when the binaries are not malicious at all, then it should actually be removed from ZeuS Tracker.

6.3 Data-sets

Data-set 1 consists of a large number of entries, however besides the number of binaries per detection no further information is given. So, the influence of other factors could not be taken into account.

Data-set 2 and 3 were derived using a wayback machine, as mentioned before this data was not up-to-date. Besides that, those data-sets only contains 86 binaries with an AV detection rate. To find detection rates for other binaries as well a search was done using VirusTotal. It was possible to find the last analysis, but it was impossible to find the first analysis. So, this information could not be used.

For the last data-set there were no dates available for the detection rate analyses. As seen in the analyses of data-set 2 and 3, it does matter at what time the analyses are performed. Thus, this may also have biased the results.

To gather improved data-sets ZeuS Tracker and VirusTotal could be contacted for a cooperation in which they share available data with the researcher. MD5 hashes could be shared with VirusTotal and in return they share the detection rates for different periods of time. When ZeuS Tracker shares a list containing the up-time for every binary, the analyses will be improved too.

6.4 Time

Finally, there is off course a time limitation for this research. A lot of time is spent on this research, but when having more time, more analyses could be done which could result in improved results. Besides that, better and complete data-sets could be gathered and more research could be done about statistical analyses.

7 Conclusions

Within this research a literature review has been performed about the ZeuS Trojan and its detection. This resulted in the problem of very low detection rates by AV programs. The literature review did not result in any information about the up-time of C&Cs and the effect on the detection rates. So, a statistical analysis was performed to investigate the correlation between the detection rates of ZeuS Trojans and the up-time of ZeuS C&Cs.

Four different data-sets were extracted from Zeus Tracker and VirusTotal and have been analyzed using different methods. For all those data-sets average detection rates were calculated. Those detection rates range from 40% and 57% and can be classified as low. This finding is supported by the found literature as mentioned in section 2.

Furthermore, it was investigated that the day the detection rate analysis was performed influences the detection rates: i.e. the earlier the detection rate was measured the higher the detection rate. But, as mentioned before this may be a limitation of the data-set. To avoid incorrect representations the detection rates should all be measured on the same day. The best approach to do so, is to perform an detection rate analysis on the day the C&C appeared on ZeuS Tracker.

The first hypothesis "the detection rate will be higher when the MD5 hash is known for a longer period of time by AV programs" was confirmed during this research. There is a strong positive linear relationship between the detection rates of the first and last analyses. The average detection rate for the last analysis is determined as 74,89%, which is circa 20% higher than the detection rate during the first analysis. This makes sense, since AV programs will not remove the signature from the database unless it is not malicious. So, only AV programs who not have added the MD5 hash yet, will add it to their database, which will result in higher detection rates. Still, not every AV program adds the MD5 hash to their database, because the detection rate never reached 100% within this research. This may be because they do not know about the malicious binary.

Using the last data-set it was concluded that there is a very weak negative correlation between the detection rate and the period the ZeuS C&C was online before it was removed. So, the second hypothesis "the detection rate will be higher when the C&C is online for a longer period of time" was rejected. This also answers the research question.

Nevertheless, as discussed in section 6 there were multiple limitations within this research, so it may be better to improve the data-sets and then perform the research again. Until then, it is recommended that AV programs updates their databases as fast as possible to make sure the detection rates increase and that other security parties or law enforcement agencies still keep taking down ZeuS C&Cs as soon a possible to limit the damage by the ZeuS Trojan.

References

Mamoun Alazab, Sitalakshmi Venkatraman, Paul Watters, Moutaz Alazab, and Ammar Alazab. Cybercrime: the case of obfuscated malware. In *Global Security, Safety and Sustainability & e-Democracy*, pages 204–211. Springer, 2012.

Hamad Binsalleeh, Thomas Ormerod, Amine Boukhtouta, Prosenjit Sinha, Amr Youssef, Mourad Debbabi, and Lingyu Wang. On the analysis of the zeus botnet crimeware toolkit. In *Privacy Security and Trust (PST), 2010 Eighth Annual International Conference on*, pages 31–38. IEEE, 2010.

Stephen Cass. Anatomy of malice [computer viruses]. *Spectrum, IEEE*, 38(11):56–60, 2001.

Kamal Dahbur, Bassil Mohammad, and Ahmad Bisher Tarakji. A survey of risks, threats and vulnerabilities in cloud computing. In *Proceedings of the 2011 International conference on intelligent semantic Web-services and applications*, page 12. ACM, 2011.

Kaspersky. Kaspersky security bulletin 2014. 2014. URL `http://securelist.com/files/2014/12/Kaspersky-Security-Bulletin-2014-EN.pdf`. Accessed: 2015-11-03.

Pandalabs. Annual report 2014. Technical report, 2014. URL `http://www.pandasecurity.com/mediacenter/src/uploads/2015/02/Pandalabs2014-DEF2-en.pdf`. Accessed: 2015-11-03.

Jay Pedersen, Dhundy Bastola, Ken Dick, Robin Gandhi, and William Mahoney. Fingerprinting malware using bioinformatics tools building a classifier for the zeus virus.

Marco Riccardi, Roberto Di Pietro, Marta Palanques, and Jorge Aguila Vila. Titans revenge: Detecting zeus via its own flaws. *Computer Networks*, 57(2):422–435, 2013.

Ronald Rivest. The md5 message-digest algorithm. 1992.

Daniel J Sanok Jr. An analysis of how antivirus methodologies are utilized in protecting computers from malicious code. In *Proceedings of the 2nd annual conference on Information security curriculum development*, pages 142–144. ACM, 2005.

MRG Effitas Online Banking / Browser Security. Certification project - q3 2015. Technical report, 2015. URL `https://www.mrg-effitas.com/wp-content/uploads/2015/10/MRG-Effitas-Online-Banking-Certification-Q3-2015.pdf`. Accessed: 2015-11-03.

Mika Stahlberg. Internet fraud prevention, June 1 2007. US Patent App. 11/806,568.

VirusTotal. Statistics, 2015. URL `https://www.virustotal.com/nl/`. Accessed: 2015-11-02.

Candid Wueest. The state of financial trojans 2014. page 24, 2015. URL `http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/the-state-of-financial-trojans-2014.pdf`. Accessed: 2015-11-03.

James Wyke. What is zeus? *Sophos, May*, 2011.

ZeuSTracker. Zeus tracker, 2015. URL `zeustracker.abuse.ch`. Accessed: 2015-11-02.