Target selection regarding financial malware attacks within the Single Euro Payments Area

Mixed methods research: The expectations of experts versus the patterns shown by the Zeus dataset

Master thesis of Daan van Moorsel (4043197)

In partial fulfilment of the requirements for the degree of Master of Science in Systems Engineering, Policy Analysis and Management at the faculty of Technology, Policy and Management, Delft University of Technology

June 2016

GRADUATION COMMITTEE:

Chairman
Prof. Dr. M.J.G. van Eeten
Professor, Faculty of Technology, Policy and Management, TU Delft

First supervisor
Dr. Ir. A.J. Klievink
Assistant professor, Faculty of Technology, Policy and Management, TU Delft

Second supervisor
Dr. Ir. W. Pieters
Assistant professor, Faculty of Technology, Policy and Management, TU Delft

Daily supervisor
R.S. van Wegberg MSc.
PhD candidate, Faculty of Technology, Policy and Management, TU Delft

External supervisor
M. Bras
Policy Officer Cybersecurity, Payment Systems Division, De Nederlandsche Bank
Acknowledgements

This thesis represents the final work for the completion of my Master Systems Engineering, Policy Analysis and Management (SEPAM) at the Delft University of Technology (TU Delft). It was a learning experience, which I wouldn’t want to have missed. The first five months of the research, were executed at the Dutch Central Bank (DNB). Hereby, I would like to thank all the colleagues from the Payment Systems Division of the DNB. Personal thanks go out to, Raymond Kleijmeer for his enthusiasm at the start of the project and Kim Gunnink for connecting me with Colleagues at DNB. Special thanks go out to my roommates; Jaap Rotte for being an inspiring and motivating person, who is always available for questions or a discussion. And Rein Kieviet, who surprised me many times with his social and happy character and his young and energetic mind. Finally, I would like to thank Maarten Bras with the supervision of the research. It was very nice, to have such a smart and dedicated external supervisor.

In addition to the support at the DNB, I am very lucky with the graduation committee of my thesis. The subject of the thesis and the composition of the committee, were triggered by the course Cyber Security Essentials of prof. Jan van den Berg. He made me very enthusiast of the subject cyber security. Furthermore, during one of the guest lectures, prof. Michel van Eeten inspired me with his presentation regarding cyber security economics. During this graduation project, Michel was the chairman. He provided me with some interesting research ideas. While at the same time, he encouraged me to make considered research decisions by myself. This was a very nice way of working.

Moreover, I would like to thank my first supervisor Dr. Ir. Bram Klievink and my second supervisor Dr. Ir. Wolter Pieters. They are both experienced researchers, who were very helpful during the project. Bram really supported me, with determining the right direction for finalizing the research project. Furthermore, he encouraged me to pursue a high academic quality of both the thesis and the scientific article. In addition, Wolter provided me with advices, which make the thesis better understandable for the readers. In addition, he contributed to the composition of the research questions. Much of the quality of this research can be attributed to those two supervisors.

In addition to the research committee, Rolf van Wegberg provided the daily supervision for this project. I would like to thank Rolf, for his intensive involvement within both defining a research project and with the execution of the project. The whole project with Rolf was an educative and nice process. Rolf was always available for questions and feedback, while he is really busy with his own research as well. Many research ideas are provided by Rolf and included in this graduation project.

Furthermore, I would like to thank Samaneh Tajalizadehkhooob and Carlos Hernandez Gañán. Samaneh supported me with understanding the Zeus dataset, which is stored in a MySQL database. Besides, this research refers many times to her previous work with the Zeus dataset. In addition, Carlos supported me with testing the hypotheses. Thereby, they increased the quality of the quantitative analyses. Moreover, I would like to thank all the other experts consulted for the purpose of this project.

Finally, I would like to thank my family and friends for their support during the graduation project. Especially, my girlfriend who also graduated during this academic year. Finally, I would like to remember my grandfather, who suddenly passed away, one week after I started as a graduation intern at the DNB. He was always interested in the study projects I executed.

Daan van Moorsel
Oegstgeest, June 2016
Summary

This research deals with the selection of financial institutions, for the purpose of financial malware attacks. Hereby, target selection by financial malware schemes is researched. The financial malware scheme is the synergy of actors and organizations that enables financial malware attacks. Note that, not all actors within the financial malware scheme are criminals. Regarding this research, target selection is defined as: Attack choices by the financial malware schemes, regarding which financial institution to attack, at which point of time and for how many weeks. Two main reasons exist to focus on financial malware. First, because financial malware is on the top of the list of cyber threats (Marinos, 2014; Europol, 2015). Secondly, malware is used to execute man-in-the-middle attacks, which are very common in the financial sector (Vrancianu et al., 2010).

Due to the advent of the internet, many services of financial institutions become available online. Although financial institutions are continuously dealing with securing those modern payment services, online financial service fraud is a serious problem. In the early years of security, information security was particularly focused on technical measures. Nowadays, there has become a general understanding that socio-technical measures are required as well (van den Berg et al., 2014). The current focus of cyber risk management, within the European financial sector, is on the cyber resilience of financial market infrastructures (CPMI, 2014; DNB, 2015). Cyber resilience already deals with the criteria technology, people, processes and communication.

However, less knowledge exists regarding the level of cyber risks, encountered by different financial market infrastructures and financial institutions. To further develop cyber risk management, it is essential to have a clear understanding of the threat landscape. Therefore, this research intends to address the following gap: A limited understanding of why certain financial institutions are more likely to be selected as target by cybercriminals.

It is possible to gather insights into target selection, by financial malware schemes, because the malware attacks are botnet based. A botnet is a flexible remote-controlled network containing computers that function together to make a platform available for fraudulent and criminal purposes (Bauer and van Eeten, 2009). Those computers need instructions, about which financial institutions to attack at which point of time. Previous researchers Tajalizadehkhoob et al. (2013), created the Zeus dataset, which contains these instructions for Zeus financial malware between 2009 and 2013q1. Furthermore, they proved that is it possible to extract intelligence from this Zeus dataset.

This research intends to exploit the possibility, by involving cyber security experts and adding new data to the dataset. Thereby, empirical insights are provided, into both technical as socio-technical characteristics of financial institutions, which could make them more likely to be selected as target by financial malware schemes. Currently, less empirical knowledge exists regarding these characteristics. Besides, these insights should support both the financial institutions with their cyber risk management and the authorities with developing cyber security criteria. The focus is on targeted financial institutions within the Single Euro Payments Area (SEPA). For the purpose of this research, the following research question is developed:

What are the expectations of experts, regarding the evolution of target selection by financial malware schemes and how did target selection actually evolve according to the Zeus dataset?

A mixed methods approach, of both qualitative and quantitative research methods, is used to answer this question. For the purpose of the first part, three experts are consulted extensively during the whole project, besides four experts are consulted ones or twice with semi-structured interviews. The outcomes of the consultations are placed within the perspective of the Routine Activity Theory (RAT). RAT proposes that crime occurs, when a suitable target is in the presence of a motivated offender and is without a capable guardian. Thereby, it supports in understanding, why certain financial institutions are more likely to be selected as target. Based on the conditions of RAT, it is determined
whether the experts assume that target selection is more focused on the suitability of a target or on the absence of a capable guardian.

The experts expect that in the first years of financial malware, target selection was particularly influenced by the suitability of a financial institution. Especially the size of the financial institutions seems to be taken into account, which – for instance - can be expressed by the number of clients. Besides, decreasing the accessibility of the financial institution is mentioned to be a short time solution. Because of the reason, that the cybercriminals can easily find out the new ways of authentication and circumvent them. Moreover, new ways of access won’t distinguish financial institutions for a long period of time, because other financial institutions can easily copy those new authentication methods.

However, according to the experts, the capability of the guardian is assumed to have more influence on target selection, over the years. It is noticed, that the capability of the guardian can both be expressed by the technical defense measures adopted by the valuable financial institutions, and the awareness of their clients. Because many of their clients have been attacked and became familiar with the risk of financial malware. The experts expect, that due to the increasing awareness and the sophisticated defense measures, big financial institutions become less interesting to target. Therefore, small financial institutions are targeted more intensively nowadays. In addition, the development of more sophisticated malware made it (economically) interesting to target smaller financial institutions.

Furthermore, the experts mention characteristics of the context of the country. Which according to them, have influence on both the suitability of the financial institution, and the capability of its guardian. From the expectation of the experts, multiple hypotheses have been developed that can be tested with the Zeus dataset. Based on the availability of data and time, some of these hypotheses are selected, and processed in order to test them. Multiple statistical analyses are executed, to test the relation between the number of clients of a financial institution, and the encountered attack intensity. Besides, this relation is analyzed on country level. Finally, the evolution of this relation is researched, by executing the analyses per year of the Zeus dataset.

Ten SEPA countries are selected for testing the relation between the number of clients and the encountered attack intensity. According to the quantitative analysis, between 2009 and 2013q1, financial institutions with more clients encountered a higher attack intensity. However, when analyzing per country, in only three of the ten selected SEPA countries, financial institutions with more clients are targeted more. In addition, when analyzing per year, it seems that only in the year 2009 this relation holds. For the years 2010, 2011 and 2012, no significant relation has been observed, between the number of clients and the encountered attack intensity. Furthermore, in the first nine weeks of 2013, financial institutions with less clients are targeted more.

By comparing the ideas of the experts, with the outcomes of the hypotheses, the outcomes of the hypotheses are placed within a broader context. It is noticed, that the size of a financial institution can be expressed by more factors than the number of clients. Moreover, the financial institutions offer many services, which all have their own number of clients. Therefore, the number of clients is a broad concept itself. Finally, the experts mention that the context of the country influences the attack intensity, while the quantitative analyses focused on the relation between the number of clients and the attack intensity.
# Table of contents

Acknowledgements ........................................................................................................ 2
Summary .......................................................................................................................... 3
Table of contents ............................................................................................................. 5
List of tables ................................................................................................................... 8
List of figures ................................................................................................................. 8

1. Introduction .................................................................................................................. 10
   1.1. General background on cybercrime within the financial sector ....................... 10
   1.2. Goal of the research and the research gap ....................................................... 11
   1.3. Previous research and the available Zeus dataset ........................................... 12
   1.4. Research questions ............................................................................................ 13
   1.5. Critical reflection of the Zeus dataset .............................................................. 14
   1.6. Research approach ............................................................................................ 15
   1.7. Scientific and societal relevance ....................................................................... 17
   1.8. Thesis outline ..................................................................................................... 18

2. The background ............................................................................................................ 19
   2.1. Online services of financial institutions ......................................................... 19
   2.2. Consequences of cybercrime ............................................................................ 20
   2.3. Financial malware and the realm of cyber-attacks in the financial sector .......... 21
      2.3.1. Zeus malware .............................................................................................. 22
      2.3.2. The development of Zeus malware ............................................................ 25
      2.3.3. State of the art malware .............................................................................. 26
   2.4. Actors within the financial malware scheme ................................................... 26
      2.4.1. The malware scheme .................................................................................. 27
      2.4.2. The cyber criminals and their business models .......................................... 29
   2.5. Cyber risk management among financial institutions ....................................... 30
      2.5.1. The development of cyber security and cyber risk management in general .................................................................................. 30
      2.5.2. Cyber risk management within the financial sector .................................... 31
      2.5.3. Cyber measures to deal with financial malware ...................................... 33
   2.6. Summarizing the essential parts for the continuation of this research ............... 34

3. Methodology ................................................................................................................. 36
   3.1. A rationale of cyber criminals, the Routine Activity Theory ............................ 36
   3.2. The consultation of experts .............................................................................. 40
   3.3. The Zeus dataset and data preparation .............................................................. 41
3.4. Quantitative analysis ........................................................................................................... 41
3.5. Comparing the outcomes of the two methods ........................................................................ 41
4. Target selection as assumed by experts, in the perspective of RAT ........................................ 42
  4.1. RAT and (Zeus) financial malware ..................................................................................... 42
  4.2. The ideas of experts regarding target selection ................................................................. 43
    4.2.1. The consultation of experts by semi-structured interviews ........................................ 43
    4.2.2. The experts’ ideas of target selection in the perspective of RAT ............................... 46
    4.2.3. The evolution of target selection .................................................................................. 55
    4.2.4. Conclusion ................................................................................................................. 58
  4.3. Hypotheses extracted from experts .................................................................................... 59
    The following hypotheses focus on targeting without pre-selection: .................................. 60
    The following hypotheses focus on characteristics of the financial institution: .................. 60
    The following hypotheses focus on the context of a country: .............................................. 60
    The following hypotheses focus on the evolution of target selection and the development of
    financial malware: ............................................................................................................. 61
  4.4. Selecting hypotheses ........................................................................................................... 61
  4.5. Developing appropriate hypotheses to test with the Zeus dataset ..................................... 63
    4.5.1. Preparation for testing the hypotheses ....................................................................... 63
    4.5.2. The hypothesis to test with the Zeus dataset ............................................................. 64
    4.5.3. The structure of the hypotheses to test ...................................................................... 65
5. Target selection according to the Zeus dataset ........................................................................ 66
  5.1. The Zeus dataset ............................................................................................................... 66
    5.1.1. The database ............................................................................................................. 67
    5.1.2. Selected tables for this research ................................................................................... 68
  5.2. Descriptive analysis ........................................................................................................... 69
    5.2.1. The distribution of targeted domains among different SEPA countries .................... 69
    5.2.2. Targeted domains within the Dutch financial sector ................................................... 72
    5.2.3. Conclusions high-level analysis .................................................................................. 73
  5.3. A metric regarding attack intensity ..................................................................................... 74
  5.4. Profound analysis .............................................................................................................. 74
    5.4.1. Preparing the dataset .................................................................................................. 75
    5.4.2. The structure of the profound analysis ...................................................................... 77
    5.4.3. Generic analysis ....................................................................................................... 78
    5.4.4. Analyzing per year ..................................................................................................... 85
    5.4.5. Financial institutions with multiple targeted domains .............................................. 86
    5.4.6. Conclusions of the profound analysis ....................................................................... 87
6. Comparing the expectations of experts with the patterns shown by the Zeus dataset ............... 91
   6.1. Consistencies .................................................................................................................. 91
   6.2. Inconsistencies ............................................................................................................. 92
   6.3. Conclusions .................................................................................................................. 92
7. Discussion and recommendations ........................................................................................ 94
   7.1. Main findings ................................................................................................................ 94
      7.1.1. Consultation of experts ......................................................................................... 94
      7.1.2. Consultation of experts in the perspective of RAT ............................................... 95
      7.1.3. Quantitative analysis with the Zeus dataset ......................................................... 95
      7.1.4. Comparison between the experts and the Zeus dataset ....................................... 96
   7.2. Limitations .................................................................................................................. 97
      7.2.1. The consultation of experts .................................................................................... 97
      7.2.2. The Zeus dataset ................................................................................................. 97
   7.3. Discussion .................................................................................................................... 98
   7.4. Recommendations ....................................................................................................... 99
      7.4.1. Recommendations for scientific research ............................................................. 99
      7.4.2. Recommendations for the authorities within the financial sector ....................... 100
      7.4.3. Recommendations for the financial institutions .................................................... 100
   7.5. Future research ........................................................................................................... 101
References ............................................................................................................................... 102
Consulted experts .................................................................................................................... 107
Appendices ............................................................................................................................. 108
   Appendix I: Countries within the Single European Payments Area (SEPA) ...................... 108
   Appendix II: Interview protocol ......................................................................................... 110
   Appendix III: The number of botnets targeting domains ................................................ 111
   Appendix IV: The number of clients per targeted domain .............................................. 112
   Appendix V: Graphs to check the assumptions of the linear regression analysis after log
               transformation and deleting outliers ..................................................................... 121
   Appendix VI: Linear regression without 2013 ................................................................. 122
   Appendix VII: Getting insight into the influence of creating groups and the influence of the
                  outliers into the found pattern ............................................................................. 123
List of tables
Table 1: Online banking fraud rates in the Netherlands .......................................................... 10
Table 2: Exclusion-criteria analysis ..................................................................................... 62
Table 3: Attacked Dutch domains between 2009 and 2013q1 including the number of the first week of the attack ......................................................................................................................... 72
Table 4: Linear regression model information after log transformation .................................. 81
Table 5: The Spearman’s rho test on the targeted domains from all the selected countries .......... 81
Table 6: Five groups with the same reach of 2.2 million clients .................................................. 82
Table 7: Multiple linear regression model information, based on the number of clients and the country .............................................................................................................................................. 84
Table 8: Multiple Linear regression model information based on the number of clients and the year86
Table 9: Targeted domains that have at least one other targeted domain that corresponding to the same financial institution .................................................................................................................. 87
Table 10: Information Wilcoxon-Mann-Whitney rank sum test .................................................. 87
Table 11: Outcomes of the hypotheses .................................................................................... 90
Table 12: Similarities of the outcomes of both analyses............................................................. 91
Table 13: Difference between the outcomes of both analyses ................................................... 92
Table 14: Linear regression information after log transformation without the year 2013 .......... 122
Table 15: Overview number of clients and attack intensity per group, based on groups of the same size ................................................................. 123
Table 16: Overview number of clients and attack intensity per group, with groups of 2.2 million clients .......................................................................................................................... 124

List of figures
Figure 1: Client-server-model adopted from (Wikipedia, 2015) ................................................. 20
Figure 2: The Zeus financial malware timeline 2006-2014 (extracted from s21sec.com) ................. 26
Figure 3: A structured overview of a simplified financial malware scheme (ecosystem), adapted from Lucas (2015) ......................................................................................................................... 28
Figure 4: Conceptualization of Cyberspace, adaptation from van den Berg et al. (2014) ............... 31
Figure 5: RAT framework adopted from Cohen and Felson (1979) .............................................. 37
Figure 6: RAT applied to Zeus malware adapted from Cohen and Felson (1979) ......................... 43
Figure 7: The number of targeted SEPA domains per week per country ..................................... 70
Figure 8: The number of targeted SEPA domains per week per country, without the four dominant countries .......................................................................................................................... 71
Figure 9: Domain weeks per number of botnets with a reach of 0 – 66 botnets .......................... 71
Figure 10: The cumulative number of botnets attacking domains of the Dutch financial institutions between 2009 and 2013q1 per week .......................................................................................... 73
Figure 11: Overview of the structure of the generic analysis ....................................................... 77
Figure 12: Overview of the structure of the analysis per year ...................................................... 77
Figure 13: Graphs to check the assumptions of normality and homoscedasticity of the linear regression analysis .......................................................................................................................... 79
Figure 14: Linear regression model after log transformation .......................................................... 80
Figure 15: The average attack intensity per group per way of creating groups, including trend lines 83
Figure 16: Targeted domains per category of clients per year ..................................................... 85
Figure 17: Domain weeks per number of botnets with a reach 10 – 66 botnets ........................... 111
1. Introduction

1.1. General background on cybercrime within the financial sector

Financial institutions become increasingly dependent on information technology and telecommunications to deliver services to consumers and business every day. For instance, banks introduced many platforms through which transactions could be done without much effort, in order to enhance their client base (Vrancianu and Popa, 2010). Although financial institutions are continuously dealing with securing those modern payment services, online financial service fraud is a serious problem. Anderson et al. (2013) mention that global losses due to financial fraud are in the billions of Euros. In 2011 United States financial service cyber fraud losses were $388 billion (Bresiger, 2013). The annual reports of the Dutch Baking associations (NVB), mention the fraud rates of internet banking in the Netherlands. From those reports the following figures are extracted (NVB, 2009-2014), see table 1.

Table 1: Online banking fraud rates in the Netherlands

<table>
<thead>
<tr>
<th>Year</th>
<th>Loss (in millions of Euros)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>1.9</td>
</tr>
<tr>
<td>2010</td>
<td>9.8</td>
</tr>
<tr>
<td>2011</td>
<td>35</td>
</tr>
<tr>
<td>2012</td>
<td>34.8*</td>
</tr>
<tr>
<td>2013</td>
<td>9.6</td>
</tr>
<tr>
<td>2014</td>
<td>4.7</td>
</tr>
</tbody>
</table>

* In the second half of 2012 the fraud rates decreases with 60% compared to the first half of 2012

According to Raghavan and Parthiban (2014), globally banks financial losses due to cybercrime are around USD 114 billion every year. Besides the losses, bank’s costs of cyber risks increase by the security measures in place. The cost spend to combat cybercrimes is globally USD 274 billion per year (Raghavan et al., 2014).

Cybercrime towards financial institutions is particularly aimed at first world countries (Galién, 2014), however, significant differences exist in the level of fraud between those countries. The FFIEC (Federal Financial Institutions Examination Council) supports this line of argumentation, by stating that the level of cyber risk exposure varies significantly across financial institutions (2014). This level is the amount of risk posed by a financial institution’s activities and connections, notwithstanding risk-mitigating controls in place (FFIEC, 2014). Examples of risk-mitigating controls financial institutions currently have in place are; authentication of clients, firewalls and real-time supervision of transactions (Asghari et al., 2014; CPMI, 2014). Those risk mitigation measures, are focused on information security and based on the preservation of Confidentiality, Integrity and Availability (CIA standards) (van den Berg et al., 2014).

According to Moore et al. (2010) information security regularly fails for non-technical reasons. Since the past couple of years there have become a general understanding that the CIA security standards alone are insufficient to combat those attacks (Van den Berg et al., 2014). The problem is that those defense measures focusing on the CIA standards are often reactive, time consuming and public. While, attackers act proactive and investigate the already existing defense knowledge and measures to find possibilities to bridge these defense measures. This can be seen as an iterative loop; usually
attackers find the most vulnerable link, defenders repair the problems, attackers find new possibilities and so a strong dynamic process is at play (Böhme and Moore, 2009; Lagazio, 2014). Furthermore, Solms and Niekerk (2013) argue that cyber security goes beyond the boundaries of traditional information security to include not only the protection of information resources, but also that of other assets. Besides, the Dutch central bank (DNB) concluded in their theme research towards information security (2014) that merely focusing on technology isn’t sufficient. Instead, a risk based and holistic approach is required to deal with the risks of online banking.

1.2. Goal of the research and the research gap
Cyber risks within the financial sector arise not only within the technical context, but also within the socio-technical context (van den Berg et al., 2014). Besides, information security fails regularly due to non-technical reasons (Moore et al., 2010). Therefore, van den Berg et al. (2014) argue that nowadays cyber risk management should both have the technical focus on information security as the focus on the risks that have emerged in the socio-technical context. These concepts can be applied to the environment of online financial services, including its cyber security needs.

For example, many financial transactions are executed via platforms (e.g. mobile application and (mobile) websites) which are connected through the internet with the financial institutions’ servers. In here, the platform together with the internet and the servers deliver the technology that makes it possible to execute the financial transactions. Technical cyber security measures that the platform could adopt, to reduce cyber risks include; two factor authentication, encrypted layer, transactions monitoring, anomaly detection, etc.

Besides, the financial transactions are not only executed on the technical level, but they take place in a broader socio-technical context. This context is constituted by the organizations and human actors that are involved in the transactions. Cyber security measures that can be adopted by financial institutions and authorities to reduce risks in the socio-technical context of the financial transactions are; getting a better understanding of the socio-technical risks, informing the human actors that are involved in the transactions with the socio-technical risks, including more authentication methods, etc. Furthermore, the technical systems to enable financial transactions, including the socio-technical context should be organized. Therefore, both the technical as the socio-technical context are governed in complex ways by many different human actors and organizations (Van den Berg et al., 2014). Examples of the actors and organizations involved in the governance of financial transaction are: Financial institutions, Central banks, payment organizations, lobby organizations, policy makers.

Both the measures in the technical context and the socio-technical context, including the governance of those contexts, need to develop constantly, to deal with the risks of the evolving sophisticated cybercrimes within the financial sector. This research intends to contribute to the development of securing online financial services. Particularly, towards cyber security development within the Single Euro Payments Area (SEPA). SEPA stands for a European Union (EU) payments integration initiative. SEPA arose from the initiative to bolster the common currency, by developing a set of harmonized payment schemes and frameworks for electronic euro payments (EPC, 2016). The jurisdictional scope of the SEPA schemes currently consists of the 28 EU Member States plus Iceland, Norway, Liechtenstein, Switzerland, Monaco and San Marino (EPC, 2016). Appendix I, shows a list of all the countries involved in SEPA.

Authorities within the financial sector are dealing with the governance of cyber risk management (DNB, 2015; ECB, 2013; Europol, 2015). Thereby, they are executing measures and developing new regulations like; (1) FI-ISAC, which is an industry forum for collaboration on critical security threats
facing the global financial services sector (Enisa, 2008, fsisac.com, 2015), (2) Regulations regarding IT risks among financial institutions (ECB, 2014) and (3) oversight of information security regarding IT risks among financial institutions and payments (DNB, 2014; ECB, 2013; SecuRe Pay Forum, 2013).

The current focus of cyber governance within the European financial sector, is on the cyber resilience of financial market infrastructures (CPMI, 2014; DNB, 2015). According to the Committee on Payments and Market Infrastructures (CPMI), it is acknowledged “that cyber resilience is not just about information and communication technologies. Rather, it has a broader impact and relevance for these organizations. Four general areas are covered in a governance framework: people, technology, processes and communication” (CPMI, 2014). However, less knowledge exists regarding the level of cyber risks, encountered by different financial market infrastructures and financial institutions. To further develop cyber risk management, it is essential to have a clear understanding of the threat landscape and the trends of cybercrime. Therefore, this thesis intends to fill the following gap: A limited understanding of why certain financial institutions are more likely to be selected as target by cybercriminals than others.

This research intends to contribute to the understanding of target selection within the Single Euro Payments Area (SEPA). Thereby, empirical insights are provided, into both technical and socio-technical characteristics of financial institutions, which make them more likely to be selected as target by financial malware schemes. Currently, less empirical knowledge exists regarding these characteristics. These insights should support both the financial institutions with their cyber risk management and the authorities with developing cyber security criteria, on order to reduce the societal losses of cybercrime.

For example, when according to this research, financial institutions with sophisticated counter measures like transaction monitoring and anomaly detection are selected less frequent as target, financial authorities could make those measures accessible for each financial institution. This regards a measure within the area technology of cyber resilience. Another example regards the process part of cyber resilience; If this research finds that financial institutions, where employees or not allowed to connect their own devices to the network, are selected less frequent as target. Each financial institution could reconsider their risk acceptance level of connecting own devices to the network. Moreover, financial authorities could set criteria that limits the number of own devices connected to the network of financial institutions.

In this thesis, target selection regarding financial malware is researched. Thereby, target selection is defined as: Attack choices by the financial malware schemes regarding which financial institution to attack, at which point of time and for how many weeks.

1.3. Previous research and the available Zeus dataset
The FFIEC mentions, that the level of cyber risk varies significantly across financial institutions (2014). Furthermore, Tajalizadehkhooob et al. (2014) describe, that significant differences exist in the level of online banking fraud between countries around the globe. Tajalizadehkhooob et al. (2013) researched some characteristics that can explain the difference in online banking fraud levels between countries and financial institutions. That research provides insight into the relation between countries their GDP rate and broadband penetration and the amount of cyber-attacks financial institutions encounter. Furthermore, the relation between the size of a financial institution, whether or not financial institutions having the English language as an option in their banking webpages, and whether or not the domain name of the financial institution contains the word “bank” and the attack intensity institutions encounter. Those insights are elaborated in chapter 4.2.
With their findings, the researchers proved that it is possible to extract intelligence from the files that Zeus financial malware uses as instructions for their operations. Furthermore, they created the Zeus dataset that consists of the time of attack, the target of attack and the attacker. Moreover, the dataset contains a reliable metric of relative attack intensity encountered by the targets. With this dataset, the researchers partially deal, with one of the main challenges in researching cybercrime. Namely, the limited collaboration between the financial institutions and the authorities, on the issue of compiling data to identify trends of cybercrime (Raghavan et al., 2014).

The mentioned proof itself is really interesting, and together with the dataset many research opportunities are provided. For instance, it can be used to identify trends of target selection, based on factors of financial institutions like geographical location and size. However, due to time constraints they only provide high-level intelligence regarding the attractiveness of different financial institutions and the distribution of attacks among them. This thesis intends to exploit the possibility of extracting intelligence by involving more cyber security experts and adding new data to the dataset. Hereby, the focus is on attacked financial institutions within the Single Euro Payments Area (SEPA).

1.4. Research questions

The goal and the research gap described in chapter 1.2, together with the previous research and the available Zeus dataset described in chapter 1.3, result in the following main research question:

What are the expectations of experts, regarding the evolution of target selection by financial malware schemes and how did target selection actually evolve according to the Zeus dataset?

The intention is, to compare the perspectives of experts regarding target selection, with target selection that is actually made according to the Zeus dataset. Hereby, the two different sources could supplement each other, however the expectations of experts are subordinate to the patterns shown by the Zeus dataset. Therefore, the Zeus dataset can be used to test the experts’ expectations. To answer the main research question, first the underlying questions have to be addressed;

1. What is financial malware, how does it work and in what way can target selection be expressed?
   a. What is the influence of financial malware in the realm of cybercrime in the financial sector?
   b. What does the financial malware scheme look like?
   c. How has cyber risk management among financial institutions developed, especially regarding malware attacks?

   Deliverable I: The background on financial malware

2. How has target selection evolved, according to the experts from the field?
   a. What are the main ideas of experts, about target selection by financial malware schemes?
   b. What hypotheses can be developed, based on the experts’ expectations?
   c. Which hypotheses are appropriate for further analysis regarding this research?
   d. What data should be added to the Zeus dataset in order to test the hypotheses?
   e. What hypotheses can finally be tested with the Zeus dataset?

   Deliverable II: Target selection and its evolution as assumed by experts, including hypotheses to test with the Zeus dataset
3. How did target selection actually evolve, according to the Zeus dataset?
   a. Which quantitative analyses are appropriate to test the hypotheses?
   b. What insights are provided by the analyses?

   *Deliverable III: Insight into the target selection that is actually made according to the Zeus dataset.*

4. Is target selection including its evolution as assumed by experts, consistent with the patterns shown by the Zeus dataset?
   a. What are the consistencies, between the experts and the Zeus dataset?
   b. What are the inconsistencies, between the experts and the Zeus dataset?
   c. What is the meaning of those (in-) consistencies?

   *Deliverable IV: Discussion and conclusions, regarding the (in-) consistencies between the assumptions of experts and target selection that is actually made, according to the Zeus dataset.*

Before the research approach is elaborated, first paragraph 1.5 critically reflects on the Zeus dataset. Since the Zeus dataset is used as the ground truth regarding target selection, it is necessary to provide insight into what the data really explains, and what the shortcomings of this dataset are.

### 1.5. Critical reflection of the Zeus dataset

For this research, the Zeus dataset is the most objective available data source that can explain target selection, by financial malware schemes. This dataset is used to test the hypotheses that arise from the experts’ ideas regarding target selection. Which means, that a certain hierarchy exists between the Zeus dataset and the expectations of experts. This paragraph describes, why the dataset provides valid insights into target selection. Furthermore, also the shortcomings of the dataset are mentioned, including the assumptions that are made.

Zeus malware can be used to exploit clients of financial institutions in multiple ways. The most common exploits are; 1) monitoring web traffic, 2) taking screenshots, 3) recording keys struck on the keyboard and mouse clicks and 4) modifying web pages (Mooiman, 2016). The latter one, is used in this research to determine target selection. Modifying web pages, is enabled by sending web injects of a financial institution’s web domain, to the infected computers. Web injects are malware configuration directives that are used to inject rogue content into the Web pages of institutions websites, to steal confidential information from the institution’s clients (Klein, 2011). The raw data which underlies the Zeus dataset, consists of such configuration directives, belonging to Zeus malware between 2009 and 2013q1.

These configuration files contain the domains that are targeted by web injects. Due to the previous work of Tajalizadekhoob et al. (2013), the Zeus dataset provides insight into the targeted domains on a weekly basis. The description above reveals some shortcomings of the Zeus dataset. First, the dataset contains only web injects, for the purpose of Zeus financial malware. While many other malware families exist. Besides, it only deals with Zeus malware between 2009 and 2013q1. Furthermore, Zeus malware can be used in multiple ways to target clients of financial institutions, while the configuration files only provide insight into the purpose of modifying web pages. In addition, it could be the case that the configuration files still contain web injects, which aren’t effective anymore.
Notwithstanding these shortcomings, it is assumed that the Zeus dataset is representative for the purpose of target selection by financial malware schemes. Because, Zeus is known as a persistent malware family, which is one of the most dominant malware families that ever existed (Europol, 2014; Tajalizadehkoob et al., 2013). Besides, Zeus was one of the first developed financial malware families that becomes very popular (Lucas, 2015). In addition, among the four common ways of exploiting clients with Zeus financial malware, inject code can be used to really attack the financial institution, on system level. Although most of the configuration files probably contain ineffective web injects, up to date web injects are required continuously for successful attacks. Therefore, many web injects within the configuration file will be effective.

Summarizing, the dataset has its shortcomings, however it is assumed that it can provide more precisely insights into target selection than the experts. Therefore, the Zeus dataset is used as a ground truth dataset for this research. A more extensive description of the Zeus dataset is provided in chapter 5.

1.6. Research approach

In order to answer the research questions mentioned in the previous paragraph, a research approach of mixed methods is developed. Consisting of both qualitative- and quantitative research methods. The main goal of the research, is to deliver empirical insights into both technical and socio-technical characteristics that influence the likeliness of a financial institution, to be targeted by financial malware. Hereby, the research focusses on the confrontation between the expectations of experts and the patterns shown by the Zeus dataset.

This research consist of four different phases; (1) a literature review to provide the background of financial malware, the financial malware scheme and the development of cyber risk management by financial institutions to deal with the malware, (2) an elaboration of the evolution of target selection as assumed by experts in the perspective of the Routine Activity Theory (RAT), (3) executing quantitative analyses with the Zeus dataset, to provide insight into target selection that is actually made between 2009 and 2013q1 (4) comparing and discussing the outcomes of the experts’ consultations with the outcomes of the quantitative analysis with the Zeus dataset. From the second phase of this research, just target selection is mentioned, which refers to attack choices by the financial malware schemes, regarding which financial institution to attack, at which point of time and for how many weeks.

Phase 1: Providing the background of financial malware risks

The first phase exists of a literature study, to provide the background of the research. This background provides a better understanding of the problem area. First the process of executing online services is mentioned. Subsequently the consequences of cybercrime on these services is described. Furthermore, the literature study focusses on the influence of financial malware in the realm of cybercrime within the financial sector. Moreover, in this phase the financial malware scheme is described. Besides, phase contains a description of the development of cyber risk management within the financial sector and the role of global authorities. Finally, this phase ends with a summary of the parts of the background that are required to understand, for the continuation of the research.
Phase 2: The evolution of target selection as assumed by the experts, in the perspective of RAT

The second phase is executed with qualitative research methods. In this phase the consultation of experts takes place. Those consultations are executed to gather experts’ assumptions and expectations regarding target selection, including the evolution of it. The experts are asked to make an educated guess about target selection for the purpose of financial malware attacks. Hereby, the output from multiple experts is combined. Moreover, the consultation of experts is an additional source of knowledge to the scarce amount of scientific literature, about target selection for the purpose of financial malware attacks.

Before the elaboration of the interviews, this phase starts with the description of the routine activity theory (RAT). This theory provides a rationale regarding why and when cybercrime arises. For the purpose of this thesis, RAT is used to argue and understand why certain financial institutions are more likely to be targeted. In this phase the assumptions of experts are connect with the RAT theory. Hereby, RAT functions as an instrument to describe the evolution of target selection and to place the assumptions of experts in a theoretical perspective.

Subsequently, the assumptions regarding (the evolution of) target selection are gathered by literature research and the consultation of experts. Within this research the consultation of experts has a crucial role. Because it provides ideas about why certain financial institutions are more likely to be selected as target than others.

The expectations from most of the experts are collected through semi-structured interviews. Which are suitable to gather perspectives of experts (Gillham, 2000). However, three of the experts are providing daily support to the research, and are consulted many times, those meetings differ in structure. The experts consulted, are from different positions and organizations in the field; Cyber Intelligence unit from The Dutch Central Bank (DNB), IT-oversight from DNB, PhD economics of cyber security at the technical university of Delft, Security specialist from ABN AMRO, PhD financial regulation at the University of Amsterdam, and head information risk management and CISO at Binckbank.

Subsequently, from the expectations and assumptions of experts, hypotheses are developed. Those hypotheses could be tested. Due to time constraints and the limited availability of data, not all of these hypotheses are tested within the quantitative analysis. Therefore, this phase includes an exclusion-criteria analysis to decide which hypotheses are suitable for this research. Finally, in this phase, the selected hypotheses are processes, in order to create appropriate hypotheses to test with the Zeus dataset.

Phase 3: Analyzing target selection that is actually made according to the Zeus dataset

This phase, quantitatively analyzes target selection according to the Zeus dataset. Hereby, the Zeus dataset is used as proxy, to determine target selection. The Zeus dataset is used as ground truth data that can confront the ideas of experts regarding target selection.

First, descriptive analyses are executed, to provide a high-level overview of the distribution of Zeus malware attacks, among financial institutions within the Single Euro Payment Area (SEPA). The distribution of attacks is defined by: which institutions were under attack in which weeks and by how many botnets. An important remark is that those descriptive analyses view the “intention of attacks”, it doesn’t say anything about the success rate of those attacks.
Subsequently, profound analysis is executed to get a better understanding of target selection, made by the Zeus malware schemes. Hereby, the hypotheses that are selected in the previous phase, are tested. Besides, financial institutions which are targeted relatively intensive, are analyzed. Those profound analyses focus on targeted domains from a selection of SEPA countries (Austria, Belgium, Bulgaria, Finland, France, Hungary, Ireland, Norway, Romania, and the Netherlands). The selection process is shown in chapter 3.

To be able to test the hypotheses, multiple factors are chosen based on which the financial institutions can be distinguished. The data regarding those factors has to be added to the Zeus dataset. Moreover, a metric is described which shows the relative attack intensity encountered by financial institutions and can be used to test the hypotheses.

**Phase 4: Comparing the expectations of experts with the patterns shown by the Zeus dataset**

Within the final phase, the (in)consistency between the experts’ expectations and the outcomes of the quantitative analyses is discussed. This phase is added to the research, in order to provide meaning to the outcomes of the hypotheses. For this purpose, the outcomes are placed within a broader context, to be able to turn the results into findings. Thereby, three significant issues are taken into account. First, the difference between the scope of the qualitative analyses and the scope of quantitative analyses. Hereby, the qualitative analysis focus on target selection regarding financial malware in general. While the data used for the quantitative analyses is from Zeus malware attacks between 2009-2013q1.

Secondly, the hypotheses that are tested with the Zeus dataset are subject to multiple transformations, in order to create appropriate hypotheses to test with the Zeus dataset. Therefore, the outcomes of the tested hypotheses cannot directly be reflected on the ideas of experts. Finally, the Zeus dataset has its own shortcomings. Which means that the outcomes of the quantitative analyses also have their limitations. For those reasons, the expectations of experts that are not consistent with the data, shouldn’t be directly discharged. Vice versa, hypotheses that are consistent with the Zeus dataset, do not directly indicate that the experts are perfectly right regarding target selection.

**1.7. Scientific and societal relevance**

The societal relevance and societal value of this project, are provided by the gathered insights into both technical and socio-technical characteristics, which make financial institutions more likely to be targeted by financial malware. Furthermore, the insight into these characteristics, provide potential benefits for both global authorities within the financial sector as for financial institution including their clients. The financial institutions and their clients will have more insight into the threat landscape of financial malware. Thereby, they should be better in dealing with the risk of financial malware. In addition, global authorities will be better informed about how to organize risk management regarding financial malware. This could support in addressing the societal losses due to financial malware.

The scientific relevance of this thesis consists of the empirical insights into both technical and socio-technical characteristics of financial institutions. Which could explain the differences within the risks encountered by financial institutions from financial malware. Those characteristics will add value to the current known characteristics like website languages of the financial institution and the GDP and
broadband penetration of the home country of the financial institution. Moreover, this research exploited the possibility of extracting intelligence from the files that financial malware used as instructions for their operations. Furthermore, the ideas of multiple experts, related to the Dutch financial sector, are combined and placed within the perspective of the Routine Activity Theory. The outcomes of the qualitative analyses, contribute to the usability of this theory within cyber space.

1.8. Thesis outline
Chapter 1 described the introduction to the problem of financial malware attacks. Besides the goal and the research gap are mentioned, also an overview of the research questions is provided. In addition, the Zeus dataset, which is fundamental to this research, is critically discussed. Finally, the research approach is elaborated.

Subsequently, chapter 2 starts with a brief overview of the history of cybercrime within the financial sector, including the consequences of cybercrime on this sector. Moreover, financial malware and particularly Zeus malware is discussed. In addition, the actors within the financial malware scheme are described and some of their business models are mentioned. Furthermore, this chapter discusses the development of cyber risk management in the financial sector, particularly focused on the mitigation of financial malware risks. Finally, a summary is provided that includes the minimum required knowledge for the continuation of the research.

Chapter 3 describes the methodology of the research. This chapter clarifies how the phases two, three and four are executed. Hereby, the Routine Activity Theory is explained, the consultation of experts is discussed, the Zeus dataset is elaborated, and the methods for quantitative analyses are described. Finally, the issues, of comparing the expectations of experts with the outcomes of the hypotheses, are taken into account.

Chapter 4 starts with a description of the routine activity theory (RAT) applied to cyber space and financial malware. Subsequently, the structure of the expert’s interviews is mentioned and the outcomes of the interviews are elaborated by placing them in the perspective of RAT. Furthermore, the hypotheses extracted from these outcomes are described. In addition, hypotheses are selected that can be tested in this research, this is done by executing exclusion-criteria analysis. Chapter four is finalized by processing the hypotheses, in order to make them appropriate to test with the Zeus dataset, including the available additional data.

Chapter 5, first discusses the Zeus dataset and the additions and adaptations to it, for the purpose of this research. Subsequently, descriptive analysis are executed to provide high-level insights into the distribution of Zeus malware attacks among financial institutions within SEPA. In addition, profound analysis are executed. Hereby, the hypotheses are tested and the financial institutions, which are targeted relative intensively, are taken into account.

In chapter 6, the outcomes of the qualitative analysis are compared with the outcomes of the quantitative analysis. This chapter discusses the consistencies and inconsistencies between the expectations of experts and the patterns shown by the Zeus dataset, by placing them in a broader context. Furthermore, these outcomes are fed back to RAT.

Finally, in chapter 7 the main findings of the research are discussed, including the limitations and the recommendations for scientific research, the financial authorities and the financial institutions. Furthermore this chapter contains a discussion and recommendations for future research.
2. The background

One of the first significant computer frauds in the financial sector that can be found on the internet, dates back to 1988, when an employee of the Chicago Bank gave out the confidential codes used to authorize transfers from the United Merrill Lynch and Brown-Forman accounts (Sector, 1988). In the years till 1995, computer frauds in the financial sector appeared only occasionally. However, due to the advent of the internet and widely utilization of it in the financial sector, from the year 1990 financial institutions started providing services online (Claessens, 2002). Furthermore, from around 1995 financial institutions started providing many services online. As a consequence, financial institutions encounter online fraud on a regular basis (Anderson et al., 2008). Nowadays, those committed online frauds are known and classified as one of the many kinds of cybercrime. The annual global financial losses of financial fraud activities are in the billions of euro (Anderson et al., 2012; Raghavan et al., 2014).

To assess the risks of cybercrime financial institutions encounter, it is essential to understand what kind of cyber-attacks cause the current losses. This chapter first discusses online banking to understand the environment where cyber risks can occur. Secondly, the different cost categories of cybercrime are mentioned. Subsequently, the most popular cyber-attacks within the financial sector are elaborated, where the influence of financial malware is introduced. Moreover, Zeus malware has been elaborated. Furthermore, the different actors that play a role regarding financial malware are mentioned. This part includes the business model of those criminals and results into a description of the financial malware ecosystem. Besides, the development of cyber risk management is elaborated. Finally, the parts of the background that are specifically used in the continuation of the research are summarized.

2.1. Online services of financial institutions

All kinds of financial services can be executed online nowadays, like; payments, insurances, stock investments, private investments, etcetera. Regarding the financial malware taken into account in this research, especially online payments are required to understand. In this paragraph, the platforms used by financial institutions to provide online payments are discussed. Financial institutions have many platforms in place through which transactions could be done without much effort (Claessens, 2002; Vrancianu and Popa, 2010). Based on the client-server-model (see figure 1) it can be clarified how this works. The platform consists of the following three main components:

1. Financial institutions’ clients (client side which connect to financial institutions via network infrastructure and by personal computers, mobile phone, tablet etc.)
2. Network infrastructure (Internet)
3. Financial institution (server side)

A client of the financial institution connects with the computer, mobile phone or tablet to the Internet, and then starts communicating with the server of the financial institution via an application or the web browser. The focus of this thesis is only on the financial services executed via the computer, because the Zeus dataset used for the research doesn’t include the mobile variant of Zeus financial malware.

For that purpose, the client first visit the home page of the financial institution and subsequently request the login page of the institution’s website. When the connection has been made with the
login page, the client has to go through an authentication process to gain access to her account details and other financial services. Subsequently, when executing one of the financial services the client has to go through an authentication process again.

Figure 1: Client-server-model adopted from (Wikipedia, 2015)

2.2. Consequences of cybercrime

The economic consequences of cybercrime are tremendous though hard to quantify (van Eeten, 2009). According to Anderson et al. (2013), the huge costs of cybercrime financial institutions suffer can be separated into three main categories. Those categories are described below, adaptation of the ideas from Anderson et al. (2013) and Lagazio al. (2014):

- **Direct losses** like monetary losses can be caused by; fraud; legal costs; recovery and clean-up cost; regulatory fine; loss of client accounts; and loss of client trust/loyalty.

- **Indirect losses** are the monetary losses and opportunity costs imposed on organizations and society when a cybercrime is carried out, no matter whether successful or not. Examples include: opportunity cost of reduced sales; cost of recovering unanticipated damage to infrastructures; overall reputational damage extending beyond a company’s own clients; competitive disadvantage due to intellectual property (IP) thefts.

- **Defense costs** are direct defense costs of development, deployment and maintenance of cybercrime measures and indirect defense costs arising from inconvenience and opportunity costs caused by the defense measures. Examples of defense costs are: cost of security measures; security services provided to individuals and industry, such as training and awareness measures; law enforcement.

The direct- and indirect costs of cybercrime seem to be the tremendous consequences for the financial institutions. However, the most significant consequences of cybercrime are the large defense costs (Raghavan et al., 2014; Anderson et al., 2013). Those large defense costs are probably pushed by the strategic position financial institutions pursue by implementing better security measures (than their competitors). According to Lagazio et al. (2014), this is a strategic position driven by the belief that competing in security measures can increase client trust and loyalty and protect their assets in case of a successful cyber-attack.
In the extension of the reputational damage financial institutions suffer, the issue of under-reporting cybercrime arises. More reports of cybercrime lead to more reputational damage, therefore financial institutions under-report the cybercrime incidents (Anderson et al., 2008). As a consequence, the impact of cybercrime is underestimated by the government. Lagazio et al. (2014) mention that this “lack of awareness about the real extent of cybercrime leads to lower “state effort to reduce cybercrime”, fewer “global enforcement measures”, lowers “organized nature and specialization in policing cybercrime” and increased “jurisdictional arbitrage”. All these will hinder “effective policing of cybercrime” and lead to growth of “cybercrime incidents””. This research intents to deal with that lack of awareness as well by showing the Zeus malware attacks that are really made (see chapter 5).

2.3. Financial malware and the realm of cyber-attacks in the financial sector

One of the main risks in the financial sector is identity theft, which can occur by using a cyber-attack like phishing, key logging, financial malware like Trojans horses or other malicious codes (Vrancianu et al., 2010; Choo, 2010). The most commonly used attacks for identity theft by cybercriminals are phishing and Trojan horses (Anderson et al., 2008; Vrancianu et al., 2010; Adham et al., 2013). Via phishing, victims are lured by an email to log on to a website that appears genuine but is actually designed to steal their passwords (Anderson et al., 2008). A whole different kind of attack is through Trojan horses, which is malicious code (malware). According to Anderson et al. (2008), the common usage is that a Trojan is a program that does something malicious (such as capturing passwords) when run by an unsuspecting client. Another major risk is represented by the authentication schemes currently in use, that base their robustness on the end-client’s decisions, which making them entirely vulnerable to social engineering attacks (Vrancianu et al., 2010).

Many financial institutions have realized that phishing and financial malware together with the authentication schemes currently in use lead to online man-in-the-middle attacks (MitM); which are the most common today (Vrancianu et al., 2010). The cybercriminals typically use MitM to take over a session once the client has authenticated herself to the bank (Anderson et al., 2008). The man in the middle (attacker) pretends to be the bank, thereby receives the client’s authentication. With that authentication he logs in on the client’s bank account (Anderson et al., 2008; Utakrit, 2009). In the Netherlands, such attacks became a serious concern from 2007, when clients of ABN Amro suffered from a MitM attack (Ringlestijn, 2007).

Besides the cyber-attacks that are focused on stealing money from financial institutions via their clients, there are also cyber-attacks that threatens the financial system as a whole (DNB, 2015). Particularly, those attacks are focused on stealing sensitive data and reducing the availability of the digital payment infrastructure (CPMI, 2014). Distributed denial-of-service attacks (DDoS) are used to degrade the systems availability. In a DDoS attack, a cybercriminal directs a flood of illegitimate service requests to overwhelm the targeted computer or network in an attempt to make the resource unavailable to intended clients, thereby seizing control of multiple systems by infecting them with malware (CPMI, 2014). In 2011 there were multiple successful DDoS attacks on the Rabobank (Bakker, 2011; Udo de Haes, 2011). In addition, the risk of attacks impacting the integrity of the software and the data of the financial market infrastructures are increasing as well (CPMI, 2014). In 2011, a targeted attack on the International Monetary Fund (IMF) was executed with the goal to steal sensitive information (Apps, 2011).
However, this thesis only focuses on man-in-the-browser (MitB) attacks which have similarities with the described MitM attacks. Specifically, on Zeus financial malware and its variants that enables MitB attacks.

2.3.1. Zeus malware
The malware identified as the current threats by EU law enforcement can be divided into three categories based on their primary functionality – ransomware, Remote Access Tools (RATs) and data stealers (Europol, 2015). This research deals with the Zeus malware family, which can be categorized as data stealer (Europol, 2015). Zeus is a Trojan horse that is able to act as a man-in-the-browser (MitB), which has some similarities to the man-in-the-middle attack. A MitB Trojan operates in the internet browser that is displayed on a client’s screen and controls the ingoing and outgoing content of information at the system level (Utakrit, 2009). Zeus malware has been primarily known for its use in financial fraud (Adham et al., 2013). According to Europol (2015), the threat level of Zeus malware is high. In this paragraph is described how Zeus financial malware works and how it developed itself into more sophisticated versions. The Zeus dataset used for this research exists of Zeus and the three variants ICE X, Citadel and Power Zeus (the peer-to-peer variant).

How Zeus works
Online financial service fraud by Zeus financial malware can be explained by the following six steps (FBI, 2010); (1) malware coding, (2) infecting the victim, (3) intercepting financial service credentials, (4) getting remote access to financial service account, (5) transferring money to mule, and (6) transferring money from mule to criminal organization. Each step is briefly described below.

(1) Malware coding
The technical development of Zeus is executed by the malware coders. Those coders write Zeus software to exploit a computer vulnerability and install a Trojan (FBI, 2010). The developed Zeus kit contains a control panel application that is used to both maintain the botnet and to retrieve/organize recovered information. In addition the Zeus kit contains an EXE builder that is used to create the Trojan binaries and encrypt the configuration files (Macdonald, 2011). Configuration files configure the parameters and initial settings for computer programs (wiki, 2015). Regarding Zeus malware, the configuration files are used as instructions for the computers within the botnet their operations. Those instructions are set by cyber criminals within the financial malware scheme and among others determine which domains of financial institutions to attack. Attacking means, sending web injects of the financial institution’s web domain to the infected computers. This information is used as proxy, to determine target selection.

(2) Infecting the victim
Clients of financial institutions get infected with Zeus malware without knowing. The infection in most of the cases is caused by clicking on spam emails containing links or infected files that start the download (Macdonald, 2011). Sometimes clients are infected during surfing on the internet by drive-by downloads, which is an unintended download of the malware (Microsoft, 2014; Wikipedia, 2015).

Notice that multiple Zeus botnets exist. Macdonald (2011) mentions that “the Zeus toolkit is a commercial product that is sold to many different users, and distributed freely to many more. Each of them can create one or more botnets of their own, so the number of Zeus botnets is likely quite
Each botnet has its own bot executable and RC4 key (password) (Macdonald, 2011). Data sent through the Zeus botnet is encrypted with RC4 encryption.

After the client’s computer is infected, the computer will turn in to a bot. The first step in building a bot executable is to edit the configuration file. The configuration tells the bot how to connect to the botnet, and it also contains information on what user data to gather and how to do so. According to Macdonald (2011), the configuration file exists of the following two parts:

1. **Static Configuration**
   “The StaticConfig is compiled into the bot by the Builder tool. It contains information that the bot will need when it is first executed. To update the StaticConfig the bots must be ordered to download a new bot version”. The available settings are (Macdonald, 2011):
   - The name of the botnet that this bot belongs to.
   - The amount of time to wait between dynamic configuration file downloads.
   - The time interval between uploads of logs and statistical information to the drop server.
   - The URL where the bot can get the dynamic config file.
   - A URL where the bot can check its own IP address, to determine if it is behind a router or firewall.
   - The encryption key that is used to hide information transmitted within the botnet.
   - A language ID list that tells the bot to go into a dormant state if the infected computer’s language is on the list.

2. **Dynamic Configuration**
   Macdonald (2011) mentions that, “The DynamicConfig is downloaded by the bot immediately after it is installed on a victim's computer. This file is downloaded at timed intervals by the bot, and can be used to change the behavior of the botnet. Most of the entries control how information is collected from the infected computer”. Available settings include (Macdonald, 2011):
   - A URL where the bot can download a new version of itself, if the command to do so is given.
   - The URL of the drop server where logs, statistics and files will be uploaded and stored.
   - Information used to inject additional fields into web pages viewed from the infected computer*.
   - A list of URLs where an emergency backup config file can be found.
   - A set of URL masks used to cause or prevent logging of information.
   - A set of URL masks to indicate that a screen image should be saved if the left mouse button is clicked.
   - A list of pairs of URLs that are used to cause redirection from the first URL to the second.
   - A set of URL masks used to collect TAN (Transaction Authentication) numbers - used by some banks for online authentication.
   - A list of IP/URL pairs that are inserted into the infected computer’s hosts file to override DNS lookups.

* Hereby web injects play a significant role. Web injects are malware configuration directives that are used to inject rogue content into the Web pages of bank websites to steal confidential information from the institution’s customers (Klein, 2011).

(3) **Intercepting banking credentials**
Subsequently, the bot starts communicate with the C&C server. When the dynamic configuration file is received, the bot will retrieve the drop server URL from it and HTTP POSTs information about itself to the drop server and puts its status “online”. This drop server component of the Zeus kit has a collection of PHP scripts that allow the bot master to monitor the status of their bots, issue commands to the bots and retrieve the information that are collected by bots. The data stolen by the bots is also sent to the drop server.

Furthermore, the C&C server monitors the user's Web browsing activities (both HTTP and HTTPS) using the browser window titles or address bar URLs as triggers for its attack and sends the logs on average every minute to the server (Wyke, 2011). The bots are able to gather online banking information of all the users of the infected computer. Then, the C&C server starts to communicate with its bot to send the commands or the web injects into web pages of the infected computer.

(4) Getting remote access to banking account

Besides, the criminals are able to insert themselves into an online banking session. The C&C server sends commands to the bot to execute unnoticed transactions during online banking or other unnoticed online interactions with financial institutions. Lawrence (2015) describes this as follows:

“First, they combed through stolen data to identify infected machines that had access to commercial bank accounts, where even large transfers wouldn’t raise alarms. Whenever one such machine logged into a bank website, the hacker could see it on an administration interface that came bundled in the ZeuS kit and piggyback onto the legitimate session. The victim might see on his screen a page citing delays because of maintenance or a box asking for a PIN code or Social Security number, while the hacker used the stolen access to clean out the account. It was a nightmare for bank security; victims had logged themselves in, so the sessions seemed 100 percent legitimate”

(5) Transferring money to mule

Hereby the client’s money is stolen by the criminal actors within the malware scheme. The money is transferred to a money mule. Transferring the money from the client to the mules can be done in multiple ways; 1) All the authentication codes may be available for the criminal, and thereby the criminal is fully authorized to execute transactions 2) Transactions can be added by the criminals on the background of the transaction lists, which are hidden by web injects 3) The transfer of money to the mules is sometimes executed by the banking clients themselves, while strange amounts are noticed within the transaction list. The client inserts the second authentication code, although unknown transactions are visible.

(6) Transferring money from mule to criminal organization

Finally the mule transfers the money to the organization after taking a small percentage. Or the mule withdraws the money, takes a percentage of it and bring the rest of it to the cyber criminals. Thereby, the criminals remain anonymous.

Cyber criminals behind the financial malware scheme can adapt their attack behavior in multiple ways. They can develop more sophisticated malware that circumvents the executed security measures or they can choose to attack other financial institutions, which for instance have weaker security measures. This thesis first provides some insights into the development of financial malware, but the latter way is the main focus of this research. Hereby, the selection of financial institutions to
attack (=target selection) by the financial malware schemes is researched. Regarding this thesis, **target selection** is defined as: *Attack choices by the financial malware schemes regarding which financial institution to attack, at which point of time and for how many weeks.*

### 2.3.2. The development of Zeus malware

Zeus first version appeared in 2006. In May 2011, the source code for the Zeus Trojan horse was leaked on the Internet. Two new major toolkits based on the leaked Zeus source code have become available: ICE IX and Citadel. Later also the peer-to-peer variant of Zeus became available. Those variants of Zeus malware are discussed below, those are also the variants of Zeus that are part of the Zeus dataset. Subsequently, figure 2 provides an overview of the Zeus time line, which clarifies the availability of different Zeus variants.

**Ice IX**

In the August 2011 (see figure 2), Ice-IX a new variant of Zeus became available. According to Europol (2015), “Ice IX is another first generation Zeus variant following the release the Zeus source code, appearing in the same time period as Citadel. Although its use appears to be in decline, several EU Member States have still actively investigated cases of its use”.

**Citadel**

In the end of 2011 Citadel became available. According to Europol (2015), Citadel is a successful variant of Zeus. It was first available for every visitor of the underground market, however later it became limited to certain groups. Citadel is not only focused on financial crime, but is also used for much more targeted (APT) attacks focusing on specific businesses or government entities (Sherstobitoff, 2013; Tamir, 2014). According to Tamir (2014), Citadel has been noted as the first malware to specifically target password management solutions.

**Power Zeus**

Stone-Gross (2012) mentions that one of the most significant developments in financial malware is the private peer-to-peer (P2P) version. This Zeus variant is known as p2p Zeus, Power Zeus and Gameover Zeus, and first identified in October 2011 (Andriesse et al., 2013). Andriesse et al. (2013) mention that this version removes the centralized command and control (C&C) infrastructure previously required to push configurations files, updates and collect information harvested from infected computers.

In P2P Zeus, the centralized C&C server, a single point of failure targeted by researchers and law enforcement is replaced with a robust P2P network (Symantec, 2012). According to Lawrance (2015) “Gameover ZeuS hid the command centers by constantly changing their location on the Internet and diverting traffic through up to 2,000 proxy servers. Gameover Zeus also deployed a decentralized structure: Infected computers could pass commands among each other, peer to peer, rather than each separately communicating with a command server”. Thereby, the distribution of configuration files, the propagation of binary updates and sending stolen data to the controllers are being done through the peers in the network instead of relying on the centralized, vulnerable C&C server (Stone-Gross, 2012).
2.3.3. State of the art malware

This thesis focuses on four variants of the Zeus Malware family; Zeus, Citadel, Ice X and Power Zeus. However, the development of malware are increasing rapidly in scale and sophistication (Lucas, 2015). To get some insights into the development of malware, two newer and more sophisticated versions of malware are described briefly in this paragraph. First the mobile Zeus malware variant Eurograbber is described and secondly a description is given of one of the newest malware variants named Carbanac, which attacks are not limited to financial institutions.

Eurograbber

The Eurograbber attack employs a very successful variation of the Zeus-In-The-Mobile Trojan (Kalige and Burkey, 2012). Kalige and Burkey (2012) mention that, “the Trojan used to attack mobile devices was developed for both the Blackberry and Android platforms in order to facilitate a wide “target market” and as such was able to infect both corporate and private banking users and illicitly transfer funds out of clients’ accounts in amounts ranging from 500 to 250,000 Euros each”.

Carbanac

The newest variant of malware is called Carbanac. Since September 2015, this malware variant targets institutions within the USA and Europe. Before that month in only targeted financial institutions from Russia (Kaspersky, 2015). The September variant not only targets financial institutions, but also Forex trading companies, online casinos and other institutions. The malware can be classified as an APT and works as follows: The cyber criminals send targeted mails including malicious content. When the computer of the target gets affected, the criminals are able to search the network until they found the systems that can be manipulated (Kaspersky, 2015).

2.4. Actors within the financial malware scheme

Financial economic crime is increasingly executed, by organized crime groups and more often executed via the digital world (cyber space). Thereby, the network of stakeholders becomes more
complex (NVB, 2009). According to Europol (2014), relationships between cybercriminals are often transient or transactional and although they may form more coherent, project-based groups, they lack the structure and hierarchy of a traditional organized crime group.

Since 2004, cybercriminal networks start emerging, known as online black markets in which the bad guys trade with each other, with criminals taking on specialized roles (Thomas and Martin, 2006). Related to that, Bauer et al. (2008) describe and Lucas (2015) mentions that an ecosystem around online banking attacks has been arisen. Within this ecosystem malware is only one part of the multiple steps required for a successful financial institution attack. To facilitate those different steps, many services and products are available within the underground economy. Those services support cybercriminals in setting up, running and maintaining a successful attack (Lucas, 2015).

Regarding financial malware, an ecosystem exists of actors that enables malware attacks. In this thesis that ecosystem is defined as a financial malware scheme. This chapter describes the malware scheme and the business models in place by the cyber criminals within the malware scheme.

2.4.1. The malware scheme
The financial malware scheme, is the synergy of actors and organizations that enables financial malware attacks. It is import to understand that within this malware scheme not all players are criminals (Bauer et al., 2008; BITS, 2011). Figure 3, shows an overview of a financial malware scheme (= ecosystem), as provided by Lucas (2015). The different actors of cybercrime within the financial sector can be categorized into four main categories; malicious exploiters, money mules, victims, and security guardians. However, the malware schemes also include the infrastructural parties which are most of the time not aware of criminal activities, like; telecom providers, financial service providers, web shops, etc. The four main categories are described below.

Malicious exploiters (Cyber criminals)
Malicious exploiters can be defined as the cyber criminals. Raghavan et al. (2014) mention that, “those malicious exploiters can be categorized into five sub categories. Innovators (who seek to find security holes in the system to overcome protection measures adopted by the banks). Amateur (who are beginners in this area and their expertise is limited to computer skills, which is exploited by the cybercriminal). Insiders (who are working within the bank to leak out important information in order to take some kind of revenge). Copy cats (they are interested in recreating simple tasks). Criminals (highly organized and very knowledgeable who may use all the above mentioned stakeholders for their own profit)”.

The cyber criminals within the Zeus malware schemes that determine which financial institutions to attack have three different options to get the web injects (= inject code), that are used to manipulate online banking sessions. 1) Different web injects could be developed by the Zeus malware scheme itself. 2) Different web injects could be gathered through the internet, especially via the underground economy. And 3) The Zeus malware schemes that buy Zeus malware, could use the web injects that are already included in the Zeus malware by the malware coders.

In figure 3, the exploit kit, the banking malware, the command and control, the control panel and the inject code, are all developed and created by a malicious exploiter. While some financial malware schemes are creating and executing all the parts by themselves, it can be done by different
exploiters. A much more complex ecosystem exists around all those activities, however that is beyond the scope of this research.

Money mules
Florêncio and Herley (2010) mentioned that the key role of money mules is to convert “reversible traceable transactions into irreversible untraceable ones”. Raghavan et al. (2014) mentions that Money Mules are “individuals recruited wittingly and often unwittingly by criminals, to facilitate illegal funds transfers from bank accounts”. In figure 3 it is shown that the money mule gathers the fraudulent transaction and transfers it further to the attacker, which is the malicious exploiter inhere.

Victims and Targets
Regarding this research, it is important to clarify who the targets of the financial malware scheme are. Although the clients of financial institutions are attacked and affected in the first place, the real targets of the financial malware scheme are the financial institutions. Therefore, in this thesis the clients of the financial institutions that have been attacked are called ‘victims’ and the financial institutions are classified as ‘targets’.

In figure 3, the victim is the client and the Bank is the target.

Security guardians
Security guardians are also one of the main actors in the online banking fraud field. They improve the existing banking systems and their vulnerabilities and develop more sophisticated security systems for mitigating the online fraud on the financial institutions. The security guardians in this problem is either the banking sector itself or third party security firms who provides security services for the banking sector.

In figure 3, the guardians are not shown, however the guardians should interrupt the process shown in the figure. Thereby, the guardian could be together with the bank to monitor fraudulent transactions. Instead regarding Zeus malware, it is really important that the victim is aware of financial malware.

Figure 3: A structured overview of a simplified financial malware scheme (ecosystem), adapted from Lucas (2015)
To get an understanding, why the criminal actors are involved in financial malware schemes, more explanation regarding the business models of the criminals contributing to that scheme is needed. Furthermore, this explanation provides the first focus on target selection.

2.4.2. The cyber criminals and their business models
This paragraph, provides a high-level explanation of the business models of cyber criminals, to understand the involvement of different actors within financial malware schemes. Furthermore, this paragraph provides the first insights into target selection. Hereby, the motives of cybercriminals, to participate within the scheme, are elaborated. However, the non-criminal actors within the malware scheme, are most of the time not aware of their involvement. Therefore, these motives do not regard those parties.

Ultimately, the intent of the cybercriminals is stealing money (Choo, 2011), however cybercriminals have multiple opportunities to get a piece of the stolen money. The supply- and value chain of financial malware can provide insight into the business models of cybercriminals (Kserthi, 2010, p. 193; BITS, 2011). The supply chain in the malware industry encompasses more than just software. It is an elaborate collection of organizations, people, technologies, processes, services, and products (BITS, 2011). The value chain provides insight into the financial flow between those different players (Bauer et al., 2008).

According to Millar (2015), twenty first century supply chains have evolved into worldwide interconnected supply- and demand networks comprising complex operations, with profound interdependencies and exposure to greater volatility in an uncertain world. The linear concept of a chain is therefore no longer valid to describe these complex international networks of suppliers, stakeholders, partners, regulators and clients. Instead, supply chain ecosystems are providing a more holistic view (Millar, 2015). Regarding financial malware, the financial malware schemes can be classified as such an ecosystem (Lucas, 2015).

In the previous paragraph, insight into the malware scheme is provided. The exploit kit, the banking malware, the command and control and the inject code, are all developed and created by a malicious exploiter. Some of the financial malware schemes are creating all these parts by themselves. Subsequently those actors execute it for their own use and keeping it away from other criminals. Hereby they can execute fraudulent transactions or gather banking credentials or other personal information which they can sell in the underground economy.

Those criminals could also sell all these developed parts of financial malware within the underground economy to other criminals. Hereby they can still execute criminal activities themselves, and additionally earn money by selling or leasing the products to other criminals. Finally, cyber criminals could earn money only developing certain parts of financial malware and selling it to other criminals. Hereby the developing criminals stay out of other criminal activities. Due to those available services, it is very easy to execute financial malware attacks. Therefore, many (small) malware schemes exists nowadays.

The criminal services and products are traded illegally within the underground economy, thereby the criminals make use of the Darkweb. Conceptually, the Darkweb is constituted by the darknets,
which include anonymization techniques used in parts of the Internet. This allows cybercriminals to communicate freely without the risk of being traced. These tools are perfectly legitimate for citizens to protect their privacy and are not only used for a criminal intent (Europol, 2014).

The final paragraph of this chapter elaborates on cyber risk management within the financial sector.

2.5. Cyber risk management among financial institutions

This paragraph discusses cyber risk management by financial institutions. First the development of cyber security in general is described. Subsequently, cyber risk management within the financial sector is discussed including its challenges. Furthermore, the role of global authorities is mentioned. Finally, cyber measures that deal with financial malware attacks are described.

2.5.1. The development of cyber security and cyber risk management in general

During the last twenty years, the society has become strongly dependent on IT systems by the creation of cyberspace (van den Berg et al., 2014). According to Friedman (2014), cyberspace of today is almost unrecognizable compared to its early beginnings. For this thesis, Friedman’s definition of cyberspace, as described in Cyber Security and Cyber War, What Everybody Needs to Know (2014), is used:

“Cyberspace is the realm of computer networks (and the clients behind them) in which information is stored, shared, and communicated online”.

The developed cyberspace has enabled wealth in the western countries, and provided many opportunities for both the governments and business in those countries (Choo, 2011). However, the development of cyberspace also created new opportunities for criminals (Anderson et al., 2008; Choo, 2011).

The first years during- and after the creation of cyberspace, security was particularly the concern of the national governments (Anderson et al., 2008). Until about 2004 the main focus on combating cybercrime shifted from governments to companies (Anderson, 2008). In those early years security was focused on information security, which was based on the preservation of Confidentiality, Integrity and Availability (CIA standards) (van den Berg et al., 2014). Inhere the focus was particularly on the technical aspects of information security.

However, since a couple of years, there have become a general understanding that the CIA security standards alone aren’t encompassing. According to van den Berg et al. (2014), “gradually, more business-oriented topics have been adopted in the security standards like ‘cyberspace assets’, ‘supplier relationships’, ‘business continuity management’, and ‘compliance’”. Those business-oriented topics focus more on the socio-technical aspects. To clarify the current intelligence of cyber security inherent risk, it is necessary to get more grip on the differences between the technical focus and the socio-technical focus regarding cyber security. Therefore, the conceptualization of cyberspace by van den Berg et al. (2014) is discussed briefly.

According to van den Berg et al. (2014), the easy client interface of applications on the internet has made it possible that there are so much cyber activities nowadays. These activities include many types of data and information exchange, information search and retrieval, e-watching and -listening, IT-enabled transactions, remote control, cyber protesting and cybercrime, up to cyber warfare (van den Berg et al., 2014).
The developed cyberspace can be conceptualized by three layers. The technology that makes the cyber activities possible forms the technical layer. The cyber activities themselves take place in the socio-technical layer. And finally, those two layers are governed -in complex ways - by a huge variety of human actors and organizations, which creates the so-termed governance layer of cyberspace (van den Berg et.al, 2014). Figure 4 presents the conceptualization of cyberspace, which is divided in several cyber sub-domains. This figure is adapted from (van den Berg et.al, 2014).

Figure 4: Conceptualization of Cyberspace, adaptation from van den Berg et al. (2014)

This thesis focuses on cybercrime within the financial sector. Financial institutions constitute a significant part of the sub domain financial sector (aligned red in figure 4). The next paragraph describes the state of the art literature regarding cyber risk management in the financial sector.

2.5.2. Cyber risk management within the financial sector
Van den Berg et al. (2014) argue that nowadays cyber risk management should both have the technical focus on information security risk management in the technical layer as the focus on the risks that have emerged in the socio-technical layer of cyberspace. Moreover, both the technical layer and the socio-technical layer are governed in a complex way. Therefore the governance of cyber risk management needs attention as well.

Regarding risk management, trade-offs between decreasing the direct- and indirect costs and increasing defense costs need to be made. Although there are more sophisticated measures on the market to combat cyber criminals, investing in those measures is not always the right thing to do (van Eeten, 2014; Moore, 2010). The vulnerabilities of those more sophisticated defense measures will be found too. Security engineers have to defend the whole infrastructure, while cyber criminals only have to find that single vulnerability. Besides, most of the defense systems are one-step behind the tools adopted by cyber criminals (Raghavan et al., 2014). Therefore, cyber risk management is known as an asymmetric warfare.

However, Moore et al. (2010) discus that information security often fails for non-technical reasons. Instead, the trade-offs financial institutions need to make are significantly influenced by the
misaligned incentives and moral-hazards regarding security decisions. In addition, currently there is no clear supervision and/or collaboration between the banks and their global authorities (Moore et al., 2010; Raghavan et al., 2014).

**Misaligned incentives and moral hazard effects**

Moore et al. (2010) mention that cyber security in the financial sector has to deal with a major problem regarding security incentives. In the financial sector those incentives are misaligned, since the organizations that defend the systems in the financial sector do not bear the full costs of failure (Moore et al., 2010). For example, the banks who make security decision regarding online banking don’t suffer the full cost of failure. In other words, the cost and benefits of the banks that make the security decisions are not consistent with the cost and benefits of the society as a whole (van Eeten, 2014).

Bohm et al. (2000) discuss how many banks have seen online banking as a mean of dumping on their clients many of the transaction risks that they previously bore in the days of cheque-based banking. This is a clear example of an externality that causes misaligned incentives. Another example of externalizing risks in the financial world is provided by van Eeten (2014). According to van Eeten, banks are externalizing risks by educating their clients to check the little padlock in the browser window (Bauer and van Eeten, 2009). Besides the misaligned incentives there is another reason for the ineffective cyber risks mitigation in the financial sector, known as the moral-hazard effect.

Anderson (2008) mentions that moral-hazard effects play an important role in information security. According to Carboni et al. (2012), moral-hazard effects occurs when agents behave in a way that benefits them at the expense of a principal. This problem is likely to occur when information asymmetry exists, meaning that agents have more information about transactions than the principal does. For instance, financial institutions take more cyber risks because others bear the risks without knowing.

Anderson (2008) mentions an interesting moral-hazard effect regarding cyber security in the financial sector: “U.S. banks faced a much fiercer liability regime than UK banks, but they actually spent less on security than UK banks did, while UK banks suffered more fraud”. This appears to have been a moral-hazard effect, and was one of the anomalies that sparked interest in security economics, secure systems need properly aligned incentives (Anderson, 2008).

In addition to the mitigation infectiveness, individual financial institutions deal with, there are also cybercrime mitigation deficiencies regarding the supervision of authorities and collaboration between the financial institutions.

**Lacking supervision and collaboration**

Currently, there doesn’t exist a global regulatory framework regarding cyber security in financial sector. Regulatory regimes regarding cyber security diverge among different countries across the globe (Anderson, 2008). The difference in liability between countries already became clear in the previous paragraph. Besides, in countries like the UK and Germany banks can deny liability, which can undermine the incentive to co-operate (Anderson et al., 2008).

Besides the diverging regulations, Raghavan and Parthiban mention that there is insufficient amount of cooperation among the banks across the world (Raghavan et al., 2014). The financial sector should
collaborate with global authorities and watchdog organizations for the development of tools and models which can be applied to counter global banking cybercrimes (Raghavan et al., 2014). Cyber security experts mention that the only way to effectively combat cybercrime, public and private sectors should join forces and combine greater regulation with international action (Apps, 2011). According to Choo (2011), the potential consequences of cybercrime within the financial sector crosses borders.

The scope of both the financial system as the internet itself are typically international. Therefore, International Cooperation is essential. Financial institutions and enforcement agencies should work together and thereby improve information sharing (DNB, 2015). Besides, global authorities should intensify collaboration as well in order to strengthen the resilience to cyber threats. In September 2014 a working group has been set up, comprising of central banks and securities regulators, with DNB acting as co-chair. The objective of the working group is to determine what further steps are necessary to raise the bar across the board for measures to enhance the resilience of individual institutions and for the system as a whole (DNB, 2015).

Furthermore, the Dutch association of banking mentions that financial cybercrime is executed by organized crime groups that are not limited by country borders. Together with the increasing complexity within the actor field, international coordination of perspectives and assembling of knowledge and experiences become essential (NVB, 2009).

2.3. Cyber measures to deal with financial malware

This paragraph describes the development of measures that deal with financial malware, based on the development of cyber security in general, as described previously in this chapter. Therefore, first the technical measures are mentioned. Historically, clients of financial institutions only needed their bank account number and password to get access to their online services. This is known as one-factor authentication and is relatively easy to bypass, since clients often choose weak passwords and can easily misplace their credentials leading to their account being compromised (Kalige and Burkey, 2012). To improve this, the financial institutions added a second authentication mechanism that validates the identity of the client and the integrity of the online transaction. Specifically, when the financial institutions’ client submits an online banking transaction, the financial institutions sends a Transaction Authentication Number (TAN) via SMS to the client’s mobile device. The client then confirms and completes their banking transaction by entering the received TAN in the screen of their online banking session (Kalige and Burkey, 2012). Those authentication methods that use besides the pc also other channels for authentication, are categorized as multi-channel authentication methods.

As described in chapter 2.3.1, notwithstanding the implementation of two factor authentication, the Zeus financial malware scheme is able to successfully attack financial institutions. To deal with this more socio-technical measures are executed as well. Those measures are focused on informing the clients of the financial institutions which are known as the weakest link (Friedman, 2014). Informing clients could be about the technical measures they can adopt themselves and informing clients could focus on the malicious links and phishing mails. Regarding the first, clients are informed that they should regularly update all computers and software that are used to conduct online banking transactions (Kalige and Burkey, 2012). Kalige and Burkey (2012) mention the following primary elements; Operating System, Antivirus software, Java, Adobe Flash, Adobe Reader, Internet Browser, and Any other tools or programs used for downloading files or web surfing. Regarding the other
focus of informing clients, social engineering has been explained and it is discussed how clients should not react on phishing mails.

In addition to those measures, the financial sector focusses on a culture of cyber resilience. According to, Symantec (2016), “cyber Resilience is about the management—not the elimination—of risk. Not only is eliminating risk impossible, but it impedes agility; an environment with an acceptable level of risk supports innovation”. Furthermore, IT Governance (2016), mentions that “Traditional cyber security is proving an increasingly inadequate response to the modern cyber threat landscape. It’s no longer sufficient to suppose that you can defend against any potential attack; you must accept that an attack will inevitably succeed. An organization’s resilience to these attacks – identifying and responding to security breaches – will become a critical survival trait in the future”.

The CPMI (2014), researched the importance of cyber resilience to better understand financial market infrastructures’ abilities and perspectives in the field of cyber resilience. From this research an extensive cyber resilience framework has developed that includes an integrated approach. This integrated approach focusses, besides the CIA triangle, also on cyber governance and a range of measures (CPMI, 2014).

Cyber governance deals with an integrative approach of people, technology, process and communication. Besides, the range of measures focus not only on prevention, instead also on detection and recovery (CPMI, 2014).

2.6. Summarizing the essential parts for the continuation of this research

Chapter 2 provided an extensive description of the background of cybercrime within the financial sector. Thereby, the focus was particularly on financial malware. For the continuation of this research, some parts of the background are required to understand. These parts are described briefly in this paragraph.

First, the client-server model can be used to understand the execution of online financial services, on a high-level (paragraph 2.1). Besides, this model supports in understanding a man-in-the-browser (MitB) attack, which can be enabled by Zeus financial malware (paragraph 2.3). Target selection for the purpose of MitB attacks, is researched in this thesis. Particularly, MitB attacks that are enabled by Zeus financial malware between 2009 and 2013q1. Therefore, a more technical understanding of Zeus malware is required for the continuation of this research.

Thereby, the creation of Zeus botnets is necessary to understand. Including the command and control (C&C) server, that sends the web injects to infected computers (paragraph 2.3.1). Computers are infected via phishing mails or drive-by downloads, the infected computers are called “bots”. Subsequently, a C&C server is able to provide instructions to these bots. Furthermore, the C&C server can send web injects to the bots. These web injects are able to modify web pages of the bots. This kind of financial service fraud is researched further in the thesis. The web injects are coded into configuration files (paragraph 2.3.1).

Furthermore, it is required to understand that financial malware schemes are complex networks, existing of actors and organizations that enable financial malware attacks. These networks also contain non-criminal actors. Actually, target selection is determined by the criminals within the malware scheme. However, for the simplicity of this thesis, target selection by financial malware schemes is researched (paragraph 2.4.1).
Finally, the development and costs of cyber security measures are necessary to understand. The costs coming along with cyber security measures, are a significant part of the total costs of cybercrime. Therefore, trade-offs are required when making cyber security decisions (paragraph 2.5.2). Currently, there has become a general understanding that cyber security should both focus on technical measures and on socio-technical measures. Currently, cyber governance in the financial sector, deals with an integrative approach of people, technology, process and communication. Besides, the range of measures focus not only on prevention, instead also on detection and recovery (paragraph 2.5.3). The continuation of this research, focusses on the influence of both technical and socio-technical characteristics of financial institutions, on target selection.
3. **Methodology**

The previous chapters mention the research gap, the research goal and the research questions. Moreover, the research approach is defined and the first phase of the research is executed. Which provides the background of financial malware and cyber risk management. This chapter clarifies how the other phases of the research are executed, to be able to answer the sub questions. By answering those questions, it is intended to compare the evolution of target selection assumed by experts from the field, with target selection that is actually made according to the Zeus dataset. The outcomes of this confrontation should contribute to financial malware risk management within the financial sector.

Below, the 11 steps are mentioned which are executed to answer the remaining three sub questions. Those steps are divided among three remaining phases of the research.

**Phase 2: Extracting assumptions regarding (the evolution of) target selection from the experts**

1. Consultation of experts.
2. Placing the experts’ expectations about target selection, within the Routine Activity Theory.
3. Developing hypotheses based on these expectations.
4. Selecting hypotheses to test with the Zeus dataset, based on exclusion-criteria analysis.
5. Creating appropriate hypotheses to test with the Zeus dataset.

**Phase 3: Analyzing (the evolution of) target selection with the Zeus dataset**

6. Mentioning the data to add to the Zeus dataset and which statistical analyses to execute, in order to test the hypotheses.
7. Executing high-level analyses, to gather a high-level overview of the distribution of Zeus malware attacks, among financial institutions within the SEPA.
8. Describing a metric to test the hypotheses.
9. Selecting SEPA countries based on different criteria for testing the hypotheses.
10. Adding data to the Zeus dataset to be able to test the hypotheses.
11. Executing statistical analyses to test the hypotheses.

**Phase 4: Comparing the evolution of target selection as assumed by the experts with the patterns shown by the Zeus dataset**

12. Comparing the outcomes of the consultation of experts with the outcomes of the tested hypotheses.
13. Placing the (in-) consistencies in a broader context.

In this chapter more insight is provided into the methods used to execute those phases.

**3.1. A rationale of cyber criminals, the Routine Activity Theory**

To be able to define target selection by the financial malware schemes, a theory is required that defines why certain financial institutions are more likely to be targeted than others. The expectations and assumptions of the experts are placed in the perspective of the theory. In here, some theories are mentioned that take criminal behavior into account. Subsequently, it is explained why the Routine Activity Theory (RAT) is most suitable for this research.
To understand target selection by the financial malware schemes, more explanation on the emergence of cybercrime is required. Several social theories exist that address why some individuals or organizations execute criminal activities; social learning theory (Burgess and Akers, 1966), social bonding theory (Hirschi, 1969), general strain theory (Agnew, 1992) and the rational choice theory. Those theories are helpful for elaborating on the financial malware scheme, by explaining why those people turn into criminals and execute criminal behavior. However, those theories don’t support in explaining why the financial malware schemes select certain targets. Besides those theories, various crime prevention theories exist that are based on the actor choice. For instance, the deterrence theory, which states that fear for formal punishments is key to restrain crime. Hereby, two factors are at play; certainty and severity. ‘Certainty’ indicates the chance of being caught and punished, in addition ‘severity’ regards the severity of the punishment. Although, this theory can be very interesting for research in the cyber domain, it isn’t appropriate for researching why certain financial institutions are more likely to be chosen as targets, comparable to other financial institutions.

For this research, the Routine Activity Theory (RAT) (Cohen and Felson, 1979) is used. This is a theory from the criminology field that clarifies “who are more likely to be victimized” (Cohen and Felson, 1979). This theory is developed for the physical world (terrestrial world). However, according to Yar (2005) and Choo (2010), this theory can be applied to the cyber domain as well. Therefore, RAT could provide an explanation of the emergence of cybercrime, and supports in explaining why certain financial institutions are more likely to be chosen as target than others. Although this theory doesn’t concentrate on the nature and the offender and its motives to commit crime, it is very useful for this research, because the offender (the financial malware scheme) is just assumed to exist. In short, RAT doesn’t concentrate on why the cybercriminals execute those crimes, but the theory supports in understanding why the cybercriminals choose certain targets.

RAT proposes that crime occurs during every-day routines when a suitable target is in the presence of a motivated offender and is without a capable guardian on a certain time and place (Cohen and Felson, 1979). In here, the presence of a suitable target and a motivated offender and the absence of a capable guardian, are mentioned as the three conditions that should exist together before crime actually takes place. Within figure 5 the framework of Cohen and Felson (1979) is provided.

![Figure 5: RAT framework adopted from Cohen and Felson (1979)](image)

In 2005, Yar concluded that “although some of the RAT core concepts can indeed be applied to cybercrime, there remain important differences between ‘virtual’ and ‘terrestrial’ worlds that limit the theory’s usefulness”. One of the main differences exist in the fact that in the terrestrial world RAT
holds that a certain time and space are central for crime, however in cyberspace the combination of
time and space isn’t crucial anymore (Yar, 2005). Yar (2005), mentions that the spatial and temporal
barriers are collapsed in cyberspace. In addition, the cyberspace has many-to-many connections,
which makes it possible to select thousands targets simultaneously instead of selecting only one
target at the time in the physical world. And finally, the online identity can be made more
anonymous, compared to the non-virtual world (Yar, 2005).

Some years later, Choo (2011) described that cybercrime prevention strategies make use of RAT as
well. Choo (2011) mentions that “Cybercrime prevention strategies using RAT, for example, target
each of these areas; (1) increasing the effort required to offend; (2) increasing the risk of getting
caught; and (3) reducing the rewards of offending”.

Below, the three conditions of RAT are elaborated, based on cybercrime within the financial sector:

**Presence of a Motivated Offender**
Due to the advent of the internet and the widely usages of it within the financial sector to provide
clients access to their accounts, the number of motivated offenders to execute financial malware
attacks have increased. Within cyberspace motivated offenders have the advantages of anonymity
and no geographical limitations. In addition, cyber criminals may avoid detection and/or criminal
charges because law enforcement agencies being hampered by jurisdictional, legislative and
evidentiary issues (Hutchings and Hayes, 2009).

Moreover, according to Duffield and Grabosky (2001), offenders may justify or rationalize their
actions. In the context of financial malware, rationalization by an offender may incorporate the fact
that the victim is usually refunded the fraudulently obtained amount by the financial institution
(Hutchings and Hayes, 2009). Since financial institutions are legally obliged by payment services
directive (PSD) 2007/64/EC to reimburse the victims, provided that the victim acted carefully. Which
limits the victim liability to 150 euros. However, currently a revised version of the payment service
directive (PSD2) has to be adapted by the member states of the European Union. In here, financial
institutions are allowed to reimburse all the costs. But the Dutch Central bank prefers to keep the
150 euros own risk as an incentive for the clients of financial institutions to behave prudently (Rotte,
2015). According to Hutchings and Hayes (2009) other rationalizing factors of offenders are
minimizing the crime (as they are only obtaining a little from a lot of people), and assuming that
offences will not be pursued by law enforcement (Hutchings and Hayes, 2009).

Furthermore, ‘Malware kits’ can be acquired online and require a minimum level of expertise to
utilize (Wyke, 2011). Therefore, many people are able to execute cyber-attacks. For all those reasons,
the ‘motivated offender’, is assumed to exist rather than analyzed by RAT in this research.

The other two conditions of RAT are key for the continuation of this research. The “presence of a
suitable target” and the “absence of a capable guardian” play a significant role in describing the
(evolution of) target selection, made by the Zeus malware schemes. Thereby, those two conditions
can be related to cyber risk management via the *probable frequency of future loss*, as defined by
Jones (2005). On high level this can explained as follows; when a financial institution is more suitable
to attack than other financial institutions, the probable frequency of future loss will be higher as well.
The same holds for the other condition, when a financial institution has more absence of a capable
guardian than other financial institutions, the probable frequency of future loss will be higher as well.
Hereby RAT is used as an extension of cyber risk management, this is explained below.
Presence of a Suitable Target

In this thesis the clients of the financial institutions that have been attacked are called ‘victims’ and the financial institutions are classified as ‘targets’. Regarding financial malware, the victims are used by the criminals to reach the targets money. According to (Cohen and Felson, 1979), a suitable target can be estimated according to its four-fold constituent properties; value, inertia, visibility and accessibility. The value of a financial institution in the area of cybercrime can be defined as the value that can be gained by the offender if the attack is successful. Regarding financial malware attacks, value can be decomposed into two factors; wealthy targets and many possibilities to attack targets. This can explain that financial institutions in richer countries (higher GDP) and with higher account balances, would be selected more often. And that financial institutions in countries with higher online banking penetration rates and with many clients would be selected more often as well.

Inertia is about barriers for the criminals to move their gains to themselves or other criminals. Regarding financial malware those gains are particularly money and financial credentials. Transferring them might depend on the ease of recruiting money mules, the financial institution’s clearing time and the financial institution’s cross-national money transfer policy (Tajalizadehkhoob et al., 2013).

Visibility is about how visible the target is to the cybercriminals. Regarding financial malware attacks it is about the reputation of the financial institutions and whether the online banking pages are easy to copy with web injects. Therefore, it might be that financial institutions that are active in multiple countries around the world, sponsoring big events around the world or having many clients will be more visible. Tajalizadehkhoob et al. (2013) mention that financial institutions with the word bank in their web domain name are more visible as well.

Finally, Accessibility is about how easy the target can be reached by cybercriminals. Regarding the financial malware attacks, authentication methods play an important role here. Financial institutions with two-factor authentication are less accessible for criminals than institutions with one factor authentication. In addition, Tajalizadehkhoob et al. (2013) describe that the language of the financial institutions’ webpages also influences the accessibility of a target. According to them, financial institutions in Europe with English as language option on their webpages are more accessible. So in short, the financial institutions with a higher value, lower inertia, higher visibility and more accessibility will be selected more often as target and thereby encounter a higher probability of future loss due to financial malware attacks.

Besides those four-fold of constituent properties, the number of suitable targets increases with the behaviors of potential victims, for instance, more financial institutions providing internet banking which is used by their clients for financial transactions. Nowadays with the intensifying mobile banking, financial malware is created that focuses on mobile platforms as well. Mobile variants of the Zeus malware are The-ZeuS-in-the-Mobile or Zitmo (Etaher and Weir, 2015). So when more financial institutions are using mobile banking, the number of suitable targets increases for this variant of financial malware. However, this research doesn’t take the mobile versions of financial malware into account.

Absence of a Capable Guardian

The third condition of RAT is the absence of a capable guardian. The term ‘capable guardian’ is used widely; it may include the owner of the property, law enforcement, Computer Emergency Response
Teams (CERTs), banks and financial institutions, or any other individual or agency that has the potential to discourage offenders (Yar 2005).

A capable guardian is closely related to the property accessibility of a suitable target, as described above. However, regarding this thesis, the main differences is that the accessibility regards measures of financial institutions which can be easily figured out by the cybercriminals, like the manner of authentication and what type of firewall the financial institutions have in place. While on the other side, the capable guardian includes the (technical) counter measures of a financial institutions that are most of the time unknown by the cybercriminals, because they are confidential. Examples of measures regarding the guardian of financial malware attacks are; SOCs including the SIEMs, Transaction Monitoring, external security partners, awareness programs for their clients and employees, etc.

In the early days of information security, all research and measures to defend against cybercrime were technically focused (Claessens, 2002; Florêncio and Herley, 2011; van den Berg et al., 2014). Technical measures developed are; anti-virus software, access control, transaction monitoring, firewalls, etc. Regarding Zeus malware attacks, it is important that the financial institutions’ clients have up-to-date anti-virus software and preferably a firewall as well. Although it seems naturally that all the technical measures belong to the condition capable guardian, some of these are more related to the property accessibility of the condition ‘presence of a suitable target’.

Nowadays, there has been an understanding that cybercrime can’t be defended by only taking technical measures (van den berg et al., 2014). For example, phishing used to infect computers with malware is often based on attacks that rely on aspects of human nature rather than technical exploits. A capable guardian can be created by arming the clients and employees of financial institutions with information and awareness, rather than security software (Hutchings and Hayes, 2009). Increasing the public’s awareness of potential victimization enhances their capacity to be capable guardians (Hutchings and Hayes, 2009). Grabosky and Smith (2001) mention that a key principle in preventing cybercrime is to raise awareness on the target side about the risks which they face.

### 3.2. The consultation of experts

Experts from the field are consulted, to gather expectations regarding target selection. The outcomes of these consultations are scoping the research further to be able to execute useful quantitative analysis. The experts are asked to provide their expectations about, why certain financial institutions are more likely to be selected by financial malware schemes, than others. Or, whether they expect that cyber criminals randomly target as many financial institutions as possible.

Three of the consulted experts are involved during the whole research project. Those consultations are unstructured. However, they have a great contribution to the result of this research. In addition, four other experts are consulted to gather broader insights, and more knowledge and experiences. Those experts are consulted once or twice by semi-structured interviews. Hereby, the conditions of RAT are used to place the outcomes of the consultations in a theoretical perspective.

The descriptions of the experts’ ideas are used to develop hypotheses. Many hypotheses are created. Together with an exclusion-criteria analysis, based on the criteria ‘suitability of time’ and ‘availability of resources’, two hypotheses are selected for the quantitative analysis. Furthermore, the selected hypotheses are processed, in order to create appropriate hypotheses to test with the Zeus dataset.
3.3. The Zeus dataset and data preparation
The third phase of this research, consists of quantitative research with the Zeus Dataset. This dataset has been created and used in previous research by Tajalizadehkhoob et al. (2013). The raw data contains instructions sent to the computers that are infected with Zeus malware. However, Tajalizadehkhoob et al. (2013) extracted the targeted domains by Zeus malware between 2009 and 2013q1. Furthermore, they extracted the time of the attack, and the unique botnet key that identifies the attacker.

For the purpose of this research, new data needs to be added and some of the existing data has to be adapted. The data consists of globally attacked domains, however, for this thesis only the targeted domains from the SEPA countries are selected (see appendix I). The Zeus dataset is stored in a MySQL database, which made it very easy to extract data, adapt data and add data.

In paragraph 5.1 the Zeus dataset is described including some of the adaptations. The extensions of the dataset, to be able to test the hypotheses, is described in paragraph 5.4.

3.4. Quantitative analysis
For the purpose of the quantitative analysis, different subsets of the Zeus dataset are extracted with MySQL. With those subsets, analyses are executed within the R programming language. First high-level analysis are executed to gather insights into the distribution of Zeus malware attack among domains from different SEPA countries. Hereby insight is provided into the number of attacked domains per country and the number of botnets that are attacking different financial institutions per week. Moreover, the high-level analysis zoom in on the Dutch domains. These high-level analyses are executed in paragraph 5.3.

Before the hypotheses are tested, a metric is described that can indicate the relative attack intensity among different attacked domains (see paragraph 5.4). Subsequently, the hypotheses are tested based on this metric. Those tests exist of (multiple) linear regression analyses and non-parametric tests. Thereby, the relation between the independent variables (provided by the factors) and the dependent variable (provided by the metric) is tested. For those analyses the R programming language is used as well.

3.5. Comparing the outcomes of the two methods
Finally, the ideas of experts regarding target selection are compared with the outcomes of the quantitative analyses. Thereby, three significant issues are taken into account. First, the scope of the qualitative analyses and the scope of quantitative analyses differ. Hereby, the qualitative analysis focus on target selection regarding financial malware in general. While the data used for the quantitative analyses is from Zeus malware attacks between 2009-2013q1.

Secondly, the hypotheses that are tested with the Zeus dataset are subject to multiple transformations, in order to create appropriate hypotheses to test with the Zeus dataset. Therefore, the outcomes of the tested hypotheses cannot directly be reflected on the ideas of experts. Finally, the Zeus dataset has its own shortcomings. Which means that the outcomes of the quantitative analyses also have their limitations. Due to those limitations, the expectations of experts that are not consistent with the data shouldn’t be directly discharged. Vice versa, hypotheses that are consistent with the Zeus dataset, do not directly indicate that the experts are perfectly right regarding target selection.
4. Target selection as assumed by experts, in the perspective of RAT

Target selection by the financial malware schemes, influences the cyber risks encountered by financial institutions. Getting a better understanding of why cybercriminals select certain financial institutions more often as target than others, will enable both the financial institutions and the supervisory authorities to create more effective cyber risk management. A general definition of risk is provided by Jones (2005), as described in the introduction to FAIR: “The probable frequency and probable magnitude of future loss, or in other words how frequently something bad is likely to happen, and how much loss is likely to result”.

Target selection directly affects “the frequency something bad is likely to happen” and thereby indirectly affects “the amount of loss that is likely to result”. Because financial institutions, which are selected as target (more often), encounter a higher frequency of attacks. Moreover, the amount of loss causes by those attacks is (probably) higher, than when the financial institutions isn’t selected as target, or is targeted less.

In this chapter, RAT is used to place the expectations of experts, regarding target selection, within a theory. Hereby, the financial malware scheme are assumed to be the likely offender. Based on that assumption, the other two conditions of RAT (a suitable target and absence of a capable guardian) are used to describe (the evolution of) target selection, as assumed by the experts. With the consultation of experts, it is researched whether they assume that target selection is more focused on the suitability of a target or on the absence of a capable guardian.

The application of RAT on Zeus malware attacks, is described in paragraph 4.1. Subsequently, in paragraph 4.2 the consultation of experts is elaborated. Those consultations lead to assumptions regarding target selection and its evolution. Hereby, RAT supports with understanding and structuring the assumptions of experts. Furthermore, in paragraph 4.3 extracted hypotheses are mentioned. In addition, in paragraph 4.4 exclusion-criteria are used, to determine which hypotheses are selected, for further analyses in this research. Unfortunately, for this research there isn’t enough time and there aren’t enough resources available to test all the hypotheses. Finally, in paragraph 4.5, the selected hypotheses are processed, in order to create appropriate hypotheses to test with the Zeus dataset. Thereby, paragraph 4.5.3 provides an overview of the hypotheses that are tested within the quantitative part of this research.

4.1. RAT and (Zeus) financial malware

In this paragraph, the Routine Activity Theory (RAT) is applied to financial malware attacks. In the next paragraph RAT is used to place the outcomes of the experts’ consultations in a theoretical perspective. RAT is a theory from the criminology field that can provide an explanation regarding who is more likely to be targeted by cybercrime (Yar, 2005). Besides, Choo (2011) mentions that there are a number of ways that criminological theories such as RAT can be applied to reduce the risk of cybercrime. These research target selection choices include: which financial institution to attack, at which point of time and for how many weeks.

RAT is applicable to cybercrime. In here is described how the conditions of RAT can be used to describe target selection by financial malware. The ‘Presence Motivated Offender’ is presented by the criminal actors within the financial malware scheme. Furthermore, the motivated offender is enhanced by cyberspace’s anonymity characteristic. Regarding financial malware, a ‘Suitable Target’
is the financial institution’s website or other type of webpages that can be included in the malware configuration files to be targeted with financial malware (Tajalizadehkhoob et al., 2013). The ‘Absence of Capable Guardian’ exists because it is technically impossible to continuously defend against all the financial malware attacks. It other words, cybercriminals will always find possibilities to circumvent security counter measures by financial institutions and the third parties who provide those measures. Moreover, it has become an understanding that socio-technical factors have influences on the absence of a capable guardian as well (van den Berg et al., 2014). In figure 6 an overview is provided of RAT applied to Zeus financial malware for the purpose of this thesis.

![Figure 6: RAT applied to Zeus malware adapted from Cohen and Felson (1979)](image)

### 4.2. The ideas of experts regarding target selection

This paragraph is key within the research, inhere target selection and its evolution, as assumed by experts from the field are elaborated. The geographical scope of target selection is determined by the SEPA countries. Insight into the evolution of target selection is gathered by researching literature and with the consultation of experts. Because the scientific literature of target selection by malware schemes is scarce, the consultation of experts is used to extend the knowledge field. Hereby the development of Zeus malware as introduced in chapter 2.3, is related to the evolution of target selection as assumed by experts from the field.

First, in paragraph 4.2.1, the approach of the consultations is described. This includes, the consulted experts, the method and structure of the interviews, the transcription of the output, and the description of the experts’ assumptions and expectations. Then in paragraph 4.2.2 the ideas of experts regarding target selection are described and placed in the theoretical perspective of RAT. Those ideas are further elaborated in subparagraphs 4.2.2.1 and 4.2.2.2. Furthermore, in paragraph 4.2.3 the evolution of target selection is discussed. Finally, paragraph 4.2.4 contains a conclusion of whole paragraph 4.2.

#### 4.2.1. The consultation of experts by semi-structured interviews

Semi-structured interviews are executed, to obtain insights into (the evolution of) target selection, as expected by experts from the field. This field exists of the financial sector and the cyber security area. By consulting experts, experiences from the field are gathered. In addition, those experts are able to enhance the knowledge of financial malware risks. The academic literature on this subject is not so saturated. Thereby the information gathered by the consultation of experts is used to enhance the knowledge about target selection by financial malware, gathered from academic literature.
Furthermore, the outcomes of the semi-structured interviews provide a scope for the quantitative research method. In the next chapter, quantitative analyses are executed to research whether the expectations of experts can be validated with the Zeus dataset. However, in this paragraph the consulted experts, the method and the structure of the interviews are clarified.

The consulted experts

Several interviews are conducted, the consulted experts can be classified within two levels of contribution. Three experts where heavily involved in the research. One of them (Rolf van Wegberg) is the daily supervisor of this project, and the two others (Maarten Bras and Raymond Kleijmeer) are external supervisors from the Dutch central bank. Besides those three experts, four other experts are interviewed once or twice. The experts are selected based on their job experiences and their relation with the financial sector. The choice has explicitly been made to not include security experts from security companies, because this research focusses on the perception of experts within the financial sector. Regarding the qualitative part of the research seven experts with diverging jobs and expertise are consulted, those are:

1) **Maarten Bras (Member of the Cyber Intelligence Unit at DNB):** Mr. Bras is part of the cyber intelligence unit of the Dutch Central Bank that facilitates this research. In addition Mr. Bras is part of my graduation committee. Thereby, Mr. Bras was involved in the qualitative part of the research on a weekly basis.

2) **Raymond Kleijmeer (Member of the Cyber Intelligence Unit at DNB):** Mr. Kleijmeer is also part of the cyber intelligence unit of the Dutch Central Bank that facilitates this research. In the beginning of the research Mr. Kleijmeer was heavily involved with discussing interesting gaps. Later in the research process he was involved with the qualitative part of the research on a monthly basis.

3) **Arne de Boer (Supervisor at IT oversight at DNB):** Mr. de Boer is part of the department IT oversight. Therefore he has profound knowledge about the cyber risk frameworks used at Dutch financial institutions. In addition has some profound knowledge about cyber security measures executed by Dutch financial institutions. Unfortunately, most of those taken measures are confidential. Mr. de Boer is consulted two times and afterwards provided some extra literature about IT standards used by the Dutch Central Bank.

4) **Rolf van Wegberg (PhD candidate Economics of Cybersecurity research group at the Technical University of Delft):** Mr. van Wegberg is a PhD researcher at the cyber security research group of the TU Delft. In addition, he is researcher at TNO – TU Delft. His research is embedded in the MalPay project, focusing on malware targeting financial service worldwide. Therefore, he has much knowledge about the strategies of cybercriminals for target selection and the interaction between these strategies and the policies of financial service providers. Mr. van Wegberg is involved during this whole research. From writing the proposal till finalizing the quantitative research. Regarding the qualitative part of the research Mr. van Wegberg provided much ideas and thoughts. Mr. van Wegberg was consulted on a monthly basis.

5) **Maarten Jak (Intelligence Specialist II at ABN AMRO | Expertise Team Analysis | Security & Intelligence Management):** Mr. Jak is Cyber intelligence specialist at ABN Amro. Moreover, he is involved in the MalPay project as well. Because Mr. Jak is working for a commercial bank, he can provide insight into the malware threats they encounter. Besides he has a network of security experts and intelligence specialists from other financial institutions where he can compare the encountered threats with. Therefore, Mr. Jak is able to describe target selection and the evolution of it in comparison with other financial institutions.
6) **Hessel Mooiman (head information risk management and CISO at Binckbank):** Mr. Mooiman is head information risk of the Dutch investment bank Binck. The Binck banks is not found in the Zeus dataset. Therefore it is very interesting to discuss the threat landscape of Binck bank. Furthermore, Binckbank is not focused on payments services, instead it focusses on security transactions as a broker. Therefore, the threat landscape for those financial institutions differs from the threat landscape of the standard banks.

7) **Robin Döttling (PhD candidate finance Group University of Amsterdam):** Mr. Döttling provided workshops for the course financial regulation. The course provides insight in the regulatory frameworks currently in place within the financial sector. Hereby the way of approaching and dealing with issues within the financial sector became clear. Besides, the course gives more insight in the phenomenon of misaligned incentives. Misaligned incentives exist, because of externalities, information asymmetry and adverse selection. Externalities have a big impact in the economics of cyber security, therefore it is very useful to get more grip on that issue. The course elaborates on the influences information asymmetry currently have within the financial sector. Those insights provide a fundament for better understanding the consequences of information asymmetry regarding cyber security within the financial sector. Finally, the consequences of adverse selection are discussed as well in the course, hereby the market for lemon theory by Akerlof was discussed as well. In literature, the market for lemon theory has also been applied to security goods.

**Method and structure of the interviews**

In here, it is described how the interviews are organized to gather the assumptions of experts regarding target selection (and its evolution). Since the Dutch central bank (DNB) is the client of this research, first two experts (Mr. Bras and Mr. Kleijmeer) from DNB’s cyber intelligence unit are consulted. During the project those two experts were continuously involved and consulted multiple times. In addition to them, Mr. van Wegberg, a PhD researcher on the economics of cyber security at the TU Delft was also continuously involved and consulted multiple times as well. For those three experts no interview protocol exists, because it were extensive discussions that took place during the whole project.

The other experts from the field are consulted ones or twice by a semi-structured interview of approximately one hour. During those interviews, the experts provided their opinion and thoughts about financial malware and the evolution of target selection by financial malware schemes. Before the interviews, experts are informed about the topic of the research and the research questions. The interview protocol with the questions is provided in appendix II.

Before the interviews, an explanation of target selection regarding this research is provided. Furthermore, the findings by Tajalizadehkhooob et al. (2013) regarding target selection are mentioned to provide the experts some baggage (Those findings are also provided in appendix II). However, the experts aren’t informed with the output from previous experts’ interviews. This is done to prevent the interviews from turning into the same direction as the previous ones. During the interviews the target selection by financial malware schemes is discussed, as assumed by the experts.

**Output of the interviews and transcription**

Together with the experts, which were continuously involved in the research, the two main ideas regarding target selection are designed. Subsequently, regarding the four experts that were consulted with semi-structured interviews, a more focused approach was used. Thereby, was specifically searched, towards characteristics of financial institutions, which could make them more likely to be selected as target.
Turning the output of semi-structured interviews into text, requires effort and thoughts. Since this kind of interviews doesn’t make use of an elaborated set of questions, the output can be very diverging. Although this makes the transcription of the interviews labor intensive, it also creates the opportunity to gather new ideas. Ideas that aren’t steered or restricted by the interviewer. After the interviews are transcribed, the outcomes are placed within the perspective of the Routine Activity Theory (RAT) as well. The interviews are transcribed, in order to provide an overview of the corresponding expectations and assumptions of experts. Furthermore, it shows the creation of meaningful hypotheses. In addition, with the transcribed interviews, the expectations of experts can be placed in the theoretical perspective, more easily.

Description of the assumed target selection
In paragraph 4.2.2 the assumptions of experts regarding target selection are elaborated. On a high-level, those assumptions regard target selection by the financial malware schemes, about which financial institution to attack, at which point of time and for how many weeks. However, the intention of the interviews is, to gather more information about target selection on a lower level. Instead of only extract experts’ assumptions, regarding which and when, the intention is to gather expectations regarding, why these financial institutions are targeted. Hereby, ideas of experts are gathered about characteristics of financial institutions, based on which cybercriminals could make target selection.

For the purpose of successful Zeus malware attacks, the Zeus malware schemes need an overlap of machines (computers of client’s) which are infected with Zeus malware and web injects of the financial institutions of the infected clients. Therefore, for the purposes of this research, target selection is actually focused on selecting financial institutions to develop web injects for. To remember: web injects are malware configuration directives that are used to inject rogue content into the Web pages of bank websites to steal confidential information from the institution’s customers (Klein, 2011).

Therefore, the general questions for the experts are; which financial institutions are selected for the development of web injects? And why are those financial institutions selected? Moreover, it is asked if and how these choices have changed over time.

4.2.2. The experts’ ideas of target selection in the perspective of RAT
Since the Zeus malware family appeared on the (underground) market, many people became able to buy and use malware for attacks (Lucas, 2015). Although, some technical skills remain required to execute an attack, developing expertise became unnecessary. Therefore, it is expected that many Zeus botnets have been active. Notice that the different malware schemes that run those botnets, can select targets in different ways. Jak (2016) mentions, that certain botnets existed, which only attacked one single financial institution.

From the expert interviews, many ideas about target selection by financial malware schemes are gathered. Those ideas represent the expectations of experts regarding target selection. On high-level, two main ideas can be distinguished; 1) Many financial institutions are targeted, without a clear (pre-) selection process. Subsequently, the criminals continue targeting the institutions, which seems vulnerable. 2) Financial malware schemes use pre-selection when targeting financial institutions. Based on certain characteristics, institutions are selected and targeted. For those institutions the web injects are optimized. The characteristics for pre-selection, can be very diverging and on different levels. Many of the characteristic supposed by the experts, are mentioned in
paragraph 4.2.2.2. However, only three of these characteristics, which are clearly mentioned by multiple consulted experts, are elaborated. These three characteristics are;

(a) **The size of a financial institution**: size can be expressed by many factors, like; net profit, total assets, total payments, number of clients, etc.

(b) **The expected vulnerability of a financial institution**: the expected vulnerability can be determined by assessing the authentication methods in place, checking the firewalls, determining the awareness of clients, etc.

(c) **The context of the country where the financial institution is established**: the context of the country can be taken into account during target selection (e.g. the level of cooperation between financial institutions and law enforcement, money transfer policies, availability of money mules).

Regarding those characteristics, a relative high level has been taken into account. This high-level is taken into account because less knowledge exists about target selection by financial malware schemes. It is therefore more useful to take a broader perspective as starting point. After describing the qualitative analysis on a relative high-level, for the purpose of the quantitative analysis the focus could be on a lower level. Furthermore, based on the outcomes of this research, it could be determined whether future research on a lower level of certain characteristics could be interesting.

The main differences between the two ideas of target selection, is the selection of targets in advance. However, at both the ideas of target selection trial and error is used to test whether financial institutions can be successfully attacked. Tajalizadehkhoob et al. (2013) mention that trial and error is likely to exists, because globally many financial institutions are attacked for a short period of time.

However, according to Bras (2015), cyber criminals attack the same financial institutions for a short period of time, because otherwise financial institutions are improving their defense measures. Bras (2015) mentions that: “*When cyber criminals attack a certain financial institution for a significant long period of time, that financial institution will execute measures to counter this attack and will invest more in threat mitigation than when an attack that isn’t that persistent. The trade-offs are different, because the cost-benefit differs when the losses are less (due to a less persistent attack)*”. Those trade-offs mentioned by Bras are explained in paragraph 2.5.2. According to Bras (2015), financial malware schemes are focusing on keeping the defense measures of financial institutions low. This belongs more to the idea of target selection that uses pre-selection.

With the trial and error approach, cyber criminals are searching for vulnerable financial institutions. However, the vulnerability of a target can be broadly interpreted. Regarding financial malware, the question whether a target is vulnerable or not, can’t be answered with ‘yes’ or ‘no’. Instead, vulnerability can be expressed in many ways, like; the number of clients that are successfully attacked and the amount of money that is successfully transferred from clients to the financial malware schemes, or a combination of these (e.g. the amount of money gained per successfully attacked client, or a ratio of these like the number of successfully attacked clients per the total number of attacked clients of the financial institution).

The experts’ ideas are elaborated in subparagraphs 4.2.2.1. and 4.2.2.2. Both subparagraphs are finalized with placing the idea of target selection within the perspective of the routine activity theory.
Hereby, it is mentioned whether the specific idea of target selection is particularly focused on the suitability of a target or more on the absence of a capable guardian.

During the interviews with Bras (2015), de Boer (2015), Kleijmeer (2015) Jak (2016) and Mooiman (2016), it is noticed that target selection evolves overtime. Thereby, the influence of the characteristics is subject to a certain evolution as well. The experts assume that target selection have changed over time, together with the development of the malware itself. This evolution of target selection is described briefly in subparagraph 4.2.3.

4.2.2.1. Target selection without pre-selection

Target selection that randomly selects targets, is the first main idea of experts that is described. This idea indicates that many financial institutions are targeted without a clear pre-selection process. Subsequently, if successfully the attack continues. For this research, it means that many financial institutions from the SEPA countries are selected at least once. All those selected financial institutions should be in the configuration files, and be shown by the Zeus dataset. This target selection approach is mentioned by Bras (2015), van Wegberg (2015) and Jak (2016) as a possible way of selection target. However, note that these experts also mention that target selection based on characteristics of a financial institutions is very likely.

Jak (2016) and Mooiman (2016) mention, that there are financial malware schemes which target many financial institutions. Those malware schemes use malware that contains huge configurations with many financial institutions in it. According to van Wegberg (2015) and Jak (2016), in the early years of Zeus malware this was the case, because financial malware schemes didn’t know which financial institutions are vulnerable. After the vulnerable targets were identified, only these financial institutions are attacked further.

Target selection without pre-selection is explained with the routine activity theory (RAT) below.

**Explaining the idea of target selection without pre-selection in the perspective of RAT:**

Based on the routine activity theory, the idea of targeting financial institutions without pre-selection, can be described as follows: first many financial institutions are selected regardless of the suitability of the target and the capable guardian of the target. Both the suitability of a target and the absence of a capable guardian are assessed after the financial institutions are attacked for the first time. Based on these assessments it is determined whether to continue with attacking the financial institution or not.

Regarding the suitability of a target, this statement has to be nuanced al little. Because, before the financial institutions can be selected as a target, the financial malware schemes should at least know of the existence of a financial institution, what the web injects should look like and how many authentication codes are required. Which means that, before the financial institution is selected, each property of a suitable target (value, inertia, visibility and accessibility) has to meet a certain minimum level.

However, it is assumed that these levels are met easily. Subsequently, after the financial institution has been targeted, the suitability of a target is determined again. Hereby, the value can be measured by the number of successfully attacked clients or the amount of money that is successfully transferred. Inertia is for most of the financial institutions in SEPA assumed to be approximately the same. Because of the money transfer policies within SEPA, only the availability of money mules could
differ per country. Visibility won’t be measured again after the first attack, because those financial institutions are selected already. Finally, accessibility is assessed by the ease of creating web injects for the domain of the financial institution, and the ease of getting authorized.

Besides, after the financial institutions are attacked for the first time, also the absence of a capable guardian is assessed. However, it is expected that this is not a straight forward task and could be done in multiple ways. Possibly it will be measured just like the value of the financial institution. For example, by the number of successful attacks, which again can be expressed by the number of successfully attacked clients or the amount of money transferred to the malware scheme. When the financial institution is valuable according to the first attack, the capability of the guardian doesn’t make sense for the criminals anymore.

4.2.2.2. Target selection that uses pre-selection based on characteristics.

In contrast with the idea of randomly selecting targets, Bras (2015), Kleijmeer (2015), van Wegberg (2015) and Jak (2016) assume that the financial malware schemes (carefully) selecting targets in advance, based on characteristics of financial institutions. According to the literature and those consulted experts, many different characteristics of financial institutions could be used to select targets. Characteristics that come across, diverge from socio-economic characteristics towards technical characteristics of the financial institutions. For example: authentication methods in place (e.g. one/two factor, or innovative ways of authentication, risk based authentication), the number of clients, the net income, accepted risk levels, private or public property, the rank in google, availability of web injects, advanced cyber security measures in place, the country where the financial institution operates, whether or not the financial institutions are vulnerable according to reviews on the (dark-) web.

This paragraph focusses on the idea of target selection that uses pre-selection, based on characteristics of a financial institution. Many characteristics that could be taken into account are mentioned by the experts (see above). Three of those characteristics which are mentioned explicitly by the experts are elaborated. First, target selection based on characteristics is explained with the routine activity theory (RAT). Subsequently, each characteristics is described in relation with target selection, moreover for each characteristic a connection with RAT is provided. Hereby, the intention is to show whether target selection based on characteristics focus more on a suitable target or more on the absence of a capable guardian. In other words, it is described whether the experts assume that the value, inertia, visibility and accessibility of a financial institution are particularly used for selecting the target, or that financial institutions are particularly attacked if they have weak cyber security.

Explaining the idea of target selection based on characteristics in the perspective of RAT:

Regarding pre-selecting targets, the suitability of a target and the absence of a capable guardian are decisive in deciding to select the financial institution as target or not. RAT provides an explanation of target selection for each characteristic as well. Hereby is explained whether the characteristic especially focuses on the condition of a capable guardian, or on the condition of a suitable target. Furthermore, it is explained on which properties of the suitable target the characteristic is focused.

a) Target selection based on the size of a financial institutions

The size of a financial institution seems to be a characteristic of financial institutions that plays a role in target selection. Jak (2016) mention that big financial institutions are more likely to be attacked
than smaller financial institutions. The size of a financial institution can be determined by multiple factors, like; asset value, number of clients, the net income, total payments, etc. In previous research by Tajalizadehkhoob et al. (2013), the size of financial institutions is determined by Alexa’s traffic volume and for U.S. financial institutions also the amount of deposits is taken into account.

Bras (2015) and van Wegberg (2015) mention, that the total payments operated by financial institutions, is a very interesting characteristic for cyber criminals to take into account. However, it is hard to gather this information. During the consultation of experts, especially the number of clients has been mentioned as a factor that influences target selection, for the purpose of financial malware.

The number of clients
According to Bras (2015), van Wegberg (2015), Jak (2016) and Mooiman (2016), the number of clients of a financial institution is taken into account by the cyber criminals, when selecting financial institutions as target for financial malware. Thereby, it is mentioned that target selection is based on the potential “success/cost” rate per attack. With the goal to attack as many clients with less effort and money.

Besides, for the purpose of successful Zeus malware attacks, the Zeus malware schemes need an overlap of machines (computers of client’s) which are infected with Zeus malware and web injects of the financial institutions of the infected clients. Therefore, from an economic perspective, it makes sense that financial malware schemes focus on developing web injects for either financial institutions with many clients or financial institutions with wealthy clients.

According to the Jak (2016) and Mooiman (2016), attacking financial institutions with many clients, was especially the case in the first years of financial malware. Jak (2016), Mooiman (2016) and van Wegberg (2015) mention that over the years, financial institutions with a small number of clients have become more interesting as target. Three reasons are identified that could explain this evolution towards institutions with less clients. First, because big financial institutions improved their counter measures. For instance, Jak (2016) mentions that, in the Netherlands first the three big Dutch institutions were attacked. Subsequently, the defense and detection measures of those financial institutions become highly developed, and smaller institutions become more interesting to attack. Moreover, Mooiman (2016) expected that Binckbank will encounter higher attack intensities due to the development of those counter measures at the big Dutch banks. However, he assumes that other variants of malware instead of Zeus will target Binckbank. The pattern that shows this evolution, can be called the waterbed effect.

The existence of a waterbed effect within the Dutch financial sector makes sense for a couple of reasons. First, the Netherlands is a rich country, according to Tajalizadehkhoob et al. (2013) banks within a rich country are attacked more. In addition, the Netherlands has a high broadband penetration, Tajalizadehkhoob et al. (2014) mention that countries with a high broadband penetration are attacked more. For those reasons, financial institutions within the Netherlands are always interesting to attack. When certain financial institutions in the Netherlands become less vulnerable, other financial institutions in the Netherlands could be the new target.

Secondly, financial malware schemes started searching for victims that weren’t under attack yet. Financial institutions that aren’t under attack yet, probably have worse defense measures and less informed clients. Those institutions that aren’t attacked yet are expected to be smaller financial institutions (with less clients). Finally, developed sophisticated malware makes it economically seen
interesting to attack financial institutions with a lower number of clients. The development of this sophisticated malware including the evolution of target selection is elaborated further in paragraph 4.2.2.3.

**Explaining the idea of target selection based on the number of clients in the perspective of RAT:**

From the interviews can be deduced that the number of clients of financial institutions influences the potential financial gain of attacking that financial institution. It is mentioned that developing web injects for financial institutions with many clients has a higher probability of gaining money. Because cyber criminals need an overlap of web injects of a financial institution and clients with infected machines of the same financial institution. When more clients have this overlap, the possibility to gain money is bigger. Therefore, financial institutions with more clients have a higher value regarding financial malware attacks. According to RAT, “value” is a property of a suitable target. From that it can be inferred that financial institutions with more clients are more suitable targets.

Besides, it is expected that the number of clients also influences the property “visibility” of the condition suitable target. First because big financial institutions are just known by the criminals. Furthermore, big financial institutions most of the time have a higher reputation. Which can be positive due to sponsoring of events around the world. Or negative, because for example frauds are publicized extremely when it concerns big financial institutions. However, this increasing visibility isn’t mentioned by experts as an influencing factor.

In addition, Jak (2016) and Mooiman (2016) assume that over the years smaller financial institutions become more interesting to select as target. One of the reasons mentioned is that the big financial institutions adopted more sophisticated cyber security measures. Which reduces the number of successful attacks. This means that the big financial institutions improved their guardian and decrease the property of accessibility.

From this can be concluded that according to the experts, cyber criminals first take the suitability of a financial institution into account. Subsequently, when the defense measures to protect this value become very sophisticated, the improved guardian is taken into account as well. Thereby, less suitable financial institutions become more interesting to select as target.

**b) Target selection that is based on the expected vulnerability of financial institutions**

According to van Wegberg (2015) and de Boer (2015), the vulnerability of a target can be of influence on target selection. The vulnerability of a financial institution can be determined by cyber criminals, based on different factors. Besides, the vulnerability of a financial institution can differ per financial malware family as well. One factor that influences the vulnerability of a financial institution and is widely mentioned by the experts, is the authentication method in place for online financial services (Bras, 2015; de Boer, 2015; Kleijmeer, 2015; van Wegberg, 2015; Jak, 2016; Mooiman, 2016). Hereby, the difference between two-factor authentication and one-factor authentication is described to be significant. It is known that one-factor authentication is more vulnerable, but it is unknown whether cyber criminals select financial institutions that use one-factor authentication for online banking more often as target. Moreover, the same counts for financial institutions that use two channel authentication for online financial services, which is communication through two different paths (e.g. pc and mobile).

Mooiman (2016) mentions that innovative ways of authentication, like fingerprint and eye print authentication and risk based authentication are interesting possibilities to deal with the
vulnerability of the current authentication methods in place. Besides, nowadays many authentication methods are implemented next to each other (Mooiman, 2016). According to van Wegberg (2015), other factors that could determine the vulnerability, are; whether or not the financial institution has a random reader in place, required to make financial transactions and whether or not the financial institution has monitoring systems in place that can identify malware on their systems. However, it is always the question if cyber criminals are able to identify these factors.

Furthermore, it is assumed by Bras (2015), de Boer (2015), Kleijmeer (2015), van Wegberg (2015) and Jak (2016), that big financial institutions are attacked as long as they are vulnerable. Moreover, Mooiman (2016) mentions that financial institutions that become less vulnerable, are even attacked more intensive to return on investment. Subsequently, all the attacks will be interrupted. Bras (2015) mentions, “when those financial institutions become less vulnerable, it is expected that some of the financial malware schemes start to attack other financial institutions”. According to Bras (2015), approximately the same amount of institutions will be under attack. Due to certain developments and events there can be a shift among the targets, but the same number of attacks will be there. For example: In the extreme, when bank B implements a new authentication method that makes it harder for cyber criminals to attack that bank, the attacks on bank B will shift towards another bank. Of course this will be more nuanced; due to improved defense measures some criminals will stop attacking the specific financial institution. However, not all cyber criminals quit attacking, but some criminals will shift the attack to other financial institutions.

When relating this to the authentication methods in place, it could be the case that when only a couple of financial institutions implement two-factor authentication, they may be selected less as target. However, when many financial institutions have two-factor authentication in place, it won’t influence target selection anymore. Because then the most suitable targets probably have two-factor authentication as well. When those targets are selected again, the norm will be to circumvent two-factor authentication. In addition, two-factor authentication is vulnerable for both man-in-the-middle attacks and Trojan horses.

More sophisticated counter measures that can be executed by financial institutions, are mentioned by the experts as well, to indicate the vulnerability (Jak, 2016; Mooiman, 2016). However, it seems that sophisticated measures like; SOCs including the SIEMs, Transaction Monitoring, external security partners, and awareness programs in place for their clients and employees, etc., are hard to find out by the cyber criminals. And are therefore hard to take into account as a factor to determine vulnerability.

An interesting way of target selection based on vulnerability is mentioned by Döttling (2015). He suggests that financial institutions that have high leverages, may also take more cyber security risks, and thereby be more vulnerable. Financial institutions with a higher leverage are taking more investment risks. Leverag is a technique to multiply gains and losses. While leverage magnifies profits when the returns from the asset more than offset the costs of borrowing, losses are magnified when the opposite is true. Although it is a little bit farfetched, financial malware schemes could use a leverage rate per financial institution as indicator for vulnerability.

According to van Wegberg (2015), the financial malware schemes can take the value and the vulnerability of a financial institution into account together, during target selection. First, a selection
of valuable financial institutions is attacked. When those financial institutions are vulnerable the financial malware scheme won’t change to other targets.

**Explaining the idea of target selection based on the expected vulnerability in the perspective of RAT:**

Target selection based on the vulnerability of a target. When placing this characteristic in the perspective of RAT, it is expected that the vulnerability of a target depends on the absence of a capable guardian. However, when analyzing target selection further with RAT, it is noticed that the measures which improve the capability of the guardian of financial institutions, are hard to determine in advance. Although the vulnerability of a financial institution in real, depends on the capable guardian, cybercriminals probably can’t use it to determine vulnerability in advance.

Measures executed by financial institutions that can be assigned to the capable guardian include; SOCs including the SIEMs, Transaction Monitoring, external security partners, awareness programs for their clients and employees, etc. For example, a capable guardian can be created by arming the clients and employees of financial institutions with information and awareness, rather than security software (Hutchings and Hayes, 2009). Increasing the public’s awareness of potential victimization enhances their capacity to be capable guardians (Hutchings and Hayes, 2009). Those measures are hard to find out by the cyber criminals.

On the other hand, the property accessibility of the condition suitable target, which is about how easy the target can be reached by cybercriminals, can be determined more easily by the cybercriminals. Regarding the financial malware attacks, authentication methods play an important role in here. Financial institutions with two-factor authentication or multiple-channel authentication are less accessible for criminals than institutions with one-factor/channel authentication. So, the way of authentication which is mentioned by experts in relation with the vulnerability of a target, is more focused on the suitability of a target than on the absence of a capable guardian. The same holds for the firewalls in place by the institutions and even by the clients. Firewalls can be easily figured out by the cybercriminals. Therefore, regarding RAT those firewalls belong to the property accessibility as well.

From this can be concluded that, based on RAT, the vulnerability of a target cannot always be assigned to the capable guarding. The security measures which can be found out so easily by the cybercriminals are more related to the property accessibility of the condition suitable target. Regarding target selection based on the characteristic vulnerability, it is more likely that cybercriminals assess the vulnerability of a financial institutions based on these known measures. Therefore, regarding this characteristic, first the suitability of a target is taken into account. Subsequently, after attacking the financial institution the capable guardian could be assessed as well.

**c) Target selection that is based on the context of a country**

Both the literature and the consultants Bras (2015), Kleijmeer (2015), van Wegberg (2015) and Jak (2016) mention that the context of a country plays a role regarding target selection. For financial malware attacks this context can be determined by the degree of cooperation between financial institutions and law enforcement, the degree of cooperation between financial institutions, money transfer policies of the country, the number of financial institutions in the country, and the availability of money mules within the country (Tajalizadehkhoob et al. 2013; Raghavan et al., 2014;
Mooiman, 2016). Mooiman (2016) mentions that the mule network becomes one of the most important parts of the financial malware ecosystem. He mentions that malware becomes obtainable for everybody, due to the developed criminal services that are available on the underground economy. However, transferring the money out of the system becomes the hardest part (Mooiman, 2016).

According to Tajalizadehkhoob et al. (2013), financial institutions in some countries are more interesting to attack than others. They mention that globally, the broadband penetration and the GDP of the home country of the financial institution influence the amount of malware attacks encountered by the financial institution.

Furthermore, the total fraud due to financial malware is increasing in Europe, however in the Netherlands it is decreasing. According to Jak (2016), Dutch financial institutions are selected sporadic nowadays. “Because Dutch banks have invested extremely in detection measures” (Jak, 2016). Hereby, the three big financial institutions in the Netherlands deal with cyber security issues together. Besides, Mooiman (2016) mentions that the collaboration between Dutch financial institutions on cyber security level is globally unique. Jak (2016) supports this line of argumentation by stating that Dutch financial institutions do not compete on cyber security level. Therefore, the level of collaboration between financial institutions on cyber security level within a certain country, determines the context of the country as well.

Furthermore, financial institutions in certain countries will always be under attack according to Kleijmeer (2015). He mentions that this is because of the innovation principle. Which states that, when attacks succeed in certain countries, it will be successful in almost all countries. When relating this to the context of a country, it could be stated that certain countries have such a difficult context for the purpose of financial malware, that when attacks success in these countries, they are possibly successful in any other SEPA country. For criminals that operate new variants of financial malware this can be an interesting attack tactic. According to Jak (2016), Scandinavian countries could be such an innovation country. He mentions that the Scandinavian countries encounter new versions of malware first, and are also the first in developing counter measures against those new variants of malware.

Explaining the idea of target selection based on the context of the country in the perspective of RAT:

The context of a country is a relative high characteristic. As described above, this context is determined by many lower level characteristics or factors. For the description with RAT, those lower level characteristics have to be taken into account.

The degree of cooperation between financial institutions and law enforcement within a country is hard to place within RAT. On one hand, it could be related to the property “inertia” of the condition suitable target. Because that cooperation could function as a watchdog and thereby decrease the ease of transferring money to cyber criminals. In addition, the collaboration between institutions and law enforcement could also be related to the capable guardian of a financial institution. Because the law enforcement could be seen as a third party that supports in protecting the financial institution from cyberattacks. This is also described by Yar (2005). Moreover, the level of collaboration between law enforcement and financial institutions is hard to assess by cybercriminals.
Both the money transfer policies of the country and the availability of money mules within the country are definitely related to the property “inertia” of the condition suitable target. Each of these lower-level characteristics influence the ease of transferring money from the target to the criminals.

The number of financial institutions in the country could be placed within both the property value and the property visibility of the condition suitable target. Less financial institutions within a country indirectly increase the value of the selective number of financial institutions in the country. Furthermore, when there are less institutions within a country, it is assumed that the selective number of financial institutions have a higher visibility.

Finally, the degree of collaboration between (big) financial institutions on cyber security level in a certain country, can be best related to the capability of the guardian. Getting insight into (the existence of) this level of cooperation is also hard for cybercriminals.

It can be concluded that regarding the characteristic “context of a country”, both the condition suitability of a target and the condition capability of a guardian are taken into account. However, the lower-level characteristics that relate to the capable guardian are hard to assess by cybercriminals (in advance). Therefore, it is expected that the suitability of a target is taken into account first. Subsequently, when criminals are caught by law enforcement, the capable guardian could be taken into account as well.

4.2.3. The evolution of target selection

Target selection based on characteristics, is described in the previous paragraph. In this paragraph is elaborated, what the evolution of target selection means in relation with these characteristics. It is known that cybercriminals learn quick and continuously developing their attack vectors. Jak (2016) and Mooiman (2016) mention, that together with the development of malware, target selection evolved as well. According to Mooiman (2016), the malware schemes including the malware itself, develop really fast. Because many specialists are involved within the malware economy. Moreover, even scrum teams exists for developing malware. The next subparagraph discusses the influence of evolution on the pre-selection of targets.

4.2.3.1. Static pre-selection and evolving pre-selection.

Due to evolution of target selection, the influence of the three different characteristics changes. In other words, over time pre-selection leads to another selection of targeted financial institutions. This evolution can be clarified with the terms “static pre-selection” and “evolving pre-selection”. Static pre-selection means, that financial institutions with certain characteristics will always be targeted. For instance, experts expect that big financial institutions remain targeted (Wegberg; 2015; Bras, 2015; Mooiman, 2016). More straightforward, it is assumed by the experts, that vulnerable financial institutions will always be targeted (Wegberg; 2015; Bras, 2015; Jak, 2016; Mooiman, 2016). Furthermore, the experts assume that financial institutions in all the SEPA countries will always be targeted, irrespectively of its less likely context (Wegberg; 2015; Bras, 2015; Kleijmeer, 2015; Mooiman, 2016).

Evolving pre-selection means, that financial institutions which didn’t met the certain characteristics in the first time, could be selected over time, even though their characteristics didn’t change. For instance, the experts expect that the proportion of targeted financial institutions with fewer clients, increases over time (de Boer, 2015; Jak, 2016; Mooiman, 2016).
4.2.3.2. Developments that cause the evolution of target selection

The evolution of target selection by financial malware schemes, can be attributed to three developments; the development of the presence of opportunities at big financial institutions, the development of hybrid malware, and shifting victims from retail customers towards business. Those are elaborated below. First, the evolution of target selection is described within the perspective of RAT.

Switching targets based on the presence of opportunities

Cyber criminals continuously execute attacks, however there are multiple targets to commit crime on. When the presence of opportunities to commit crime, is greater at another financial institution some cyber criminals will shift to that target. For example, the big financial institutions are improving their defense measures because of the encountered attack intensity. When the cyber defense of the attacked financial institution become more capable, due to the implementation of advanced security measures, less attacks will be successful. Subsequently, some of the cyber criminals will shift towards other targets. Jak (2016) and Mooiman (2016) mention that this shift between targets is especially the case between big financial institutions that implement strong security measures and small institutions that don’t have such a strong cyber defense yet.

Another development regarding the presence of opportunities, is caused when many botnets are attacking the same financial institutions. The opportunities on those financial institutions will decrease. Because, many clients have been attacked together and thereby become familiar with the risk of malware. Subsequently, other financial institutions become more interesting over time. This could be smaller financial institutions and financial institutions in other countries.

For example, when many financial institutions have been under attack in a certain country have been under attack heavily, many clients become familiar with the attack and the financial institutions will adopt better security measures. Subsequently the financial malware schemes start focusing on financial institutions from other countries (Jak, 2016; Mooiman, 2016). According to the experts, the countries with a less interesting context, will become more interesting over time.

According to the expectations of experts, regarding the number of clients, this means that after a certain period of time institutions with a lower number of clients become more interesting. It is expected by Jak (2016) and Mooiman (2016) that smaller financial institutions will be selected even more in the coming years. In addition, regarding the authentication methods, institutions with two factor authentication will become more interesting over time as well. All these expectations are based on the assumption, that the most interesting financial institutions to target will become less interesting as target when they have targeted intensively. Because due to the relatively high attack intensity, both the clients of the financial institution become informed with the risks of financial malware and the financial institutions themselves will improve their counter measures.

Due to the developments of the defense measures, both socio-technical and technical, the big financial institutions become more resilience. Therefore, other financial institutions are selected as targets. However, besides the evolution of target selection, developments take place at the malware sight. Malware becomes more and more sophisticated, which results in other interesting targets. This development is described next, including the opinion of experts about the influence of this development on target selection.
From automated malware to a hybrid version of malware

According to Jak (2016), the shift towards smaller institutions becomes interesting due to sophisticated malware and the ease of internet banking nowadays. With those sophisticated malware, the attacks become more target specific nowadays. An interesting financial malware development is that of the hybrid versions of malware. Mooiman (2016) mentions this as well. He mentioned that first the modes operandi of malware was focused on attacking many clients at the same time. However, the modes operandi of innovative version of malware makes it possible to monitor infected clients. And subsequently execute more targeted attacks (Mooiman, 2016).

According to Jak (2016), hybrid malware means that the malware isn’t totally automated anymore, but some manual intervention is required. This kind of malware is able to execute target specific attacks. With this hybrid malware, it is not anymore the general malware that tries to steal credentials and money of as many clients as possible, but instead the sophisticated malware monitors many infected clients, and if a client meets certain criteria (e.g. the minimum account balance), the client is selected as victim. For the purpose of this hybrid malware attack, small web injects with a pop-up screen are used to gather authentication codes, while the man-in-the-middle manually add transactions. According to Jak (2016), this development started 4 years ago.

Due to this hybrid form of malware small financial institutions became more likely to be selected as target. Jak (2016) refers to the relative small German banks that became very interesting to attack with this variant of financial malware. The hybrid versions of sophisticated malware select many financial institutions as well. Because a hybrid form of malware makes it able to attack clients from many different financial institutions.

In addition, with the sophisticated forms of malware the characteristics of the client become more interesting than the characteristics of the financial institution. Instead searching for financial institutions that are interesting to attacked, the question is whether the client is appropriate for an attack. According to this evolution, factors like the number of clients of a financial institution, and the way of authentication become less important.

The different variants of Zeus are compared with the development of hybrid malware and the simultaneous evolution of target selection. Zeus 1, Zeus 2, ICE IX, Citadel and P2P Zeus are the variants of Zeus that were developed and are available within the Zeus dataset. Especially, Citadel and P2P Zeus seem to be more target specific variants of Zeus malware. They both became available in the end of 2011.

From retail to business

The last evolution regarding target selection is gathered from the interviews with Kleijmeer (2015) and Jak (2016). This evolution discusses the shift of victims, from retail clients towards business. Among business more money is available, that can be stolen with financial malware. To be able to attack companies, a slightly different form of malware is required. Instead of financial malware, remote access tools are used. Those tools enable criminals to take over PCs entirely. Subsequently, accounting packages become available to the criminals including the linkages with the banking systems. Then the criminals are able to execute new transactions. Citadel is mentioned as the Zeus variant that starts focusing on business as well.
For financial institutions it is really hard to recognize those fraudulent transactions, because most of the companies execute a batch of transactions at the same time (Jak, 2016). So when only one extra transaction is added, it already becomes hard to detect. Moreover, when the criminal only changes the account number in the transactions that already exist, then the money mules receive the payment. In this situation, it becomes even harder to detect the transactions as fraudulent. The money mule network becomes extremely important, and many mules are required for this kind of financial cybercrime. However, the gains are extremely (e.g. invoices can be high amounts).

4.2.3.3. The evolution of target selection in the perspective of RAT

According to the experts, cybercriminals select targets based on the suitability of a target. However, due to developments within the capable guardian of financial institutions, new targets have to be selected. For this new pool of targets, the financial malware itself had to be developed as well. The experts mention, that the suitability of the clients becomes more important, instead of the suitability of the financial institution.

In the first years of financial malware, it seems that the suitability plays a significant role. According to the experts, big financial institutions, which – for instance - can be expressed by the number of clients are targeted more (valuable targets). Moreover, countries with a context that benefits the inertia of a target (many available money mules, and efficient transaction policies), are targeted more intensively. In addition, particularly the financial institutions that are easily accessible are targeted.

Over the years, the capable guardian of a target becomes more important. According to the de Boer (2015), Jak (2016) and Mooiman (2016), big financial institutions develop their security measures, and currently have sophisticated measurers in place that improve their guardian significantly. Furthermore, financial institutions in certain countries start collaborating with each other on cybersecurity level and with law enforcement. Due to the development of the capable guardian, smaller financial institutions, which are valued less, but have weaker guardians become more interesting. Besides, the clients of those financial institutions are less aware of financial malware attacks, because they aren’t targeted yet and those institutions put less effort in awareness programs for their clients.

More sophisticated malware makes it possible to focus on the suitability of the clients, instead of the suitability of financial institutions. Moreover, with this malware, business can be selected as target instead of individuals.

4.2.4. Conclusion

Two main ideas of target selection can be distinguished from the output of experts’ interviews. One of the ideas states, that financial malware schemes randomly target as many financial institutions as possible, and continue attacking the institutions that are valuable and vulnerable. In the perspective of RAT, this idea first focuses on selecting many financial institutions as target, and subsequently assess both the suitability and the guardian’s capability of those institutions. Based on the assessment, the financial malware schemes determine to continue with attacking the financial institution or not.

According to the other main idea of target selection, the financial malware schemes select targets in advance based on certain characteristics of the financial institutions. The experts mention many
characteristics of financial institutions that could be taken into account, like; authentication methods in place, the number of clients, the net income, accepted risk levels, private or public property, the rank in google, availability of web injects, advanced cyber security measures in place, the country where the financial institution operates, whether or not the financial institutions are vulnerable according to reviews on the (dark-)web, etcetera.

Three characteristics that are mentioned explicitly by multiple experts are elaborated and placed within the perspective of RAT. Which are; 1) the size of financial institution, based on the number of clients, 2) the vulnerability of the financial institution, and 3) the country where the financial institution is established. These relatively high level characteristics, are clarified by lower-level characteristics. Thereby, some conclusions can be drawn. According to the experts, in the first years of financial malware, cybercriminals select targets based on certain characteristics which are focused on the suitability of the target. Because the characteristics that are focusing on the capability of the guardian are hard to determine by criminals.

However, when attacks seem to be less successful, or when even botnets are taken down, due to improvements of the guardian. Cybercriminals start taking the guardian into account. Therefore, new targets are selected, which in first instance were assessed as less suitable. This evolution in target selection is discussed with experts as well. From this discussion can be concluded that target selection reacts on improvements of the financial institutions’ guardian. Hereby, more sophisticated variants of financial malware are developed as well. For example, the development of hybrid forms of financial malware created the possibility to attack clients from many different financial institutions. Furthermore, the suitability of an individual client becomes more important than the suitability of a financial institution.

The evolution of target selection can be explained, by static pre-selection and evolving pre-selection. Static pre-selection indicates, that financial institutions with certain characteristics will always be under attack. For instance, big financial institutions and institutions within the western countries. However, evolving pre-selection suggests, that other financial institutions are targeted over time, although they are less suitable before.

With the elaboration of the outcomes of the experts’ consultations, it is noticed that RAT can be used to differentiate assumptions of experts regarding target selection and its evolution. Thereby, the experts expect, that the suitability of the targets was decisive regarding target selection in the first years of financial malware. However, the experts assume, that after the big financial institutions improved their security with technical measures and by improving the awareness of their clients. The cybercriminals start taking the capability of the financial institutions guardian into account as well. Therefore, in the last years of financial malware the capability of the guardian became more decisive regarding target selection.

4.3. Hypotheses extracted from experts
From the elaboration of the experts’ main ideas regarding target selection within SEPA, underlying hypotheses are developed. Those hypotheses provide an overview of the expectations of the experts regarding target selection and its evolution. It can be stated that the consulted experts together could use those hypotheses for organizing financial malware risk management. Subsequently, those hypotheses are tested with the Zeus dataset, after adding more data to it. Thereby, some of the
experts’ expectations are confronted with target selection, that is actually made by the Zeus malware schemes between 2009 and 2013q1. Note that multiple hypotheses could hold simultaneously.

**The following hypothesis focusses on targeting without pre-selection:**

*Hypothesis 1*: Financial malware schemes randomly select many financial institutions as target.

In here, randomly means that cybercriminals don’t pre-select targets based on certain criteria. According to Wegberg (2015), when this hypothesis holds, the data should show many attacked financial institutions in the first weeks after an update of Zeus is out. Subsequently, after a certain period of time the number of attacked financial institutions should decline.

Without taken the newer versions of Zeus into account, the hypothesis is much more simplified. Then it could be tested whether the first year of the dataset contains much more attacked financial institutions, that the last years. However, hereby it is assumed that financial malware schemes with new variants of Zeus don’t try to attack financial institutions that aren’t vulnerable according to the previous versions of Zeus malware.

**The following hypotheses focus on characteristics of the financial institution:**

*Hypothesis 2a*: Financial institutions with more clients are selected more by financial malware schemes than financial institutions with less clients.

It is assumed that big financial institutions are attacked more than smaller financial institutions. The number of clients is mentioned by many experts as a measure of size regarding financial malware attacks. When this hypothesis holds, the data should show that institutions with many clients are indeed attacked more. However, some experts assumed that the number of clients of a financial institution became less important over time. Therefore, it is expected that this hypothesis holds in the first years of financial malware, but won’t hold for the last years.

*Hypothesis 2b*: Financial institutions that improve their cyber security measures are attacked less.

Some experts mention that financial institutions which improve their cyber security measures, will be attacked less. Testing this hypothesis can create many value according to the experts from the field.

*Hypothesis 2c*: Financial institutions with a higher leverage are selected more as target.

During the consultation of experts it is assumed that financial institution that are willing to take more investment risks are also willing to take more cyber risks. A ratio that can provide insight into the investment risks taken by financial institutions is the leverage ratio. Therefore, it might be interesting to test whether financial institutions with a higher leverage are selected more often as target.

**The following hypotheses focus on the context of a country:**

These hypotheses are based on the outcomes from the consultation of experts. These outcomes are elaborated in paragraph 4.2.2.2 under heading c) *target selection that is based on the context of a country.*
Hypothesis 3a: Financial institutions established in countries where financial institutions cooperate on cyber security level are selected less as target.

Hypothesis 3b: Financial institutions established in countries where financial institutions and law enforcement cooperate on cyber security level are selected less as target.

Hypothesis 3c: Financial institutions established in countries with an organized network of money mules are selected more as target.

Hypothesis 3d: Countries with less financial institutions encounter a higher rate of “attacked financial institutions/total number of institutions”.

Hypothesis 3e: Financial institutions established in countries with clear money transfer policies are selected more as target.

The following hypotheses focus on the evolution of target selection and the development of financial malware:

Hypothesis 4a: Over the years, small financial institutions become more likely to be selected as target of Zeus malware schemes.

As mentioned in the description of the previous hypothesis, it is expected by some experts that the number of clients of financial institutions become less important over the years. In addition, sophisticated versions of malware are able to focus on the characteristics of individual clients instead of financial institutions. Therefore, small financial institutions become more likely to be selected as target over the years.

Hypothesis 4b: Over the years financial institutions with two-factor/channel authentication are selected as much as financial institutions with one-factor authentication.

When many financial institutions have two-factor and/or two channel authentication in place, it become impossible to differentiate on it from other financial institutions. Besides, two-factor authentication is vulnerable for both man-in-the-middle attacks and Trojan horses. Therefore, it is expected that over the years financial institutions with two-factor authentication will be selected as much by financial malware as one-factor/channel authentication.

4.4. Selecting hypotheses

Four categories of hypotheses are described in the previous paragraph. Within this paragraph, it is determined which of the hypotheses to select for further research. Due to time constraints and the availability of resources not all hypotheses can be researched in this thesis. Exclusion-criteria analysis is used, to decide which strategies are further researched within this thesis. Criteria has been developed, which are described below:

1) Feasibility to research the hypothesis within three months

Since this is research is a master thesis of 30 ECTS, the quantitative part of the research should be realized within approximately three months. Although data and resources might be available, it could be impossible to gather this data and execute analysis on it within the available time.

2) Availability of the data and other resources required to research the hypothesis
The Zeus dataset, contains many targeted financial institutions within SEPA, from many different countries. For most of the hypotheses, additional data is required before they can be tested. This means that for each financial institution, or for each country data has to be gathered about a certain factor or characteristic. Sometimes, the data isn’t available at all and in some cases, the data is not provided to researchers. In addition, researching financial institutions is very sensitive. The actual figures of financial institutions, sometimes differ from the figures that are publicly available. Furthermore, data regarding security measures of financial institutions is most of the time confidential. Even within the Dutch financial sector, it seems impossible to research this characteristic further as a master’s student.

Some experts mention that financial institutions which improve their cyber security measures, will be targeted less (hypothesis 2b). Testing this hypothesis, can create many value according to the experts from the field. Thereby, it could be observed, whether the implemented counter measures decrease the attack intensity of Zeus malware attack. In other words, it could be tested whether the cyber criminals react on counter measures. Those outcomes could support in making trade-offs regarding cyber security investments. Unfortunately, it was impossible to gather cyber security measures taken by financial institutions. The consulted experts from the Dutch central bank and the ABN AMRO wouldn’t provide the specific cyber security measures that have been taken.

For all the hypotheses, exclusion-criteria are used to determine, whether the hypothesis is suitable for further analysis within this research. Table 2, shows the outcomes of the exclusion-criteria analysis. Only hypotheses 2a and 4a are selected for profound analyses.

Table 2: Exclusion-criteria analysis

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Hypothesis</th>
<th>Criteria</th>
<th>Feasibility of time</th>
<th>Availability of data and other resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Randomly) many targets are selected, and only the valuable and vulnerable institutions remain targeted</td>
<td>Yes</td>
<td>Hardly</td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>Financial malware schemes select financial institutions with many clients more often as target, than financial institutions with less clients</td>
<td>Yes</td>
<td>Mostly</td>
<td></td>
</tr>
<tr>
<td>2b</td>
<td>Financial institutions that improve their cyber security measures are targeted less</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>2c</td>
<td>Financial institutions with a higher leverage are targeted more</td>
<td>Yes</td>
<td>Hardly</td>
<td></td>
</tr>
<tr>
<td>3a</td>
<td>Financial institutions established in countries, where financial institutions cooperate on cyber security level, are targeted less</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>3b</td>
<td>Financial institutions established in countries, where financial institutions and law enforcement cooperate on cyber security, are targeted less.</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>3c</td>
<td>Financial institutions established in countries with an organized network of money mules, are targeted more.</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>3d</td>
<td>Countries with less financial institutions encounter a higher rate of “targeted financial institutions/total number of institutions”</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>
4.5. Developing appropriate hypotheses to test with the Zeus dataset

This paragraph provides insight into the process of creating the hypotheses, which can actually be tested with the Zeus dataset and other available data. In addition, the structure to test the hypotheses, is elaborated. An overview of the hypotheses that are tested in chapter 5, is provided in sub paragraph 4.5.3. Those are tested in chapter 5.4.

4.5.1. Preparation for testing the hypotheses

The hypotheses 2a and 4a (see table 2), are selected for further analysis are:

2a: Financial institutions with more clients are selected more as target by financial malware schemes, than financial institutions with less clients.

4a: Over the years, financial institutions with a small number of clients become more likely to be selected as target of Zeus malware schemes.

Unfortunately, according to the exclusion-criteria analysis none of the hypotheses regarding the context of the country seems plausible for this research. However, because many experts explicitly mention that the country has influence on the attack intensity encountered by financial institutions, the factor country will be taken into account on a high level. For this reason, it is tested whether, within different SEPA counties, the financial institutions encounter a different relation between the number clients and the attack intensity.

Analyzing on country level is kept very broad and high-level, because researching lower level characteristics of SEPA countries seem unrealistic for this research. Especially, regarding the time and the available data for a master thesis. Therefore, it is researched whether different SEPA countries encounter different patterns regarding the first hypothesis. Thereby, hypothesis 5 is as follows:

5. Within different SEPA counties, the financial institutions encounter a different relation between the number clients and the attack intensity.

First, this paragraph describes the data that should be added to the Zeus dataset, in order to test the hypotheses. Both the number of clients of the financial institution and the country, where the financial institution is established, are factors used for testing the hypotheses. According to the experts, those two factors are expected to be of influence on target selection. In addition, the Zeus dataset and the availability of other resources and time, seem to be suitable for analyzing those factors.
The number of clients
To remind, for the purpose of successful Zeus malware attacks, the Zeus malware schemes need an overlap of machines (computers of client's) which are infected with Zeus malware and web injects of the financial institutions of the infected clients. Therefore, developing web injects for financial institutions with many clients has a bigger chance of this overlap, than developing web injects for financial institutions with less clients.

Based on the required overlap sketched above, the results from previous research and the expectations of experts, the first factor that has been chosen for the quantitative part of this research is the financial institution’s number of clients. It is expected that financial institutions with more clients are always targeted by financial malware. Furthermore, it is expected that in the last years, smaller financial institutions have become likely to be selected as target.

Country
Analyzing lower characteristics of the context of a country seems implausible for this research. However, based on the consultation of experts and the findings of Tajalizadehkhoob et al. (2013), the country has been taken into account on high level. The influence of the country, where the financial institution is established, on the attack intensity encountered by the financial institutions, is researched with the quantitative analysis. It is expected that the context of a country influences the encountered attack intensity.

4.5.2. The hypothesis to test with the Zeus dataset
Per hypothesis is described, how to test whether it holds or not.

**Hypothesis 2a:** Financial institutions with more clients are selected more as target by financial malware schemes, than financial institutions with less clients.

It is expected that financial institutions with more clients are selected more often as target by financial malware schemes, than financial institutions with less clients. This hypothesis is tested over the data from the years 2009-2013q1.

**Hypothesis 4a:** Over the years, small financial institutions become more likely to be selected as target of Zeus malware schemes.

For the purpose of this hypothesis, it is tested whether over years the influence of the number of clients on target selection decreases. And whether financial institutions with less clients even become targeted more often, than financial institutions with more clients. Therefore, hypothesis 2a is analyzed per year of the Zeus dataset.

**Hypothesis 5:** Within different SEPA counties, the financial institutions encounter a different relation between the number clients and the attack intensity.

With testing this hypothesis, it is analyzed whether attack intensity differs per country. Therefore, hypothesis 2a is analyzed over the years 2009-2013q1, per SEPA country. Below, the structure to test those hypotheses is, elaborated.
4.5.3. The structure of the hypotheses to test
The hypotheses 2a and 5, are tested over all the years of the dataset. Therefore, the hypotheses 2a and 5 are tested generically. The generic hypotheses are tested first. From here, these generic hypotheses are;

1. Financial institutions with more clients are targeted more by financial malware schemes, than financial institutions with fewer clients.
2. Within different SEPA counties, the financial institutions encounter a different relation between the number clients and the attack intensity.

Besides, hypothesis 4a is tested per year of the dataset. This hypothesis provides insight into evolving pre-selection. From here, this is hypothesis 3.

3. Over the years, financial institutions with fewer clients become more likely to be selected as target, by Zeus malware schemes.

Before those hypotheses are tested, first in paragraph 5.1 the Zeus dataset is discussed. Furthermore, in paragraph 5.2 descriptive analysis are executed to gather high-level insight into the distribution of Zeus malware attacks among SEPA countries. Subsequently in paragraph 5.3 a metric is described which can determine the attack intensity encountered by financial institutions. This metric is used for testing the hypotheses.
5. **Target selection according to the Zeus dataset**

Chapter 5 deals with the quantitative part of the research. In here, target selection is analyzed with the Zeus dataset. The focus of the quantitative research, is on the targeted financial institutions within the Singe Euro Payments area (SEPA). The outcomes of these analyses provide input for the comparison/confrontation with the expectations of experts in chapter 6. This confrontation places the outcomes of the hypotheses in a broader context.

This chapter has two goals. First, descriptive analyses are executed, to provide a high-level overview of the distribution of Zeus malware attacks among financial institutions within the SEPA. The distribution of attacks is defined by: which institutions were under attack in which weeks between 2009-2013q1, and by how many botnets. These botnets represent financial malware schemes. Although one malware scheme can operate multiple botnets, it provides an indication of the popularity of domains among SEPA countries. Besides, the distribution of target selection is related to cyber risk management of both financial institutions and their authorities. Moreover, this high-level overview creates a feeling for the Zeus dataset and supports in understanding the further analysis.

The second goal of this chapter, is to provide empirical insights, into characteristics of financial institutions that make them more likely to be targeted. Here for, the hypotheses that are described in paragraph 4.5.3 are tested. The outcomes of the hypotheses are used to confront the assumptions of experts. However, the comparison between the outcomes of the qualitative analysis and the quantitative analysis, is executed in the next chapter. An important remark is, that those analyses view the “intention of attacks”, it can’t conclude about the success rate of those attacks.

The structure of this chapter is as follows. First, in paragraph 5.1 the Zeus dataset is elaborated. Furthermore, in paragraph 5.2 the high-level analyses are executed. Subsequently, in paragraph 5.3 the metric is described, to determine the relative attack intensity between the financial institutions. This metric is used for testing all the selected hypotheses. Finally, in paragraph 5.4 the hypotheses are tested. This paragraph contains the preparation of the Zeus dataset, the structure of the profound analysis, the execution of the analyses and a conclusion.

### 5.1. The Zeus dataset

Zeus is an easy to obtain malware kit that contains the tools required to build and control a botnet (Wyke, 2011). A botnet is a flexible remote-controlled network containing computers that function together to make a platform available for fraudulent and criminal purposes (Bauer & van Eeten, 2009). Configuration files configure the parameters and initial settings for computer programs (wiki, 2015). Regarding Zeus malware, the configuration files are used as instructions for the computers within the botnet their operations. Those instructions are set by cyber criminals within the financial malware scheme and among others determine which financial institutions to attack. This information is used as proxy, to determine target selection.

Previous research by Tajalizadehkhoob et al. (2013) made use of the so called “Zeus dataset”. The raw data of their research consisted of configuration files of the Zeus malware. The configuration files, used for their research, are gathered over a period of more than four years (2009-2013Q1) by using honeypots located all over the world. The configuration files were collected using two different methods: they are gathered by running live Zeus samples or by emulating the malware to download
configuration files (Tajalizadehkhoob et al., 2013). This means that the dataset contains real life data, of attacks that were really there. Note that from this dataset can’t be extracted whether the attacks were successful or not.

Tajalizadehkhoob et al. (2013) preprocessed the data for their research into a MySQL database that contains different tables. One part of this data preprocessing is decrypting the messages send through the Zeus botnet. This is done by using the botnet’s unique encryption key (RC4 key). The messages, which are included into the configuration files, contain the commands regarding which financial institution to attack. In essence, this means that those files contain the web injects. Thereby, the researchers gather all the domains for which web injects were developed. These domains all are attacked by the Zeus malware. Not all of these domains belong to financial institutions.

So, from the raw data, many web domains are extracted that were under attack by Zeus malware between 2009 and 2013Q1. In addition, from the configuration files can be extracted in which weeks and by which and how many botnets the domains were attacked. Subsequently, the dataset is supplemented with the country where the domain belongs to, the GDP rate and broadband penetration of the existing countries in the database. Moreover, the size and the website language, for each domain in the database, were determined. In here size is determined by Alexa’s traffic volume and for U.S. financial institutions also the amount of deposits are taken into account.

5.1.1. The database
Tajalizadehkhoob et al. (2013) made use of more than four years of data (2009-2013Q1), which contain 2412 domains (not all are financial institutions). In their research, information about “the targeted domains”, “time/duration of attack” and “botnet unique identifier” has been extracted. The dataset is stored in MySQL tables.

The targeted domains
The target domains that can be extracted from the Zeus malware configuration files can be categorized into the following groups:

- Personal banking
- Corporate online banking
- Investment and online trading sites
- Credit card services
- Security software
- Extremely popular global websites (like, Facebook)

Determining the country of the domain
The country of each attacked domain was determined manually by Tajalizadehkhoob et al. (2014). Hereby, four data sources are used:

a) Where the server/infrastructure is located according to MaxMind (MaxMind, 2014);
b) Where the traffic of the domain comes from, according to Alexa (Alexa Internet, 2014);
c) Where the site owner’s headquarter is located, according to the domain’s homepage;
d) The top-level suffix of the domain (TLD).
According to Tajalizadehkhoob et al. (2013), most of the time, these sources consistently pointed to the same county. They manually checked the domains where they didn’t match.

Time/duration of attack
From the configuration text files, the infection time per domain is available in the ‘Zeus’ database. A ‘first seen’ column exists, that displays the time when the configuration file is first seen. Furthermore, Tajalizadehkhoob et al. (2013) developed an indicator that shows which weeks the different domains were under attack.

Tajalizadehkhoob et al. (2013) mention that “this metric could be calculated in two ways. Either, the included weeks are only those when configuration files have been received by a botnet, or the weeks between the times that a configuration file is received by a botnet are also counted with the assumption that the botnet has been active in these intermediate weeks as well. The second assumption is closer to reality because as C&C servers might not often update configuration files when the attack is stable. However, in the periods that no configuration file is sent we can assume that the configuration file did not require to be updated and cybercriminals are still working with the previous version. This simply implies that previous targets in the target list of configuration file are still under attack”.

Botnet unique identifier
Each Zeus botnet has his own RC4 key to decrypt the encrypted configuration files. The RC4 key is different even for the same Zeus variants. This RC4 key is used as unique identifier of the botnet.

5.1.2. Selected tables for this research
For this thesis the following tables from the Zeus database are selected:

1. **ds2_v2_domains**: this table contains all the domains that have been attacked between 2009-2013q1. This table provides insight into whether the domain is a financial institution according to Alexa. Alexa is a company that provides commercial web traffic data and analytics.

2. **ds2_v2_domain_country_magic**: this table provides insight into the home country of the domain.

3. **ds2_v2_botnetsattacked_perweek_peryear**: This table provides also insight into the specific year and weeks from 2009 that a domain was under attack and into the number of botnets attacks the domain in those specific weeks.

4. **ds2_v2_attacks_domain_yearly**: This table provides insight into the attacked domains per year, and the attack intensity they encounter. The attack intensity is determined in number of weeks under attack per year, number of botnets that attacked the domain per year, the number of configuration files send per year that included the domain, and the number of botnet weeks attacked per year. This latter metric for intensity is explained further in paragraph 5.4.

The data in those tables is fundamental for this thesis. However, this dataset has to be extended for the profound analyses that are executed to test the hypotheses. In paragraph 5.5.1 is elaborated how the dataset is extended for the profound analyses.

Furthermore, Alexa does not classify all domains, therefore some domains aren’t classified as financial institution, although they are a financial institution according to this research. Therefore,
within the table \texttt{ds2\_v2\_domains} an extra column ‘FI’ (Financial Institution) is added, which stated ‘yes’ or ‘no’.

5.2. Descriptive analysis
In this paragraph, descriptive analyses are executed to provide a high-level overview of the distribution of Zeus malware attacks among financial institutions within the SEPA. This overview provides the first insights into target selection by Zeus malware schemes. Furthermore, it creates a feeling for the Zeus dataset and supports in understanding the further analysis. Those analyses show the number of targeted financial institutions within SEPA, the number of weeks those institutions are targeted and the number of botnets simultaneously targeting those institutions per week. In addition, the analyses provide insight into the distribution of the targeted financial institutions among the SEPA countries.

Furthermore, this paragraph is finalized with descriptive analyses of the targeted Dutch financial institutions. Because the client of this report is the Dutch Central bank. Moreover, most of the experts are from the Dutch financial sector. Therefore, most of the examples provided by the experts about the evolution of target selection and the development of cyber security regard the Dutch financial sector. An important remark is that those descriptive analyses view the “intention of attacks”, it doesn’t say anything about the success rate of the attacks.

Analyzing domains instead of financial institutions
For the quantitative analysis of this research, all the targeted domains that belong to financial institutions from SEPA countries are taken into account, instead of the attacked financial institutions themselves. It has been chosen to analyze on domain level, because the domains are really attacked. Tajalizadehkhoob et al. (2013) gathered all the domains that appeared in the Zeus configuration files between 2009 and 2013q1.

The domains are linked with financial institutions, but the attack intensity is determined on domain level. Multiple domains can be linked with the same financial institution. The factors that distinguish financial institutions are linked with the domains as well. Therefore, the domains that belong to the same financial institution also have the same factors. In the concluding paragraph 5.2.2, of the high-level analysis, the outcomes are translated to the financial institutions. Besides, in that paragraph the contribution of the outcomes to the understanding of target selection is described, including a relation with cyber risk management.

First in paragraph 5.2.1, insight is provided into the number of targeted domains from SEPA countries and the frequency of botnets attacking the domains per week. Besides, the relative attractiveness of different countries is shown, by providing insight into the distribution of countries among the number of attacked domains and the number of botnets attacking the domains per week. Note that these analyses only focus on targeted domains. The outcomes don’t say anything about the domains of financial institutions that aren’t targeted.

5.2.1. The distribution of targeted domains among different SEPA countries
Between 2009 and 2013q1, 959 unique domains from 26 SEPA countries are targeted. Previous research by Tajalizadehkhoob et al. (2013) classified that 345 of these domains belonging to a financial institution. This are the domains that are classified as financial institution by Alexa. An overview of the number of targeted SEPA domains per week, is shown in figure 7. The number of targeted domains in this figure, are probably an overestimation of the real number of attacks. This is
caused by the method used to determine an attack. Hereby, all the unique domains found in all the configuration files of each week are counted. However, sometimes the configuration files just contain an update of the configuration files, which is not per definition a new attack.

Inhere the proportion of the targeted domains, per country is added to the overview. This provides some first insights into the relative attack intensity per SEPA country. However, it should be noticed that the financial sectors of different SEPA countries are varying. For instance, regarding the number of institutions per country and the number of clients per financial institution in those countries, the online banking penetration per country, etc. Therefore, those relative attack intensities have to be taken into account carefully.

![Number of attacked Domains of financial institutions in SEPA countries from 2009 per week](image)

Figure 7: The number of targeted SEPA domains per week per country

Because of the many SEPA countries (26), the overview is hard to interpret. However, some findings are provided by the histogram. First, it is noticed that, between the years 2009 and 2011Q3 each week between 50 and 140 domains were targeted. Subsequently, a decrease is shown between 2011Q3 and 2012Q3, to 10 till 100 targeted domains per week. Tajalizadehkhooob et al. (2013) assume, that this is caused by the Zeus take down efforts of law enforcement. After this period, an increase of targeted domains per week is shown to between 50 and 120 attacked domains per week.

In addition, four countries are dominant regarding the number of attacked domains belonging to financial institutions. Those countries seem to be; United Kingdom, Spain, Italy and Germany. Another finding is that the pattern of those four countries seems to be approximately the same. Which means that the number of selected domains in those countries decreases and increases accordingly per week.

Subsequently, in figure 8 the same distribution is shown, without the four dominant countries United Kingdom, Spain, Italy and Germany. This distribution shows that without those four countries, approximately 80 percent less domains are selected as target per week. Furthermore, without these countries it seems that some of the other countries contain relatively many selected domains in certain weeks.
Furthermore, to get a better understanding of how often different domains are targeted per week, insight is provided into the number of botnets attacking the 345 domains per week. The botnets represent financial malware schemes. Note that one malware scheme can operate multiple botnets. Besides, insight is provided into the distribution of SEPA countries, among the number of botnets targeting the domains per week. Hereby, it can be analyzed whether the countries with the most targeted domains, are also targeted by more botnets per week. In other words, it can be observed whether the countries with many targeted financial institutions, are also targeted by many botnets.

The number botnets targeting domains per week among SEPA countries

The Zeus dataset shows that the 345 domains, belonging to financial institutions in SEPA countries, together were 19,716 weeks under attack between 2009 and 2013q1. Hereby, a new unit is developed that provides insight into the combination of domains and the weeks that these domains were under attack: “domain weeks”. Thereby, two histograms are provided that show the number of “domain weeks” that are targeted by n botnets. In other words, insight is provided into how many weeks a certain domain is attack by 1 botnet, by 2 botnets, .. and by n botnets.

The histogram in figure 9 provides insight into the number of “domain weeks” that are targeted by n botnets, distributed among the SEPA countries. It shows that most of the time, domains are targeted by one botnet. More than 6000 “domain weeks” were targeted by one botnet, which means that the 345 domains together were 6000 weeks under attack by one botnet. This is almost one third of the total attacks. Besides, more than 2500 weeks the 345 SEPA domains were attacked by two botnets, almost 1500 weeks the 345 SEPA domains were attacked by 3 botnets.
However, in this histogram the vertical scale is much stretched, due to the frequency of one botnet targeting a domain per week. Therefore, to deal with the scale issue, two histograms are provided with a varying range of the number of botnets. Those histograms are provided by appendix III.

Those histograms show that the United Kingdom, Spain, Italy and Germany are targeted the most. This makes sense that, because the frequency is determined in “domain weeks”, because the previous analyses show that these countries contain the most targeted domains. Furthermore, it is observed that, certain domains in the United Kingdom, Spain, Bulgaria, France, Germany, Ireland, Italy and Portugal were targeted by more than 10 botnets per week. Besides, it is noticed that certain domains from the United Kingdom, Spain, Italy and Ireland are targeted by 35 botnets per week. Remarkable is the presence of Ireland, one domain from Ireland is targeted in one week by 35 botnets.

Finally, this chapter analyzes the targeted domains in the Dutch financial sector specifically.

5.2.2. Targeted domains within the Dutch financial sector
Table 3 provides an overview of the first weeks that different domains belonging to Dutch financial institutions became under attack. Hereby, it is noticed that the Postbank is chosen as target first. Subsequently the ING and then the ABN AMRO and the Rabobank are chosen as target.

Subsequently, in figure 10 an overview is provided of the cumulative number of botnets attacking the Dutch domains between 2009 and 2013q1. The only line that stops increasing after the domain is attacked is that of Postbank.nl. Together with figure 10, it is noticed that Postbank.nl has not been selected as target anymore after week 140. The Postbank merged with the ING already in 2009. From the consultation with de Boer (2015), it is known that this domain has been migrated with ing.nl in multiple waves. This migration process is called the TANGO program.

The TANGO program realized the technical integration of the systems of ING Bank and Postbank after they legally merged to ING. Within three waves, the communication to the clients had been integrated. First, wave 1 realized the integration of the communication with private clients. Secondly in wave 2, the communication with SMEs had be integrated. And finally the communication with private banking and the rest of the first two waves had been integrated (AdVanched, 2016; Brotesse, 2010). Thereby, it can be concluded that between 2009 and 2013q1, Dutch domains that are targeted once, are not selected as target anymore only when the domain stops to exist.

Table 3: Attacked Dutch domains between 2009 and 2013q1 including the number of the first week of the attack

<table>
<thead>
<tr>
<th>domain</th>
<th>week_from_2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>postbank.nl</td>
<td>5</td>
</tr>
<tr>
<td>ing.nl</td>
<td>18</td>
</tr>
<tr>
<td>rabobank.nl</td>
<td>78</td>
</tr>
<tr>
<td>abnamro.nl</td>
<td>105</td>
</tr>
<tr>
<td>asnbank.nl</td>
<td>131</td>
</tr>
<tr>
<td>snsbank.nl</td>
<td>148</td>
</tr>
<tr>
<td>iecards.nl</td>
<td>187</td>
</tr>
<tr>
<td>regiobank.nl</td>
<td>187</td>
</tr>
<tr>
<td>snsreal.nl</td>
<td>193</td>
</tr>
<tr>
<td>abnamro.com</td>
<td>193</td>
</tr>
</tbody>
</table>
According to the Zeus dataset 959 unique domains from 26 SEPA countries are selected as target by the Zeus malware schemes between 2009 and 2013q1. Based on the Alexa classification, 345 of these domains belong to financial institutions from SEPA countries. Between 2009 and 2013q1, those 345 domains together were 19.716 weeks under attack. In 2011 the number of attacked domains decreased. Subsequently, after approximately a year the number of attacked domains increased again. From these outcomes can be concluded that domains of financial institutions from SEPA countries are very interesting targets. Therefore, each financial institution within SEPA should expect that they are attacked by financial malware, when organizing their financial malware risk management. For instance, by informing their clients with the risks of financial malware. Among all the SEPA countries, most of the selected domains are from the countries; the United Kingdom, Spain, Italy and Germany. However, the financial sectors of different SEPA countries are varying. Therefore, no hard conclusions can be drawn from the fact that most of the attacked domains belong to a small selection of SEPA countries.

A unit “domain weeks” has been developed to express the combination of domains and the number of weeks they were under attack. This expression can be used to show the number of botnets attacking domains per week. On a weekly basis, most of the domains were attacked by one botnet. It is shown that 6000 domain weeks were under attack by one botnet. This is almost one third of the total attacks. Furthermore, one till ten botnets attacking a domain per week occurred very frequently, while more than 35 botnets attacking the same domain occurred very sporadic.

Domains from Bulgaria, France, Germany, Ireland, Italy, Portugal, Spain and the United Kingdom are attacked by more than ten botnets in certain weeks. Moreover, domains from Ireland, Italy, Spain and the United Kingdom are attacked by more than 35 botnets in certain weeks. Therefore, domains in these countries are more likely to be targeted than domains from other SEPA countries. The authorities of those counties should together with the concerning financial institutions assess the encountered financial malware risks. In addition, based on the assessment the financial malware risk management of those institutions could be adapted.
Besides, from the descriptive analyses of the attacked Dutch domains can be stated that first the big Dutch institutions are targeted and subsequently the smaller institutions are targeted. It isn’t analyzed whether the selection of those smaller institutions has a relation with the improved cyber security measures of the big Dutch financial institutions. However, assuming that the big Dutch financial institutions have adopted more sophisticated counter measures between 2009 and 2013, it is noticed that this doesn’t influence the number of botnets targeting the domains of these Dutch financial institution. Because, the domains of Dutch financial institutions that are selected once, are only not selected anymore when the domain stops to exist. Based on that, Dutch financial institutions should assume that they are always likely to be targeted, notwithstanding executing (sophisticated) counter measures. The financial authorities, who develop financial malware risk management criteria, should take this into account. Further research should analyze whether this can be generalized to other SEPA countries as well.

In paragraph 5.4, profound analyses are executed towards target selection by testing the hypotheses. In addition, some of the domains that are targeted more than average are taken into account. The outcomes of these analyses should provide more clarity about why certain financial institutions are more likely to be targeted than others. This clarity could support the authorities and financial institutions even more with their financial malware risk management. However, first in paragraph 5.3 a metric is presented that can provide insight into the relative attack intensity of different targeted domains.

5.3. A metric regarding attack intensity

In the high-level analysis, overviews are provided that show the number of domains attacked per week, and the frequency of the number of botnets attacking these domains per week. With further analysis, the intention is to get a better understanding of target selection and its evolution. To be able to analyze the intensity encountered by different domains, a metric is required that can provide insight into the relative attack intensity.

The metric used for further analysis is a combination of the number of weeks that domains were targeted and the number of botnets attacking the domains per week. This metric can be described as: The average number of botnets attacking a domain per week. Tajalizadehkoob et al. (2013) mention that, “to compare over longer periods, one could add up the count for each week (“botnet weeks”) or average them”. Within this research the number of botnet weeks attacking a domain per year is used as metric.

Tajalizadehkoob et al. (2013) developed this metric, to deal with the overestimation due to just counting the send configuration files that contain domains. This overestimation is caused by the frequency different configuration files are updated. According to Tajalizadehkoob et al. (2013), the update frequencies of different configuration files are subject to a high variance. Therefore, the number of configuration files send does not closely relate to the number of times a financial institution has been really selected as target. According to Tajalizadehkoob et al. (2013), “this metric eliminates some of the limitations of overestimating by normalizing the data: it merges all configurations for a single botnet sent during a week and then counts the number of botnets attacking that domain in the particular week”.

5.4. Profound analysis

In this paragraph the profound analyses are executed. The profound analysis is particularly focused on testing the hypotheses, to get a better understanding of target selection. Furthermore, the domains, which are targeted relatively intensive, are taken into account individually. The outcomes
of those analyses are used for the confrontation with the expectations of experts. Subsequently, the outcomes of this confrontation could support both the financial institutions and the authorities with organizing financial malware risk management.

The outline of this paragraph is as follows. First the preparations to the Zeus dataset are described in subparagraph 5.4.1. Subsequently, subparagraph 5.4.2 mentions the structure of the profound analysis. Then in subparagraph 5.4.3 the first part of the profound analysis is executed. Subsequently, in subparagraph 5.5.4 the second part of the profound analysis is executed. Furthermore, subparagraph 5.5.5 analyses an extra characteristic of financial institutions that could influence the encountered attack intensity. Finally, conclusions are drawn in subparagraph 5.5.6.

5.4.1. Preparing the dataset
In paragraph 4.5, it is concluded that regarding this research, the number of clients of the financial institution and the country where the financial institution is established, are the most interesting factors to research further. Therefore, first the Zeus dataset has to be extended with those factors. The number of clients has to be determined for all the targeted domains. In addition, it is noticed that regarding this research actually more domains are classified as financial institution than Alexa does. Therefore, it is manually checked whether the domain has to be classified as financial institution or non-financial institution.

An important decision for the continuation of this research is which domains to include for the profound analysis. Preferable all the targeted domains of financial institutions from the SEPA countries are used for the profound analysis. Unfortunately, this is unrealistic regarding the available time of this research and other barriers to research the number of clients belonging to financial institutions.

The selected domains
For the profound analysis only the targeted domains from certain countries are taken into account. Tradeoffs had to be made regarding which countries to select. Only for the targeted domains from the selected countries it is manually checked whether they are classified as financial institutions and what the number of clients is of the financial institution. The tradeoffs are based on the following criteria:

1) Multiple SEPA countries: The influence of the country on (the evolution of) target selection is intended to be researched. Therefore, multiple countries have to be selected for this research. Moreover, by selecting the countries, the geographical location has been taken into account. The intention was to select countries spread across whole Europe (provided that they are included in SEPA).

2) Outcomes of the high-level analysis: The outcomes of the high-level analysis, executed in paragraph 5.2, are used to determine which countries can be interesting to select. Countries that contain domains which are targeted by many botnets per week could be interesting to further analyze. Because the number of botnets targeting a domain, can provide some insights into likely targets.

3) Available time: The quantitative analyses are executed within three months by one researcher. Researching the number of clients for each targeted financial institutions is a time consuming task. Therefore, time constraints play an exclusive role within the tradeoffs. Which means, that countries with more than 25 targeted financial institutions, are unrealistic for the purpose of this research.
4) **Language barriers:** Because the number of clients have to be researched of institutions from different countries, language plays a significant role in the possibility to find a representative number.

5) **Countries mentioned by experts:** Because the expectations and assumptions of experts are researched, the countries mentioned by the experts during the consultations are selected first (when they met the other criteria as well).

The countries that are finally selected for further research are; Austria, Belgium, Bulgaria, Finland, France, Hungary, Ireland, Norway, Romania, and the Netherlands. Those countries are called “the selected SEPA countries”. Subsequently, for those countries all the domains are manually checked whether it is a financial institution or not. Furthermore, for all the domains that belong to a financial institution, the number of clients are researched.

**The number of clients**

The number of clients per financial institution are gathered from different data sources invoked through the internet. Hereby the total number of clients is researched, which includes both business clients and private clients of the institution in the specific country. For example: ING in the Netherlands has a different number of clients as ING in Belgium. Unfortunately, it seems impossible to gather data regarding the number of clients of each financial institution for each year from 2009 till 2013. Therefore, the number of clients from a single year is used, it is decided to use the year that is most closely to 2011 (which is the middle year of the dataset).

The preference of the source of the data has the following order: 1) the authority’s website is preferred most, 2) subsequently the institution’s website, 3) then a third source has been searched and 4) finally Wikipedia has been used. However, for most of the institutions only one data source had been found. For the domains, where no number of clients has been found at all, it is determined whether this financial institution is expected to be small, middle or large. Based on this classification, the domain gets assigned a number of clients. When classified as small, it gets the smallest number of clients found in the country, when classified as medium it gets the average number of clients found within the country and when classified as large it gets the highest number of clients found in the country among the targeted financial institutions. Appendix IV, shows the number of clients per country per targeted domain.

**The dataset used for further analysis**

The dataset used for further analysis is a subset of the Zeus dataset, which consists of the target domains from the following countries; Austria, Belgium, Bulgaria, Finland, France, Hungary, Ireland, Norway, Romania, and the Netherlands. The dataset contains data about the weeks between 2009 and 2013q1 that those domains are attacked, the number of botnets attack in domains, the unique botnet numbers attacking the domains, and the attack intensity per year those domains encountered in botnetweeks per year. Subsequently, for each of these domains, the number of clients of the financial institution where the domain belongs to, is added to the dataset.

Summarizing, this dataset contains 106 unique domains. Note that some of those domains belong to the same financial institutions, and therefore get assigned with the same number of clients. The dataset contains 2141 ‘domain weeks’, this means that the 106 domains are together attacked in 2141 weeks between 2009 and 2013q1.
5.4.2. The structure of the profound analysis

The structure of the profound analysis exists of two parts, and intends to test the three hypotheses that are mentioned in paragraph 4.5.3. Furthermore, it analyzes domains, which are targeted relatively intensive. Note, that these analyses only observe the targeted domains. Therefore, the outcomes don’t say anything about the domains of financial institutions that aren’t targeted according to the Zeus dataset. Moreover, the outcomes can’t conclude about the success rate of the attacks.

Within the first part of the profound analysis, the hypotheses 1 and 2 are tested generically. In here, generic means, over all the years in the Zeus dataset. Hereby, it is tested whether among all the targeted domains from the selected countries, the domains with more clients are targeted more than domains with less clients. In addition, the targeted domains that encounter relatively intensive attack intensities, are further taken into account. This relative attack intensity is measured among the targeted domains from the selected countries. For analyzing these domains, groups are created in two different ways, to show the influence of the domains on the patterns shown by the Zeus dataset. In addition, the country of these domains is shown. Finally, insight is provided into the relative size of the institution within its country, based on the number of clients.

Furthermore, in this first part generic analyses are executed per country. Hereby, it is per country analyzed whether the domains with more clients are targeted more, than the domains with less clients. An overview of the first part of the profound analysis is provided by figure 11.

![Figure 11: Overview of the structure of the generic analysis](image)

Subsequently, after the generic analysis, the second part of the profound analysis is executed. This part observes the relation, between the number of clients and the attack intensity, per year. This analysis is executed to test hypothesis 3, and provides insight into the evolution of target selection. Hereby, the attack intensity per year of all the targeted domains, from the selected countries is taken into account. An overview of the second part of the profound analysis is shown by figure 12.

![Figure 12: Overview of the structure of the analysis per year](image)
5.4.3. Generic analysis
Within this paragraph, the first part of the profound analysis is elaborated. Hereby, the first hypotheses are tested. In addition, the domains that are targeted relatively intensive are taken into account.

1. **Financial institutions with more clients are selected more as target by financial malware schemes, than financial institutions with less clients.**
2. **Within different SEPA counties, the financial institutions encounter a different relation between the number clients and the attack intensity.**

First, the relation between the number of clients belonging to a financial institution and the attack intensity encountered by the institution is analyzed generically. For the attack intensity, the number of “botnet weeks” attacking a domain per year is used as a metric. Among others, a linear regression is executed, to explain the relation between the number of clients and the encountered attack intensity.

The linear regression assumes that the relation is linear. Mathematically, this linear relationship can be written as follows: \( Y = \beta_0 + \beta_1X \) (James et al., 2013). Hereby, \( Y \) is the attack intensity and \( X \) is the number of clients. \( \beta_0 \) and \( \beta_1 \) are two unknown constants that represent the intercept and slope terms in the linear model (James et al., 2013). If \( \beta_1 \) is positive (which is expected) than an increase in the number of clients results in a higher attack intensity. When \( \beta_1 \) is negative, then more clients results in a lower attack intensity. Summarizing, when the number of clients of a financial institution increases with factor \( X \), than the encountered attack intensity increases with factor \( \beta_1 X \).

The linear regression line is fitted using the least squares approach. In other words, \( \beta_0 \) and \( \beta_1 \) are defined by the least squares approach. This approach limits the distance between the linear line and the data points that the line is trying to explain (James et al., 2013). The coefficient of determination (\( R^2 \)) is a measure that can be used to assess how close the data are to the fitted regression line. In other words, \( t \) provides a measure of how well observed outcomes are replicated by the model. The \( R^2 \) is a number between 0 and 1. Where, \( R^2 = 0 \) means that the regression line isn’t able at all to describe the observed outcomes. And \( R^2 = 1 \) means that the regression line is totally able to describe the overserved outcomes. In addition, the significance is calculated.

However, linear regression can only be used when the data meets the assumptions of linear regression. The first assumption is normally distributed data of all the variables. The second assumption is that the variance in \( y \) is homogeneous over all \( x \) values and independence, which means that the attack intensity at a certain financial institutions should not influence other attack intensities. This assumption is called homoscedasticity.

5.4.3.1. Analyzing the relation between the number of clients and the attack intensity

1. **Financial institutions with more clients are selected more as target by financial malware schemes, than financial institutions with less clients.**

Inhere the first hypothesis is tested. First, a linear regression is executed and subsequently the Spearman’s Rho is used for testing the following relation: **Does a significant positive relation exists between the number of clients of a financial institutions and the attack intensity encountered by the financial institutions?**
Within the selected countries 106 domains have been attacked between 2009 and 2013q1. For those 106 domains, the attack intensity, botnet weeks per year, is determined over each of the years. Based on those yearly attack intensities the regression line is modelled. Because some of the domains are targeted in multiple years, the dataset contains 220 values. Those 220 values are called “domain years”, which can be calculated by counting the unique targeted domains per year, and then taking the sum of the attacked domains in all those years. Domain years, is a new developed unit, based on which the following analyses are executed.

**Linear regression analysis**

Before executing the linear regression analysis, it is tested whether the data meets the assumptions of this statistical model. The four graphs in figure 13, provide insight into the assumptions of homoscedasticity and homogeneity of the data.

![Residuals vs Fitted](image1)

![Normal Q-Q](image2)

![Scale Location](image3)

![Residuals vs Leverage](image4)

**Figure 13: Graphs to check the assumptions of normality and homoscedasticity of the linear regression analysis**

The **residuals vs fitted graph** and the **scale location graph** provide insight into the reasonability of the linear relationship, the equality of the variance in the error terms and into outliers. With that information the assumption of homoscedasticity can be tested. Both the residuals vs fitted graph and the scale location graph should show no pattern. Otherwise, the data is not homogenously distributed. However, the two graphs in figure 13 show a pattern.

When further analyzing the residuals vs fitted graph, the existence of a linear relationship can be doubted, because the residuals are not distributed randomly around the zero line (dotted line). Furthermore, the variance of the error terms are clearly not equal. Finally, some residuals are clearly outliers. Those outliers are provided by the values; 11 (aib.ie in 2009), 34 (banquepopulaire.fr in 2009), and 161 (procreditbank.bg in 2009). So it can be concluded that the data is not homogenous distributed and thereby doesn’t met the assumption of homoscedasticity.
Besides those two graphs, the **normal Q-Q graph** provides insight into the normal distribution of the data. The assumption is met when the values are distributed on a straight line. That is definitely not the case here. So the assumption of normality is probably not met.

Finally, the **residuals vs leverage graph** provides insight into the influence of certain y values (attack intensities). Each value is removed at a time and the new model is compared to the one with the point, if the point is very influential then it will have a high leverage value. Points with too high leverage value should be removed from the dataset to remove their outlying effect on the model. Hereby an extra outlier has been identified; value 35 banquepopulaire.fr in 2010.

There are a few mathematical transformations, which can be applied to non-normal or heterogeneous data, to meet the assumptions. One of these is log transformation, which is executed for the purpose of linear regression analysis in this research. Hereby, the log is taken of the dependent variable (attack intensity) to create homogenous data. When executing the linear regression model again, it is noticed that the linear model performs much better regarding normality and heterogeneity. It is concluded that the assumptions are met. Appendix V, provides insight into the improvements of the data.

**Linear regression after log transformation of the attack intensity**

Figure 14 provides a scatter plot of all the 220 values. Each dot represents a domain year. On the horizontal axis, the number of clients is provided and on the vertical axis the attack intensity is shown. The linear regression line is drawn in the figure as well, which has a positive slope. An increase of one million clients leads to an increase in attack intensity of $e^{1*0.15} (= 1.16)$ botnet weeks per year. The relation is significant according to the p-value shown in table 5. Unfortunately, the R squared is still very low. Only 8.3% of the variance can be determined by the regression line.

Remind that, for 2013 only nine weeks of attacks have been gathered. This could have a negative influence on the linear regression model. Therefore, this linear regression analysis is also executed without the year 2013. In appendix VI, the outcomes of this regression can be found. The relation between the number of clients and the attack intensity is still significant and positive. Besides, both the regression analyses are slightly better without the year 2013 than with the year 2013. After the log transformation, the regression line is able to explain 11.2% of the variance.
Table 4: Linear regression model information after log transformation

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Regression coefficient</th>
<th>Std. Error</th>
<th>T value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.2815</td>
<td>0.1611</td>
<td>7.954</td>
<td>9.73e-14 ***</td>
</tr>
<tr>
<td>Number of clients</td>
<td>0.1534</td>
<td>0.0346</td>
<td>4.434</td>
<td>1.47e-05 ***</td>
</tr>
</tbody>
</table>

Note: ***p<0.001, **p<0.01, *p<0.05; 218 degrees of freedom; R-squared = 0.07851

To further improve this distribution, the outliers could be deleted as well. However, note that within this research outliers are very interesting to research, because these are the domains that are selected as target relatively intensive. Since the data exists of attacks that were actually there, those domains are really targeted that intensive. Besides, transforming the data with a log transformation, non-parametric tests could be executed with data that isn’t normal distributed or homogenous.

Among the non-parametric tests, the Spearman’s Rho is less sensitive for outliers, therefore it is used to analyze the relation between the number of clients and the attack intensity. Moreover, the Spearman’s Rho can be used to analyze non-linear relations.

**Spearman’s Rho test**

Finally, the Spearman’s rho test is executed, to analyze the level of significance, between the number of clients of the domains and the attack intensity encountered by the domains. Hereby, the strength of the monotonic relation is measured. Regarding the Spearman’s rho test a monotonic relation is assumed. Which means that when the number of clients increases, the attack intensity will never decrease.

According to the Spearman’s rho test the relation between the number of clients and the attack intensity is significant (p-value < 0.05). Thereby, the relation is 0.33, which is a weak (monotonic) relation. The outcomes of the Spearman’s rho test are in table 6.

**Spearman’s rho test**

Finally, the Spearman’s rho test is executed, to analyze the level of significance, between the number of clients of the domains and the attack intensity encountered by the domains. Hereby, the strength of the monotonic relation is measured. Regarding the Spearman’s rho test a monotonic relation is assumed. Which means that when the number of clients increases, the attack intensity will never decrease.

According to the Spearman’s rho test the relation between the number of clients and the attack intensity is significant (p-value < 0.05). Thereby, the relation is 0.33, which is a weak (monotonic) relation. The outcomes of the Spearman’s rho test are in table 6.

Table 5: The Spearman’s rho test on the targeted domains from all the selected countries

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-value</td>
<td>6.123e-07</td>
</tr>
<tr>
<td>Rho</td>
<td>0.3287248</td>
</tr>
</tbody>
</table>

Both the Spearman’s Rho test and the regression analysis show that a significant positive relation exists between the number of clients and the encountered attack intensity. Therefore the first hypothesis can be accepted. However, the regression line is not able to explain the relation well. Furthermore, the found monotonic relation with the Spearman’s rho is weak. Therefore, further analyses are required to be able to draw stronger conclusion about the relation between the number of clients and the encountered attack intensity. For these further analysis, groups of domains are created to analyze the (influence of) the relatively intensive targeted domains.

### 5.4.3.2. Analyzing the targeted domains based on groups of domain years

For further analysis, five groups are created for which the average number of clients and the attack intensity are calculated. Groups are created in two different ways, to get insight into the influence of the relatively intensively targeted domains, on the linearity and the monotonicity of the relation, between the number of clients and the attack intensity. Groups are created manually in the following two ways:
1) Groups are created by sorting the domain years based on the assigned number of clients and subsequently cutting those groups into five groups with the same size (44 domain years).

2) Regarding the second way, five groups/clusters are created based on the number of clients. The smallest number of clients among the selected domains is 0.05 million, and the biggest number of clients is 11 million. Therefore five categories are created based on the size of 2.2 million clients (see table 7). Each of the domain years has to be divided into these five categories.

Table 6: Five groups with the same reach of 2.2 million clients

<table>
<thead>
<tr>
<th>Groups including the reach of the number of clients</th>
<th>Minimum # of clients (in millions)</th>
<th>Maximum # of clients (in millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>2,3</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>4,3</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>6,7</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>8,9</td>
</tr>
</tbody>
</table>

Insights gathered from manually created groups

Groups have been created manually in two ways. For both ways, the means of the number of clients and the means of attack intensity of the different groups are compared. Appendix VII, provides insight into the means including an elaboration. Note, that inhere the domains that encounter relatively high attack intensities are analyzed. No hypotheses are tested.

On average, the targeted domains from the selected countries, are attacked by 47.5 botnet weeks per year. Remind that for the year 2013, only the first nine months has been taken into account. Therefore, the mean of attack intensity will probably be higher when whole 2013 could be taken into account, or when 2013 is extrapolated. However, the choice has been made to take only the available data into account, and not to extrapolate. Because the quantitative analysis concentrate on the attacks that are really executed by the Zeus malware schemes. Moreover, the strength of this quantitative analysis is that it makes use of real data. After extrapolating, the data won’t be totally real. Besides, within the rest of 2013 other domains could be targeted as well. Those domains won’t be created by extrapolation.

When analyzing the average number of clients and the attack intensity of those groups and, a linear relation seems to exist. Where the financial institutions with less clients encounter a lower attack intensity, and financial institutions with more clients encounter a higher attack intensity. Besides, a monotonic relation seems to exist as well (see the trend line in figure 15).

Subsequently, regarding the second way of creating groups, again the average number of clients and the attack intensity of each group are analyzed. Although the trend line provided by this analysis looks almost the same with the previous analysis, the attack intensities per group totally differ from attack intensities per group shown by the previous analysis. Furthermore, there doesn’t exist a monotonic relation between the groups and the attack intensity (see figure 15).
By further analyzing the groups 2 and 5 (with the relatively high attack intensities), it is shown that three domains (aib.ie, procreditbank.bg and banquepopulare.fr) are selected by relative many botnets, in all the years presented in the dataset (2009-2013q1). These domain years, clearly influence the attack intensity, encountered by the groups. According to the regression analysis, executed earlier in this paragraph, for the year 2009 these domains seem to be the outliers. Therefore, those domains are analyzed further.

Analyzing the domains that encounter relatively high attack intensities
First, the domain aib.ie is analyzed, which belongs to the Allied Irish Banks, containing 4 million clients and thereby has the most clients of the attacked institutions in Ireland. This domain has a relatively high attack intensity for the year 2009, namely of 920 botnet weeks. Also within the year 2010, this domain has been targeted relatively intensive, namely by 769 botnet weeks. In addition, in the years 2011 and 2012 the domain has still been under attack by respectively 427 and 487 botnet weeks. Finally, in the year 2013, which van only be analyzed for the first quarter, the attack intensity is still 123 botnet weeks (which is almost 500 botnet weeks per year, when multiplied by four).

Another outlier, is provided by the domain procreditbank.bg. This is a Bulgarian bank belonging to the Procredit group which has banks in 19 countries. This financial institution has the most clients among the selected Bulgarian financial institutions for Zeus malware attacks. Within 2009 the domain is attacked for 877 botnet weeks. And in 2010 till 2013 for respectively 671, 392, 473 and 123 botnet weeks. The latter will be almost 500 botnet weeks as well, if the quarter is multiplied by four.

Finally, banquepopulare.fr is taken into account. This domain is banquepopulare.fr, which belongs to a group of French banks containing 9.4 million clients. Thereby it is one of the two attacked French institutions that contain the most clients. The domain has been targeted relatively intensive, in the years 2009 and 2010, by respectively 822 and 705 botnet weeks. In the years 2011 till 2013 the domain was still under attack by respectively 423, 490 and 124 botnet weeks. The latter one will be around the 500 botnet weeks if multiplied by 4 quarters.
It can be concluded, that only the number of clients of a financial institution within the Single Euro Payments Area, is not able to describe and clarify target selection very well. Besides, it is noticed that the three domains that encountered relatively intensive attack intensities, all having the most clients among the attacked financial institutions in their own country. The following subparagraph, analyzes the relation between the number of clients and the attack intensity on country level.

5.4.3.3. Analyzing the relation between the number of clients and the attack intensity per selected country

2. The financial institutions within different SEPA counties encounter different attack intensities.

This hypothesis is tested by multiple regression analysis to test to following relation for each selected country: Does a significant relation exists between the number of clients of a financial institution and the attack intensity encountered by the financial institution?

For the multiple regression analysis, the country is added as extra independent variable. Thereby, regression analyses are executed per selected country. To remind, the ten countries that are selected for further research are; Austria, Belgium, Bulgaria, Finland, France, Hungary, Ireland, Norway, Romania, and the Netherlands. Some countries contain less than 10 domain years. This small number of domain years has limitations for the linear regression analysis. Because when the regression line is created with a small amount of values, each single value has much influence on the created line. Moreover, since some of the domain years belong to the same domain, there are even less varying x values.

However, the data doesn’t met the assumptions of homoscedasticity and normality. Therefore, the log transformation is executed again. The outcomes of the analysis are provided in table 7.

Table 7: Multiple linear regression model information, based on the number of clients and the country

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Regression coefficients</th>
<th>Std. Error</th>
<th>t value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.90747</td>
<td>0.26540</td>
<td>3.419</td>
<td>0.000755 ***</td>
</tr>
<tr>
<td>NOC (Austria)</td>
<td>0.18328</td>
<td>0.03657</td>
<td>5.012</td>
<td>1.14e-06 ***</td>
</tr>
<tr>
<td>Belgium</td>
<td>-0.32384</td>
<td>0.36196</td>
<td>-0.895</td>
<td>0.371990</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>3.75729</td>
<td>0.60969</td>
<td>6.163</td>
<td>3.63e-09 ***</td>
</tr>
<tr>
<td>Finland</td>
<td>-0.35405</td>
<td>0.42473</td>
<td>-0.834</td>
<td>0.405472</td>
</tr>
<tr>
<td>France</td>
<td>-0.11907</td>
<td>0.34070</td>
<td>-0.349</td>
<td>0.727070</td>
</tr>
<tr>
<td>Hungary</td>
<td>-0.25665</td>
<td>0.60962</td>
<td>-0.421</td>
<td>0.674188</td>
</tr>
<tr>
<td>Ireland</td>
<td>2.29640</td>
<td>0.39085</td>
<td>5.875</td>
<td>1.64e-08 ***</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>0.50736</td>
<td>0.37974</td>
<td>1.336</td>
<td>0.182982</td>
</tr>
<tr>
<td>Norway</td>
<td>-0.07745</td>
<td>0.42766</td>
<td>-0.181</td>
<td>0.856462</td>
</tr>
<tr>
<td>Romania</td>
<td>0.06129</td>
<td>0.54278</td>
<td>0.113</td>
<td>0.910201</td>
</tr>
</tbody>
</table>

Note: ***p<0.001, **p<0.01, *p<0.05; 209 degrees of freedom; R-squared = 0.3581

It is shown, that almost 36% of the variance can be explained by the model based on the number of clients and the country. This is a strong improvement, comparing to the model that is based on only the number of clients. Furthermore, characteristics that could explain the remaining 64% of the variance of attack intensity, could be researched. According to the experts, the sophisticated technical measures in place, the adopted authentication methods and the awareness level of the clients, are variables that could explain attack intensity.
In addition, it can be concluded, that within Austria, Bulgaria and Ireland, a significant linear relation exists between the number of clients and the encountered attack intensity. Which means, that in these countries an increase of the number of clients, results into an increasing attack intensity. However, the slope of those three regression lines varies. Therefore, among those three countries, an increase with one million clients results into a different increase of the attack intensity.

Besides, in the countries Belgium, Finland, France, Hungary, The Netherlands, Norway and Romania there isn’t a significant linear relation between the number of clients and the encountered attack intensity. This indicates, that in only three of the ten selected SEPA countries, the number of clients of the financial institution has influence on the encountered attack intensity. Characteristics that could explain the significant relation, are the number of institutions in the country, or the availability of a money mule network within the country. Moreover, two out three countries that encounter a positive relation, contain a domain that is targeted relatively intensive. This could determine the found significance as well. In that case, it would be interesting to research characteristics of those specific financial institutions. Or characteristics that could explain the context of the country they are located.

Next, the evolution of target selection based on the number of clients is analyzed. Therefore, in the next paragraph for each year in the dataset is analyzed whether a significant relation exists between the number of clients and the yearly attack intensity.

5.4.4. Analyzing per year

Within this paragraph, the second part of the profound analysis is elaborated. Hereby, the third hypothesis is tested.

3. Over the years, small financial institutions become more likely to be selected as target of Zeus malware schemes.

First, figure 2 provides insight into the targeted domains per category of clients, per year. Hereby, categories are developed of 3 million clients. It is noticed, that the number of targeted domains in the category till 3 million clients, increases every year. Note, that only in the first nine weeks of 2013, that category already contains 24 domains.

Subsequently, the second hypothesis is tested by multiple regression analysis. Hereby, the year of the attack is added as an extra independent variable, besides the number of clients. The dataset consist of the years 2009 -2013q1, regarding 2013 only the first 9 weeks are taken into account. So
for 2013 it is checked whether a significant relation exists between the number of clients and the attack intensity within these 9 weeks. Again first the log transformation is executed for the attack intensity.

In table 8, the outcomes of the multiple regression analysis are presented. The model is able to explain 10% of the variance of the encountered attack intensity. This is an increase, relative to only using the number of clients. Which means, that the relation between the number of clients and the encountered attack intensity, differs per year. Furthermore, a significant relation exists within the years 2009 and 2013. This indicates that in these countries, financial institutions seem to be selected based on the number of clients. Within the year 2009, an increase of one million clients, results into an increasing attack intensity of \( e^{0.16} \approx 1.17 \) botnet weeks per year. However, within 2013, an increase of one million clients results into a decreasing attack intensity of \( e^{1.16} \approx 3.19 \) botnet weeks per year. In table 4, the outcomes of the multiple-regression analysis are presented.

**Table 8: Multiple Linear regression model information based on the number of clients and the year**

<table>
<thead>
<tr>
<th></th>
<th>Regression coefficient</th>
<th>Std. Error</th>
<th>t value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.81473</td>
<td>0.31278</td>
<td>5.802</td>
<td>2.34e-08 ***</td>
</tr>
<tr>
<td>NOC (2009)</td>
<td>0.15745</td>
<td>0.03469</td>
<td>4.539</td>
<td>9.42e-06 ***</td>
</tr>
<tr>
<td>2010</td>
<td>-0.39059</td>
<td>0.39767</td>
<td>-0.982</td>
<td>0.32711</td>
</tr>
<tr>
<td>2011</td>
<td>-0.68350</td>
<td>0.35660</td>
<td>-1.917</td>
<td>0.05661</td>
</tr>
<tr>
<td>2012</td>
<td>-0.47953</td>
<td>0.34584</td>
<td>-1.387</td>
<td>0.16701</td>
</tr>
<tr>
<td>2013</td>
<td>-1.16402</td>
<td>0.40087</td>
<td>-2.904</td>
<td>0.00407 **</td>
</tr>
</tbody>
</table>

Note: ***p<0.001, **p<0.01, *p<0.05; 2014 degrees of freedom; R-squared = 0.1003

It can be concluded that hypothesis 3 holds. Because, in 2009 a positive relation exists between the number of clients and the encountered attack intensity. While four years later, a negative relation exists between the number of clients and the attack intensity.

Finally, it is noticed that some targeted domains belong to the same financial institution. It seems interesting to research whether those domains encounter a higher attack intensity. Therefore the following paragraph analyzes whether financial institutions with more attack domains, also encounter a higher attack intensity.

**5.4.5. Financial institutions with multiple targeted domains**

In this paragraph is analyzed, whether financial institutions that contain more attacked domains are also encountering a higher attack intensity. Thereby a fourth hypothesis is tested:

**4. Financial institutions with two or more targeted domains, encounter a higher attack intensity, than the other targeted financial institutions.**

First, it is analyzed for each targeted domain from the selected countries, whether another of the targeted domains from the same country belongs to the same financial institution. Below an overview is provided of the 17 attacked domains that all have at least one other attacked domain that corresponds to the same financial institution, see table 9 (on page 87).
Table 9: Targeted domains that have at least one other targeted domain that corresponding to the same financial institution

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ulsterbank.ie</td>
</tr>
<tr>
<td>2</td>
<td>ulsterbankanytimebanking.ie</td>
</tr>
<tr>
<td>3</td>
<td>dnb.no</td>
</tr>
<tr>
<td>4</td>
<td>nordlandsbanken.no</td>
</tr>
<tr>
<td>5</td>
<td>bnpparisbas.net</td>
</tr>
<tr>
<td>6</td>
<td>groupamabanque.fr</td>
</tr>
<tr>
<td>7</td>
<td>bnpparisbas.com</td>
</tr>
<tr>
<td>8</td>
<td>gdbdirect.fr</td>
</tr>
<tr>
<td>9</td>
<td>bawag.com</td>
</tr>
<tr>
<td>10</td>
<td>psk.co.at</td>
</tr>
<tr>
<td>11</td>
<td>bawagpsk.com</td>
</tr>
<tr>
<td>12</td>
<td>otpbank.hu</td>
</tr>
<tr>
<td>13</td>
<td>otpbankdirect.hu</td>
</tr>
<tr>
<td>14</td>
<td>snsbank.nl</td>
</tr>
<tr>
<td>15</td>
<td>snsreaal.nl</td>
</tr>
<tr>
<td>16</td>
<td>Abnamro.com</td>
</tr>
<tr>
<td>17</td>
<td>Abnamro.nl</td>
</tr>
</tbody>
</table>

A variable called ‘multiple’ is created to be able to distinguish the domains in table 9, from the other targeted domains from the selected countries. For these domains, the variable is labeled with ‘yes’ and the other domains with ‘no’. Hereby an extra class is developed with two levels. With comparing the median of the group it is analyzed whether the attack intensity differs per class. This test is executed with a Wilcoxon-Mann-Whitney rank sum test. The information of this test is provided in table 10. Hereby, it is shown that there is no difference between the median of the two groups. Therefore, it can be stated that the attacked domains belonging to financial institutions that contain two or more attacked domains, do not encounter a higher attack intensity.

Table 10: Information Wilcoxon-Mann-Whitney rank sum test

<table>
<thead>
<tr>
<th>Coefficient</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>3208.5</td>
</tr>
<tr>
<td>P-value</td>
<td>0.3735</td>
</tr>
</tbody>
</table>

5.4.6. Conclusions of the profound analysis

Finally, the following four hypotheses have been tested with the Zeus dataset, including additional data:

1. **Financial institutions with more clients are targeted more by financial malware schemes, than financial institutions with less clients.**

2. **Within different SEPA counties, the financial institutions encounter a different relation between the number clients and the attack intensity.**

3. **Over the years, financial institutions with less clients, become more likely to be selected as target.**
4. *Financial institutions with two or more targeted domains, encounter a higher attack intensity, than the other targeted financial institutions.*

Table 11 (on page 90), provides an overview of the tested hypotheses, including the scope, the executed tests and the outcomes. Furthermore, the domains with relatively high attack intensities are analyzed. This paragraph, summarizes the outcomes of the analyses and elaborates what those outcomes mean in the light of the hypotheses.

According to the linear regression analysis and the Spearman’s Rho the **first hypothesis holds**. So, between 2009 and 2013q1, *financial institutions from the selected SEPA countries with more clients are targeted more by Zeus malware schemes.*

However, the linear regression model is not able to explain many of the variance within the attack intensity (only 7.8%). Although the regression analysis intents to predict human behavior (target selection of cybercriminals), which is very hard, a prediction power of 7.8% is low. Furthermore, the Spearman’s Rho test shows a weak monotonic relation (0.33), between the number of clients and the attack intensity. This means that not every financial institutions with more clients, indeed encounter a higher attack intensity, than certain institutions with less clients. From those two analyses can be concluded, that based on the number of clients, bigger financial institutions within SEPA are more likely to be targeted. The other 92.2 % could be explained by other variables, or by a random factor. According to the experts, the context of the country and the year of the attack could explain more of the attack intensity.

Furthermore, the influence of the domains, which are targeted relatively intensive, is analyzed. It is observed, that aib.ie, procreditbank.bg and banquepopulare.fr, influence the relation between the number of clients and the attack intensity. In addition, those domains all have the most clients among the targeted domains in their country.

Besides, it can be stated that the **second hypothesis holds**: *within different SEPA counties, the financial institutions encounter a different relation between the number clients and the attack intensity.* This conclusion is based on the increasing explain ability, with the country as an extra independent variable. In addition, the varying slopes of the significant relations in different countries. Furthermore, the non-significant relation within 70% of the selected countries.

A multiple linear regression analysis is used, to test the second hypothesis. Hereby, it is noticed that based on the number of clients and the country, the linear regression analysis is able to explain almost 36% of the variance within the attack intensity. Furthermore, it is observed that within three of the ten selected SEPA countries, a significant positive relation exists between the number of clients and the yearly attack intensity. The slopes of those relations are varying. Besides, in the other seven countries no significant relation has been found. This means that in 3 of the selected SEPA countries, financial institutions are probably selected based on their number of clients. Thereby, financial institutions with more clients, are targeted more intensive.

According to the experts, characteristics that could explain the remaining 64% of the variance of attack intensity, are; the sophisticated technical measures in place, the adopted authentication methods and the awareness level of the clients, are variables that could explain attack intensity. Moreover, two out three countries that encounter a positive relation, contain a domain that is targeted relatively intensive. This could determine the found significance as well. In that case, it
would be interesting to research characteristics of those specific financial institutions. Or characteristics that could explain the context of the country they are located.

In addition, it can be concluded that the third hypothesis holds. So based on the number of clients, over the years, small financial institutions become more likely to be targeted by Zeus malware schemes. However, together with the number of clients and the year of the attack, the linear regression is only able to explain 12.1% of the variance within the attack intensity. Which means that the year doesn’t explain so much of the relation between the number of clients and the attack intensity.

For the purpose of the third hypothesis, a multiple regression analysis has been executed, with the number of clients and the year as independent variables. From this analysis can be concluded that in 2009, financial institutions with more clients encountered a higher attack intensity. Subsequently, in the years 2010, 2011 and 2012 no significant relation exists. Finally, in 2013 financial institutions with less clients encountered a higher attack intensity than financial institutions with more clients.

Finally, during executing the analyses it was noticed that some financial institutions contain multiple attacked domains. Therefore, it seems interesting to test whether financial institutions, that contain multiple targeted domains, encounter a higher attack intensity than the other targeted financial institutions. A fourth hypothesis has been developed. Together with a Wilcoxon-Mann-Whitney rank sum test, it is tested whether financial institutions that contain two or more attacked domains, also encounter a higher attack intensity on these domains. However, there seems to be no significant relation. So, the fourth hypothesis does not hold. Which means that when cybercriminals used multiple ways to target a financial institution, this not necessarily indicates that those institutions are also targeted more than others. However, financial institutions should notice that cybercriminals focus on multiple domains to target certain financial institutions.
<table>
<thead>
<tr>
<th>HYPOTHESIS</th>
<th>SCOPE</th>
<th>TEST</th>
<th>OUTCOME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial institutions with more clients are targeted more by financial</td>
<td>All the years and all the selected domains</td>
<td>Regression analysis</td>
<td>Positive significant</td>
</tr>
<tr>
<td>malware schemes, than financial institutions with less clients.</td>
<td>All the years and all the selected domains</td>
<td>Spearman’s Rho test</td>
<td>relation</td>
</tr>
<tr>
<td>&quot;</td>
<td>All the years and the domains from Austria</td>
<td>Regression analysis</td>
<td>Positive significant</td>
</tr>
<tr>
<td>&quot;</td>
<td>All the years and the domains from Belgium</td>
<td>Regression analysis</td>
<td>relation</td>
</tr>
<tr>
<td>&quot;</td>
<td>All the years and the domains from Bulgaria</td>
<td>Regression analysis</td>
<td>Not significant</td>
</tr>
<tr>
<td>&quot;</td>
<td>All the years and the domains from Finland</td>
<td>Regression analysis</td>
<td>Not significant</td>
</tr>
<tr>
<td>&quot;</td>
<td>All the years and the domains from France</td>
<td>Regression analysis</td>
<td>Not significant</td>
</tr>
<tr>
<td>&quot;</td>
<td>All the years and the domains from Hungary</td>
<td>Regression analysis</td>
<td>Positive significant</td>
</tr>
<tr>
<td>&quot;</td>
<td>All the years and the domains from Ireland</td>
<td>Regression analysis</td>
<td>relation</td>
</tr>
<tr>
<td>&quot;</td>
<td>All the years and the domains from Norway</td>
<td>Regression analysis</td>
<td>Not significant</td>
</tr>
<tr>
<td>&quot;</td>
<td>All the years and the domains from Romania</td>
<td>Regression analysis</td>
<td>Not significant</td>
</tr>
<tr>
<td>&quot;</td>
<td>2009 and all the selected domains</td>
<td>Regression analysis</td>
<td>Positive significant</td>
</tr>
<tr>
<td>&quot;</td>
<td>2010 and all the selected domains</td>
<td>Regression analysis</td>
<td>relation</td>
</tr>
<tr>
<td>&quot;</td>
<td>2011 and all the selected domains</td>
<td>Regression analysis</td>
<td>Not significant</td>
</tr>
<tr>
<td>&quot;</td>
<td>2012 and all the selected domains</td>
<td>Regression analysis</td>
<td>Not significant</td>
</tr>
<tr>
<td>&quot;</td>
<td>2013 and all the selected domains</td>
<td>Regression analysis</td>
<td>Negative significant</td>
</tr>
<tr>
<td>Financial institutions with two or more attacked domains encounter a</td>
<td>All the years and all the selected domains</td>
<td>Wilcoxon-Mann-Whitney rank sum test</td>
<td>Not significant</td>
</tr>
<tr>
<td>higher attack intensity on these domains.</td>
<td>All the years and all the selected domains</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6. Comparing the expectations of experts with the patterns shown by the Zeus dataset

In this chapter, the expectations of the experts are compared with the outcomes of the quantitative analyses. According the quantitative analyses, multiple consistencies and some inconsistencies exist, between the experts and the Zeus dataset. However, these (in-) consistencies exist, between the specific consulted experts and the specific dataset. Besides, the scopes of both analyses differs. Furthermore, the ideas of experts are processed, in order to create hypotheses that could be tested with the dataset. Therefore, the meaning of the (in-) consistencies is placed in a broader context.

The outcomes of the qualitative analysis, which are elaborated in chapter 4, contain the expectations and assumptions of experts regarding target selection by financial malware schemes. The outcomes of the quantitative analysis, described in chapter 5, provide insight into target selection that is actually made by Zeus malware schemes between 2009 and 2013q1. Hereby, high-level analyses are executed, four hypotheses are tested and the domains are analyzed, which encountered a relatively high attack intensity. The comparison of the outcomes of those analyses, provides the confrontation of experts’ expectations, with target selection that is actually made.

The comparison of these outcomes tries to answer research question four. In paragraph 6.1 the consistencies are provided. Subsequently, in paragraph 6.2 an overview of the inconsistencies is shown. Finally, in paragraph 6.3, conclusions are drawn and the (in-) consistencies are placed within a broader context. In other words, the meaning of those (in-) consistencies is provided.

6.1. Consistencies

Table 12 provides insight, into the similarities of the outcomes of both analyses. Besides, some limitations are identified.

<table>
<thead>
<tr>
<th>Qualitative analysis</th>
<th>Quantitative analysis</th>
<th>Limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bigger financial institutions are targeted more.</td>
<td>According to the regression analysis, a significant positive relation exists between the number of clients and the attack intensity among the domains from the selected SEPA countries.</td>
<td>The number of clients is not the only measure that determines the size of a financial institution. The SEPA countries with big financial sectors aren’t taken into account.</td>
</tr>
<tr>
<td>Bigger financial institutions are targeted more in the first years of financial malware.</td>
<td>According to the multiple regression analysis, a significant positive relation exists between the number of clients and the attack intensity within 2009, which is the first year of the dataset for quantitative analysis.</td>
<td>Only a small selection of domains has been taken into account. The number of clients is not the only measure, to determine the size of a financial institution.</td>
</tr>
</tbody>
</table>
Over the years small financial institutions become more likely to be selected as target. According to the multiple regression analysis, a significant negative relation exists between the number of clients and the attack intensity within 2013, which is the last year of the dataset for quantitative analysis. The data for 2013 exists only of nine weeks. Only a small selection of domains has been taken into account. The number of clients is not the only measure, to determine the size of a financial institution.

The attack intensity differs per country. According to the multiple regression analysis, the relation between the number of clients and the attack intensity is significantly positive in some countries. However in some countries no significant relation exists. Furthermore, the slope of the positive relations are varying.

The quantitative analyses didn’t directly test the expectation of the experts.

Trends

Scandinavian countries are less interesting to attack because of their strong security measures. According to the high-level analysis the Scandinavian countries encounter an average attack intensity.

6.2. Inconsistencies

Table 13 provides insight, into the difference between the outcomes of both analyses. Besides, some limitations are identified.

Table 13: Difference between the outcomes of both analyses

<table>
<thead>
<tr>
<th>Qualitative analysis</th>
<th>Quantitative analysis</th>
<th>Limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significantly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bigger financial institutions are targeted more.</td>
<td>From the multiple linear regression analysis, with the number of clients and the country as independent variable, it is noticed that not in all SEPA countries financial institutions with more clients encounter significantly a higher attack intensity.</td>
<td>The number of clients is not the only measure that determines the size of a financial institution. Some countries contain a small number of targeted domains, thereby the relation is determined based on less data.</td>
</tr>
</tbody>
</table>

Trends

<table>
<thead>
<tr>
<th>Qualitative analysis</th>
<th>Quantitative analysis</th>
<th>Limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>After improving their cyber security, the three big Dutch financial institution haven’t been selected as target anymore.</td>
<td>According to the high-level analysis, the big three financial institutions are selected as target at least till March 2013 (the end of the dataset).</td>
<td>The experts could mean that the attacks were not successful anymore. Instead of not being selected as target at all. It could be that the experts refer to the number of encountered attacks in the last two years, which are not in the dataset.</td>
</tr>
<tr>
<td>A waterbed effect could exists between the big Dutch financial institutions and smaller Dutch financial institutions.</td>
<td>According to the high-level analysis, it seems that between 2009 and 2013q1 no waterbed effect exists.</td>
<td></td>
</tr>
</tbody>
</table>

6.3. Conclusions

The following question is addressed in this chapter:
Is target selection including its evolution as assumed by experts, consistent with the patterns shown by the Zeus dataset?

In this paragraph, the findings that answer this question are elaborated. According to the tested hypotheses, bigger financial institutions encounter a higher attack intensity. Furthermore, financial institutions from different SEPA countries, encounter a different relation between the number of clients and the encountered attack intensity. Hereby, should be noticed that not in each SEPA country, a significant relation has been found. In addition, smaller financial institutions, become more targeted over the years.

The expectations of experts regarding target selection, are placed within the perspective of the routine activity theory (RAT). RAT proposes that crime occurs, when a suitable target is in the presence of a motivated offender and is without a capable guardian. Hereby, the financial malware schemes are assumed to be the motivated offender.

Besides, the experts expect that in the first years of financial malware, target selection was particularly focused on the suitability of a financial institution. Especially, the size of the financial institutions is assumed to be taken into account. Which – for instance - can be expressed by the number of clients. Furthermore, the capability of the guardian is assumed to have more influence on target selection, over the years. It is expected, that the capability of the guardian can both be expressed by the financial institutions’ technical defense measures, and the awareness of their clients. Due to the increasing awareness and the sophisticated defense measures, the presence of opportunities decreased at the big financial institutions. Therefore, experts expect that smaller financial institutions are targeted more nowadays.

Regarding the hypotheses, the number of clients is used to express size the financial institutions. However, many other factors can determine the size of a financial institution. For instance, the net income of the institution or its total assets. Moreover, nowadays there are many varying financial services. Which all have their own number of clients, in this research the total number of clients has been taken into account. Regarding the possibilities of Zeus financial malware, the number of clients holding a payment account, would be most realistic to take into account for target selection.

In addition, according to the experts, smaller financial institutions became more interesting as target, because the bigger financial institutions improved their counter measures. However, that assumption couldn’t be tested with the quantitative analyses.

Furthermore, the experts assume that the context of the country, influences the attack intensities, encountered by the financial institutions. Instead, the hypotheses tested whether the relation between the number of clients and the attack intensity differs per country. For testing whether financial institutions in certain countries are targeted more, different relations should be tested and different metrics are required. For instance, the relation between the number of targeted financial institutions and the total number of institutions per country. Besides, a metric could be used, which combines the average attack intensity encountered by the targeted institutions in a certain country, with the online banking penetration of the country.
7. Discussion and recommendations

In the previous chapters, the research questions are answered. First, the background of financial malware and cyber risk management by financial institutions is provided. Furthermore, mixed methods are used to answer the following research question:

What are the expectations of experts, regarding the evolution of target selection by financial malware schemes and how did target selection actually evolve according to the Zeus dataset?

The following 13 steps are executed to answer this question:

1. Consultation of experts.
2. Placing the experts’ expectations about target selection, within the Routine Activity Theory.
3. Developing hypotheses based on these expectations.
4. Selecting hypotheses to test with the Zeus dataset, based on exclusion-criteria analysis.
5. Creating appropriate hypotheses to test with the Zeus dataset.
6. Mentioning the data to add to the Zeus dataset and which statistical analyses to execute, in order to test the hypotheses.
7. Executing high-level analyses, to gather a high-level overview of the distribution of Zeus malware attacks, among financial institutions within the SEPA.
8. Describing a metric to test the hypotheses.
9. Selecting SEPA countries based on different criteria for testing the hypotheses.
10. Adding data to the Zeus dataset to be able to test the hypotheses.
11. Executing statistical analyses to test the hypotheses.
12. Comparing the outcomes of the consultation of experts with the outcomes of the tested hypotheses.
13. Placing the (in-) consistencies in a broader context.

This chapter discuss the main findings of the research in paragraph 7.1, the recommendations based on the research in paragraph 7.2, and the limitations of the research in paragraph 7.3. In addition, it provides a discussion in paragraph 7.4, inhere the outcomes of both the research methods including the comparison are related to financial malware risk management. Finally recommendations for future research are mentioned in paragraph 7.5.

7.1. Main findings

This paragraph discusses the main findings of the research. Hereby, the expectations of experts are elaborated and the observations from the quantitative analyses are described. Besides, the findings from the comparison of the two analyses are described.

7.1.1. Consultation of experts

The findings described here, are expectations and assumptions that are mentioned by multiple experts:

- Big financial institutions are selected more as target by financial malware schemes.
- Big financial institutions are selected more as target in the early years of financial malware, over the years smaller financial institutions become more likely to be selected as target.
- The following characteristics of a country are mentioned to have influence on target selection:
  - Money transfer policies
Cooperation between law enforcement and financial institutions
Cooperation among financial institutions
Money mule network becomes extremely important

- Certain countries are always targeted based on the innovation principle (Scandinavian countries could be such a country).
- After the big Dutch financial institutions improved their guardian, other Dutch financial institutions are selected as target.
- After the Dutch financial institutions improved their guardian, financial institutions from other countries are selected as target.
- Scandinavian countries are early adapters of sophisticated counter measures. Therefore, they encounter a lower attack intensity.
- Characteristics of the clients are becoming more important, instead of the characteristics of the financial institutions.
- Business are becoming interesting as targets for financial malware schemes.

7.1.2. Consultation of experts in the perspective of RAT
- In the first years of financial malware, targets are selected based on their suitability.
- The capability of the financial institutions guardian is not the first focus of financial malware schemes when selecting targets.
- The authentication methods are focused on the accessibility of a financial institution and thereby on the suitability of an institution, instead of on the capability of the guardian.
- Over the years, the capability of the financial institutions guardian became more important regarding target selection.
- The capability of the guardian is expressed by both the sophisticated counter measures in place and the awareness of financial institutions’ clients.

7.1.3. Quantitative analysis with the Zeus dataset
Among the quantitative analysis, a distinction is made between the trends that are identified during the analysis and the significant outcomes.

Trends
- The United Kingdom, Germany, Italy and Spain contain the most targeted domains, within SEPA.
- In approximately 30% of the domain weeks (6000) a domain was targeted by 1 botnet.
- Domains belonging to financial institutions from Ireland, Italy, Spain and the United Kingdom are attacked by more than 35 botnets in certain weeks.
- The only Dutch domain that haven’t been selected as target anymore, after it has been selected once, is a domain that stopped to exist (Postbank.nl).
- Scandinavian countries encounter an average attack intensity among the selected SEPA countries.

Significantly
- Among the selected SEPA countries, between 2009 and 2013q1 a positive significant linear relation exists, between the number of clients of the attacked domains and the encountered attack intensity. However, this relation is not able to explain many of the variance within the attack intensity. (Based on the linear regression analysis)
- Among the selected countries, between 2009 and 2013q1, a weak significant monotonic relation exists between the number of clients of the attacked domains and the encountered attack intensity. (Based on the Spearmans’ Rho test)
Among the selected SEPA countries, in the year **2009** a **positive** significant relation exists between the number of clients of the attacked domains and the encountered attack intensity. (Based on the multiple linear regression analysis).

Among the selected SEPA countries, in the years **2010, 2011 and 2012** there is **no** significant relation between the number of clients and the encountered attack intensity. (Based on the multiple linear regression analysis).

Among the selected SEPA countries, in the first nine weeks of **2013**, a **negative** relation exists between the number of clients of the attacked domains and the encountered attack intensity. (Based on the multiple linear regression analysis).

Within the countries Austria, Bulgaria and Ireland, between **2009 and 2013q1** a **significant positive** relation exists between the numbers of clients assigned to a domain and the encountered attack intensity. (Based on the multiple linear regression analysis).

Within the countries Belgium, Finland, France, Hungary, The Netherlands, Norway and Romania, between **2009 and 2013q1**, there is **no** significant relation exists between the number of clients assigned to a domain and the encountered attack intensity. (Based on the multiple linear regression analysis).

Together with the country and the number of clients of a financial institution, the linear regression model is able to explain almost **36%** of the variance within the attack intensity. (Based on the multiple linear regression analysis).

Financial institutions that contain more targeted domains, **don’t** encounter a higher attack intensity, than the other financial institutions from the selected SEPA countries. (Based on the multiple linear regression analysis).

### 7.1.4. Comparison between the experts and the Zeus dataset

When comparing the outcomes of the experts’ consultations with the outcomes of the quantitative analysis based on the Zeus dataset, both consistencies and inconsistencies between the outcomes are identified.

**Similarities**

1. Among the selected SEPA countries, financial institutions with many clients are indeed targeted more intensively.
2. Over the years (in the first nine weeks of 2013), small financial institutions are selected more as target.
3. Differences regarding target selection exist among the selected SEPA countries.
4. (Relative) small Dutch financial institutions become targeted more over the years.

**Differences**

1. Not in all SEPA countries financial institutions with more clients encounter a higher attack intensity.
2. (Big) Dutch financial institutions aren’t selected less as target over the years.

**Conclusion in the perspective of RAT**

- Between **2009** and **2013q1** more suitable targets, based on the number of clients, are targeted more.
- In **2009** more suitable targets, based on the number of clients, are targeted more.
- In the first nine weeks of **2013q1**, less suitable targets, based on the number of clients, are targeted more.
- It is unknown whether this evolution is indeed caused by the improving guardian of the suitable financial institutions.
7.2. Limitations

This paragraph describes the limitations of the experts’ consultations and the shortcomings of the data used for the quantitative analyses.

7.2.1. The consultation of experts

Three experts were heavily involved in the research. Hereby many information and knowledge is gathered from them. Those experts are related to the Technical University of Delft, TNO – Delft and the cyber intelligence group of the Dutch Central Bank, which all are relatively objective institutions and departments. However, thereby their assumptions and expectations have a decisive influence on the outcomes of the qualitative analysis.

Furthermore, an IT supervisor of the Dutch Central Bank has consulted, a security specialist from the ABN AMRO, the chief information security officer of the Binck bank and a researcher of the University of Amsterdam. One limitation regarding the composition of experts, is the lack of a cyber security expert from a security company. The results of the qualitative analysis would probably be influenced by them. However, the opposite argumentation is, that currently this research shows the difference between the expectations from experts from the financial sector (qualitative research) with target selection that is actually made (quantitative research). Hereby the differences are based on the perception of experts from the financial sector, who are in the end the ones that determine how to deal with different cyber-attack vectors and mitigate the cyber risks.

Finally, the last limitation is that all the consulted experts are connected with Dutch organizations. Those experts have many equivalent sources and thereby have corresponding knowledge. While the influence of financial malware within whole SEPA is researched.

7.2.2. The Zeus dataset

The raw data for the Zeus dataset is gathered by the Security Company FOX IT. This is done by placing honey pots all over the world. However, more honeypots are placed within the western countries. Subsequently, the raw data was aggregated by employees of FOX IT together with researchers from the TU Delft. They mentioned the limitations of their research and the dataset within table 20 Tajalizadehkhoob et al. (2013). Before starting with this research, those limitations were known. Some of those limitations were tackled, for instance by manually checking the country of the domains.

The main limitations of the dataset regards the outdated data. The consulted experts are discussing target selection of the last five years, and particularly focus on the last two/three years. While the Zeus dataset currently contains the years 2009- 2013q1. Which doesn’t contain information about the last three years. Furthermore, the experts mention all their expectations of financial malware. While many financial malware families exist, the Zeus dataset only contains information of configuration files regarding the Zeus malware family. The last main shortcoming is that this data currently doesn’t contain information about the mobile variant of Zeus. Mobile variants are becoming more and more interesting for cyber criminals, because many financial services are provided that can be reached via mobile.

The metric

The metric indicates the number of botnet weeks attacking a domain per year. This metric was developed and used by Tajalizadehkhoob et al. (2013), because it deals with the overestimation, due to just counting the send configuration files that contain domains. According to Tajalizadehkhoob et
al. (2013), “this metric eliminates some of the limitations of overestimating by normalizing the data: it merges all configurations for a single botnet sent during a week and then counts the number of botnets attacking that domain in the particular week”. This argumentation already indicates that the metric still deals with some limitations of overestimating the attack intensity.

The selected countries for the profound analysis
For the purpose of testing hypothesis, attacked domains from 10 of the 35 SEPA countries are selected. Although, not all the SEPA countries contain targeted domains, most of the countries contain targeted domains. Therefore, the lack of these other countries is a shortcoming by itself. Furthermore, none of the four countries that contain the most targeted domains, are selected for testing the hypotheses. Thereby, the subset used for the quantitative analysis, becomes small.

The number of clients
Researching the number of clients of financial institutions was a hard and an unstructured task. Because target selection among different SEPA countries was intended to research, the number of clients from institutions from different countries had to be gathered. It depends on the country how easy a certain number of clients could be gathered, however most of the financial institutions are not clearly reporting this number. Therefore, sometimes Wikipedia was the only source that provided a number. Which decreases the reliability of those numbers.

Besides, the number of clients couldn’t be found for each year of the dataset. Therefore, the year closest to 2011, has been taken for all the financial institutions. Moreover, the difference between business customers and individual clients wasn’t clear either. Therefore, the total number has been taken into account. However, this research focusses on Zeus malware attacking individual clients, and therefore is a point of discussion as well.

In addition, financial institutions are providing many services. Which all have their own number of clients. Therefore, the number of clients is a broad concept, which can be defined in many ways.

7.3. Discussion
This research, contributed empirical insights to the field of target selection, for the purpose of financial malware attacks. Hereby, characteristics of financial institutions are researched, which make them more likely to be targeted. It delivers additional findings to previous observations with the Zeus dataset, by Tajalizadehkhoo et al. (2013). Furthermore, it provides insight into the differences of the cyber fraud level, between some first world countries. Which is assumed to exist, by the Federal Financial Institutions Examination Council (2014).

First, the value of the qualitative research is discussed. By combining the outcomes of the consulted experts, value is created, by potential characteristics of financial institutions that could influence target selection. The value of those consultations is limited, because only experts are consulted from the Dutch financial sector and from Dutch research institutions. None of the consulted experts is from another financial sector within SEPA. Furthermore, none of them is from a security company. However, this research contributes to the work of Tajalizadehkhoo et al. (2013), which consulted experts from the security company FOX IT.

In addition, RAT is developed in the physical world. By using it in cyber space, it could simplify this complex artificial space, too much. Which could result in biased and short-sighted research. However, in this research, RAT seems to contribute to the explanation of target selection and its evolution. Although, the two terms “static pre-selection” and “evolving pre-selection”, had to be created to explain the evolution of target selection. Besides, a certain bias could exist, between what
the experts intend to articulate, and how the researchers interpreted their assumptions. For instance, the experts could mean that the attacks were not successful anymore. Instead of, financial institutions are not being targeted at all. Furthermore, the experts could also refer to the encountered attack intensity in last two years, which is not part of the dataset.

Secondly, the Zeus dataset is used, to test the hypotheses that arise from the experts’ ideas regarding target selection. This is valuable data, because it provides ground truth data regarding target selection by Zeus malware schemes. However, the dataset contains only web injects for the purpose of Zeus financial malware. While, many other malware families exist. Besides, it only contains data between 2009 and 2013q1. Furthermore, Zeus malware can be used in multiple ways to target clients of financial institutions, while the web injects only provide insight into the purpose of modifying webpages. In addition, it could be the case that the data still contains web injects, which aren’t effective anymore. Furthermore, only a small subset of the Zeus dataset is used for testing the hypotheses. Moreover, five of the biggest financial sectors in SEPA, haven’t been taken into account. Finally, the data for 2013 only exists of nine weeks.

Notwithstanding these shortcomings, it is assumed that the Zeus dataset is representative for the purpose of target selection by financial malware schemes. Because, Zeus is known as a persistent malware family, which is one of the most dominant malware families that ever existed (Europol, 2014; Tajalizadehkhoob et al., 2013). Besides, Zeus was one of the first developed financial malware families that becomes very popular (Lucas, 2015). In addition, among the four common ways of exploiting clients with Zeus financial malware, inject code can be used to really attack the financial institution. Besides, although the data probably contains ineffective web injects, new web injects are required continuously. For that reason, many web injects within the dataset will be effective. Finally, this research observed the influence of the number of clients on the encountered attack intensity, on country level. Due to this low level, the subsets of the data become small. However, it adds value to the findings of Tajalizadehkhoob et al. (2013), which are on global scale and on EU scale.

7.4. Recommendations
In here, the recommendations following from the research are elaborated. Those recommendations are based on both the outcomes of the research and on the research process. During the research process insight is gathered into the possibilities and barriers of researching this subject at the Dutch Central Bank.

7.4.1. Recommendations for scientific research
This is the second thesis that researched target selection by financial malware, based on the Zeus dataset. However, there are still many insights in this field that could be gathered with the Zeus dataset. Especially when more years of data are added to it. Moreover, executing analysis with the Zeus dataset is an experience and very instructive for data analysis with tools like MySQL and R. Therefore, it is recommended to continue and improve researching with the Zeus dataset. More generalized, it is recommended to continue and improve research with the instructions send to computers that are infected with financial malware. Recommendations for future research are provided in paragraph 7.5

Further scientific research with this dataset, would tremendously benefit from increasing collaboration with financial institutions and financial authorities. However, due to the confidentiality issues, the agreements about the collaboration should be established on management level, instead of by the students.
For the purpose of further research, the number of clients could be researched from targeted financial institutions of more countries. With the help of authorities this could be done much faster, more efficient and more reliable. In addition, the influence of lower-level characteristics regarding the context of a country, is very opportune for further research in this field. Certain characteristics of this context are mentioned by the experts:

- The degree of cooperation between financial institutions and law enforcement within a country
- The money transfer policies of the country
- The availability of money mules within the country
- The number of financial institutions in the country
- The degree of collaboration between (big) financial institutions on cyber security level within a country

Furthermore, the evolution of target selection is very interesting for further research. Besides, the outcomes of that research will be valuable. With the qualitative analysis many development, which are expected to be of influence on target selection, are identified. Those developments couldn’t be compared with the Zeus dataset, yet. Therefore, extensions of the Zeus dataset are required. Some of the data could be provided by the security company FOX IT. Other data could be gathered via financial institutions or authorities. For example, the characteristics of clients are expected to become decisive. I would be very interesting to gather some data of attacked clients. However, thereby privacy issues will arise.

7.4.2. Recommendations for the authorities within the financial sector

- Keep supporting smaller financial institutions with their cyber security measures. Not only with implementing technical measures, but also with socio-technical measures. Hereby, the awareness of employees and clients is really important.
- Keep dealing with cyber security together with the authorities from different SEPA countries. Because differences exist between countries, they have to learn from each other.
- Be willing to share information. Although, the authorities do have their professional secrecy and confidentiality, more collaboration with research institutions on cyber security level, could contribute to the knowledge level of target selection and its evolution.
- Improve collaboration and information sharing between research institutions and authorities within the financial sector. For researching characteristics that are expected by experts to be of influence on target selection. Both sides will benefit from this collaboration, and therefore is an item for both parties.

7.4.3. Recommendations for the financial institutions

- Keep supporting business and clients with their cyber security awareness.
- Be willing to cooperate with research institution and search for opportunities to share information. Don’t look at the short time, instead take the benefits of cooperation on the long term into account.
- Improve collaboration and information sharing between research institutions and financial institutions, for researching characteristics that are expected by experts to be of influence on target selection. Both sides will benefit from this collaboration, and therefore is an item for both parties.
7.5. Future research

- It would be interesting to research target selection on botnet level. Because, each botnet represents a malware scheme. Thereby the choices of individual malware schemes are taken into account. It could be researched which botnets attack the same cluster of financial institutions. Or which financial institutions are often selected together as target. It might even be possible to assign different botnets to the same financial malware scheme.

- Executing the same quantitative analysis based on configuration files of other financial malware families would be very interesting. Hereby the different malware families could be compared among each other. Furthermore, similarities and differences with the outcomes of the consultation of experts can be used for drawing stronger conclusions.

- Executing the same quantitative analysis based on configuration files of Zeus financial malware attacks from 2013 – 2016q2. Hereby the evolution of target selection could be analyzed further. Which provides more insight into the trends of target selection regarding financial malware.

- It would be interesting, to research the influence of certain (sophisticated) security measures implemented by financial institutions, on the encountered attack intensity. However, this is very sensitive information for the financial institutions.

- A less sensitive characteristic, which is interesting for future research, regards the executed awareness programs by the countries or by the financial institutions. The influence of those programs on the encountered attack intensity is interesting for future research.
References


Bras, 2015, *Consultation of Member of the Cyber Intelligence Unit at DNB*


Committee on Payments and Market Infrastructures (CPMI), 2014, *Cyber resilience in financial market infrastructures*, BANK FOR INTERNATIONAL SETTLEMENTS

De Nederlandsche Bank (DNB), 2014, *theme research towards information security*, extracted from: https://www.pvib.nl/download/?id=17700312


Europol, 2015, *The internet organized crime threat assessment (iOCTA)*, European Cybercrime Centre.


Lawrence, 2015, *Financial Industry’s Most-Wanted Hacker; The malware known as ZeuS and its rogue creator have been at the cutting edge of cyber-crime for nearly a decade*, extracted from:

Stone-Gross, B., 2012, Threat analysis; The life cycle of peer to peer (Gameover Zeus), Dell SecureWorks Counter Threat Unit Threat Intelligence, Extracted from https://www.secureworks.com/research/the_lifecycle_of_peer_to-peer_gameover_zeus, on September 2015.


Van den Berg, J., 2015, Lecture for SPM4450, Fundamentals of data analytics, Faculty of Technology, Policy and Management, Delft University of Technology


**Consulted experts**
Maarten Bras (Member of the Cyber Intelligence Unit at DNB)
Raymond Kleijmeer (Member of the Cyber Intelligence Unit at DNB)
Arne de Boer (Supervisor at IT oversight at DNB)
Rolf van Wegberg (PhD candidate Economics of Cybersecurity research group at the Technical University of Delft)
Maarten Jak (Intelligence Specialist II at ABN AMRO | Expertise Team Analysis | Security & Intelligence Management)
Hessel Mooiman (head information risk management and CISO at Binckbank)
Robin Döttling (PhD candidate finance Group University of Amsterdam)
Appendices

Appendix I: Countries within the Single European Payments Area (SEPA)
This appendix provides an overview of the SEPA countries, as updated by the European Payments Council in April 2016.

<table>
<thead>
<tr>
<th>Country/territory</th>
<th>BIC</th>
<th>IBAN</th>
<th>Currency code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Åland Islands</td>
<td>FI</td>
<td>FI</td>
<td>EUR</td>
</tr>
<tr>
<td>Austria</td>
<td>AT</td>
<td>AT</td>
<td>EUR</td>
</tr>
<tr>
<td>Azores</td>
<td>PT</td>
<td>PT</td>
<td>EUR</td>
</tr>
<tr>
<td>Belgium</td>
<td>BE</td>
<td>BE</td>
<td>EUR</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>BG</td>
<td>BG</td>
<td>BGN</td>
</tr>
<tr>
<td>Canary Islands</td>
<td>ES</td>
<td>ES</td>
<td>EUR</td>
</tr>
<tr>
<td>Croatia</td>
<td>HR</td>
<td>HR</td>
<td>HRK</td>
</tr>
<tr>
<td>Cyprus</td>
<td>CY</td>
<td>CY</td>
<td>EUR</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>CZ</td>
<td>CZ</td>
<td>CZK</td>
</tr>
<tr>
<td>Denmark</td>
<td>DK</td>
<td>DK</td>
<td>DKK</td>
</tr>
<tr>
<td>Estonia</td>
<td>FE</td>
<td>FE</td>
<td>EUR</td>
</tr>
<tr>
<td>Finland</td>
<td>FI</td>
<td>FI</td>
<td>EUR</td>
</tr>
<tr>
<td>France</td>
<td>FR</td>
<td>FR</td>
<td>EUR</td>
</tr>
<tr>
<td>French Guiana</td>
<td>GF</td>
<td>FR</td>
<td>EUR</td>
</tr>
<tr>
<td>Germany</td>
<td>DE</td>
<td>DE</td>
<td>EUR</td>
</tr>
<tr>
<td>Gibraltar</td>
<td>GI</td>
<td>GI</td>
<td>GIP</td>
</tr>
<tr>
<td>Greece</td>
<td>GR</td>
<td>GR</td>
<td>EUR</td>
</tr>
<tr>
<td>Guadeloupe</td>
<td>GP</td>
<td>FR</td>
<td>EUR</td>
</tr>
<tr>
<td>Guernsey</td>
<td>GG</td>
<td>GB</td>
<td>GBP</td>
</tr>
<tr>
<td>Hungary</td>
<td>HU</td>
<td>HU</td>
<td>HUF</td>
</tr>
<tr>
<td>Iceland</td>
<td>IS</td>
<td>IS</td>
<td>ISK</td>
</tr>
<tr>
<td>Ireland</td>
<td>IE</td>
<td>IE</td>
<td>EUR</td>
</tr>
<tr>
<td>Isle of Man</td>
<td>IM</td>
<td>GB</td>
<td>GBP</td>
</tr>
<tr>
<td>Italy</td>
<td>IT</td>
<td>IT</td>
<td>EUR</td>
</tr>
<tr>
<td>Jersey</td>
<td>JE</td>
<td>GB</td>
<td>GBP</td>
</tr>
<tr>
<td>Latvia</td>
<td>LV</td>
<td>LV</td>
<td>EUR</td>
</tr>
<tr>
<td>Liechtenstein</td>
<td>LI</td>
<td>LI</td>
<td>CHF</td>
</tr>
<tr>
<td>Lithuania</td>
<td>LT</td>
<td>LT</td>
<td>EUR</td>
</tr>
<tr>
<td>Country/territory</td>
<td>BIC</td>
<td>IBAN</td>
<td>Currency code</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----</td>
<td>------</td>
<td>---------------</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>LU</td>
<td>LU</td>
<td>EUR</td>
</tr>
<tr>
<td>Madeira</td>
<td>PT</td>
<td>PT</td>
<td>EUR</td>
</tr>
<tr>
<td>Malta</td>
<td>MT</td>
<td>MT</td>
<td>EUR</td>
</tr>
<tr>
<td>Martinique</td>
<td>MQ</td>
<td>FR</td>
<td>EUR</td>
</tr>
<tr>
<td>Mayotte</td>
<td>YT</td>
<td>FR</td>
<td>EUR</td>
</tr>
<tr>
<td>Monaco</td>
<td>MC</td>
<td>MC</td>
<td>EUR</td>
</tr>
<tr>
<td>Netherlands</td>
<td>NL</td>
<td>NL</td>
<td>EUR</td>
</tr>
<tr>
<td>Norway</td>
<td>NO</td>
<td>NO</td>
<td>NOK</td>
</tr>
<tr>
<td>Poland</td>
<td>PL</td>
<td>PL</td>
<td>PLN</td>
</tr>
<tr>
<td>Portugal</td>
<td>PT</td>
<td>PT</td>
<td>EUR</td>
</tr>
<tr>
<td>Réunion</td>
<td>RE</td>
<td>FR</td>
<td>EUR</td>
</tr>
<tr>
<td>Romania</td>
<td>RO</td>
<td>RO</td>
<td>RON</td>
</tr>
<tr>
<td>Saint Barthélemy</td>
<td>BL</td>
<td>FR</td>
<td>EUR</td>
</tr>
<tr>
<td>Saint Martin (French part)</td>
<td>MF</td>
<td>FR</td>
<td>EUR</td>
</tr>
<tr>
<td>Saint Pierre and Miquelon</td>
<td>PM</td>
<td>FR</td>
<td>EUR</td>
</tr>
<tr>
<td>San Marino</td>
<td>SM</td>
<td>SM</td>
<td>EUR</td>
</tr>
<tr>
<td>Slovakia</td>
<td>SK</td>
<td>SK</td>
<td>EUR</td>
</tr>
<tr>
<td>Slovenia</td>
<td>SI</td>
<td>SI</td>
<td>EUR</td>
</tr>
<tr>
<td>Spain</td>
<td>ES</td>
<td>ES</td>
<td>EUR</td>
</tr>
<tr>
<td>Sweden</td>
<td>SE</td>
<td>SE</td>
<td>SEK</td>
</tr>
<tr>
<td>Switzerland</td>
<td>CH</td>
<td>CH</td>
<td>CHF</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>GB</td>
<td>GB</td>
<td>GBP</td>
</tr>
</tbody>
</table>
Appendix II: Interview protocol

Before the interview
- Introducing myself and the master thesis

Start of the interview
- Explaining the subject of target selection by financial malware schemes
- Providing the research question of the thesis
- Explaining the mix-methods of qualitative and quantitative research
- Providing the outcomes of Tajalizadehkhoob et al. (2013)

Questions
1. How do you think that cyber criminals are selecting their targets?
   a. On what characteristics or factors of the financial institution might criminals focus?
   b. Or, do you think that cybercriminals, don’t take characteristics into account. Instead, just randomly target as many the financial institutions they know?
2. Are you informed, about certain mitigation measures, financial institutions have in place to defend against financial malware?
3. Do you expect that cyber criminals take those defense measures into account, while selecting financial institutions as target?

Summarizing the interview
- Asking to repeat some interesting things
- Mentioning that the final version of the report will be provided to the expert

The significant findings, by Tajalizadehkhoob et al. (2013):
   a. The word “bank” within the domain names of financial institutions doesn’t relate to the encountered attack intensity.
   b. English webpages for online banking only have a relation with the encountered attack intensity for financial institutions within EU countries.
   c. Globally, the broadband penetration of a country has a positive relation with the attack intensity encountered by the financial institutions established in the country.
   d. Globally, the GDP of a country has a positive relation with the attack intensity encountered by the financial institutions established in the country.
Appendix III: The number of botnets targeting domains

The first histogram has a range from 10 – 66 botnets (figure 17) and the second histogram has a range from 35 – 66 botnets (figure 18).

Figure 17: Domain weeks per number of botnets with a reach 10 – 66 botnets

Figure 18: Domain weeks per number of botnets with a reach of 35 – 66 botnets
## Appendix IV: The number of clients per targeted domain

### The Netherlands

<table>
<thead>
<tr>
<th>Financial Inst.</th>
<th>Number of clients</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>ING 9,2</td>
<td>Bank website</td>
<td>8,7</td>
</tr>
<tr>
<td>ABN AMRO 4,5</td>
<td>Different</td>
<td>4,5</td>
</tr>
<tr>
<td>Rabobank 7,5</td>
<td>For analysis</td>
<td>7,5</td>
</tr>
<tr>
<td>SNSbank 1,5</td>
<td>1,5</td>
<td><a href="https://www.snsbanknv.nl/uploads/media/downloads/0001/01/8e617c5988654511eb2529ee37880de5946d94bb.pdf">https://www.snsbanknv.nl/uploads/media/downloads/0001/01/8e617c5988654511eb2529ee37880de5946d94bb.pdf</a></td>
</tr>
<tr>
<td>Regiobank 0,56</td>
<td>Bank website</td>
<td>0,56</td>
</tr>
<tr>
<td>snsreaal* 1,5</td>
<td>Different</td>
<td>1,5</td>
</tr>
</tbody>
</table>

* snsreaal en snsbank are the same institution

### Belgium

<table>
<thead>
<tr>
<th>Financial Inst.</th>
<th>Number of clients</th>
<th>References</th>
</tr>
</thead>
</table>
| dexia | Different | 3,8 | [https://www.beste-rente-](https://www.beste-rente-)}
<table>
<thead>
<tr>
<th>Financial Inst.</th>
<th>Number of clients</th>
<th>Bank website</th>
<th>Differe nt</th>
<th>For analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>ing.be</td>
<td>2,21</td>
<td><a href="http://www.standaard.be/cnt/dmf20120114_017">http://www.standaard.be/cnt/dmf20120114_017</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fortisbanking.com</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bpo</td>
<td>1,1</td>
<td><a href="http://www.bpost.be/site/fr/residential/finance/bank/">http://www.bpost.be/site/fr/residential/finance/bank/</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cbc</td>
<td>kleine bank</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>obk</td>
<td>kleine bank</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>brownshipley-online.com</td>
<td>UK bank</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ebonline</td>
<td>0,512</td>
<td><a href="https://www.beste-rente-spaarrekening.be/banken/Europabank.html">https://www.beste-rente-spaarrekening.be/banken/Europabank.html</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>argenta</td>
<td>1,54</td>
<td><a href="https://www.beste-rente-spaarrekening.be/banken/argenta.html">https://www.beste-rente-spaarrekening.be/banken/argenta.html</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>citibank</td>
<td>0,5</td>
<td><a href="https://www.beste-rente-spaarrekening.be/banken/BeObank.html">https://www.beste-rente-spaarrekening.be/banken/BeObank.html</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ethias</td>
<td>verzekering</td>
<td>1.01</td>
<td></td>
<td><a href="https://nl.wikipedia.org/wiki/Ethias">https://nl.wikipedia.org/wiki/Ethias</a></td>
</tr>
<tr>
<td>deutschebank.be</td>
<td>0,32</td>
<td><a href="https://www.beste-rente-spaarrekening.be/banken/deutsche-bank.html">https://www.beste-rente-spaarrekening.be/banken/deutsche-bank.html</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Romania**

<table>
<thead>
<tr>
<th>Financial Inst.</th>
<th>Number of clients</th>
<th>Bank website</th>
<th>Differe nt</th>
<th>For analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>piraeusonline.ro</td>
<td>0,3</td>
<td><a href="http://piraeusbank.ro/Bank/About-us/Piraeus-Bank-Romania-En/PBR-presentation.html">http://piraeusbank.ro/Bank/About-us/Piraeus-Bank-Romania-En/PBR-presentation.html</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bancpost.ro</td>
<td>1</td>
<td><a href="https://www.bancpost.ro/Bancpost-En/About-Bancpost/Company-profile">https://www.bancpost.ro/Bancpost-En/About-Bancpost/Company-profile</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ceconline.ro</td>
<td>2,7</td>
<td><a href="https://en.wikipedia.org/wiki/CEC_Bank">https://en.wikipedia.org/wiki/CEC_Bank</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

113
<table>
<thead>
<tr>
<th>Financial Inst.</th>
<th>Number of clients</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>bcr.ro</td>
<td>4</td>
<td><a href="https://en.wikipedia.org/wiki/Banca_Comercial%C4%83_Rom%C3%A2n%C4%83">https://en.wikipedia.org/wiki/Banca_Comercial%C4%83_Rom%C3%A2n%C4%83</a></td>
</tr>
<tr>
<td>brdoffice.ro</td>
<td>2,5</td>
<td><a href="https://ro.wikipedia.org/wiki/BRD_-_Groupe_Soci%C3%A9t%C3%A9_G%C3%A9n%C3%A9rale#cite_note-ang9_2010-08-05-2">https://ro.wikipedia.org/wiki/BRD_-_Groupe_Soci%C3%A9t%C3%A9_G%C3%A9n%C3%A9rale#cite_note-ang9_2010-08-05-2</a></td>
</tr>
</tbody>
</table>

**Hungary**

<table>
<thead>
<tr>
<th>Financial Inst.</th>
<th>Number of clients</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>otpbank.hu</td>
<td>2</td>
<td><a href="https://www.otpbank.hu/portal/en/Retail">https://www.otpbank.hu/portal/en/Retail</a></td>
</tr>
<tr>
<td>raiffeisen.hu</td>
<td>0,58</td>
<td><a href="http://www.rbinternational.com/eBusiness/01_template1/829189266947841370-829188968716049154-829810062596306464-NA-2-EN.html">http://www.rbinternational.com/eBusiness/01_template1/829189266947841370-829188968716049154-829810062596306464-NA-2-EN.html</a></td>
</tr>
<tr>
<td>citibank.hu</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>bank.hu</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>otpbankdirect.hu</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

*Austria*

<table>
<thead>
<tr>
<th>Financial Inst.</th>
<th>Number of clients</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>wiki</td>
<td>Bank</td>
<td>Differe</td>
</tr>
<tr>
<td>otspbank.direct.hu</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Website</td>
<td>N/A</td>
<td>NT</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----</td>
<td>----</td>
</tr>
<tr>
<td>banking.co.at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bawag.com*</td>
<td></td>
<td>1,2</td>
</tr>
<tr>
<td>psk.co.at*</td>
<td></td>
<td>1,2</td>
</tr>
<tr>
<td>bawagpsk.com*</td>
<td></td>
<td>1,2</td>
</tr>
<tr>
<td>sparkasse.at</td>
<td></td>
<td>3,3</td>
</tr>
<tr>
<td>servicebank.at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>raiffeisen.at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sparda.at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>oberbank-banking.at</td>
<td>1,6673</td>
<td>33</td>
</tr>
<tr>
<td>valartis.at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>easybank.at</td>
<td></td>
<td>0,556</td>
</tr>
<tr>
<td>loyalbank.com</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sparda-vi.co.at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>eset.at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bankaustria.at</td>
<td></td>
<td>1,7</td>
</tr>
<tr>
<td>raiffeisenclub.at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ikarus-software</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spardabank-linz.at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hypo.at</td>
<td></td>
<td>1,15</td>
</tr>
<tr>
<td>itarus</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* merged in 2005
### Finland

<table>
<thead>
<tr>
<th>Financial Inst.</th>
<th>Number of clients</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>wiki</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bank website</td>
<td>Differe nt</td>
<td>For analysis</td>
</tr>
<tr>
<td>f-secure</td>
<td>geen bank</td>
<td></td>
</tr>
<tr>
<td>virus.fi</td>
<td>is f-secure</td>
<td></td>
</tr>
<tr>
<td>handelsbanken .fi</td>
<td>0,4</td>
<td><a href="http://www.talousanomat.fi/tiedote/2/1261862">http://www.talousanomat.fi/tiedote/2/1261862</a></td>
</tr>
</tbody>
</table>

### France

<table>
<thead>
<tr>
<th>Financial Inst.</th>
<th>Number of clients</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>wiki</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bank website</td>
<td>Differe nt</td>
<td>For analysis</td>
</tr>
<tr>
<td>Domain</td>
<td>Score</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>calyx-dc</td>
<td></td>
<td>information security</td>
</tr>
<tr>
<td>betclic.fr</td>
<td></td>
<td>betting</td>
</tr>
<tr>
<td>sophos.fr</td>
<td></td>
<td>security</td>
</tr>
<tr>
<td>axabanque.fr</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>creditmutuel.fr</td>
<td>7,4</td>
<td>11</td>
</tr>
<tr>
<td>smc.fr</td>
<td>0,35</td>
<td></td>
</tr>
<tr>
<td>banquepopulaire.fr</td>
<td>9,4</td>
<td></td>
</tr>
<tr>
<td>ukrssibbank.com</td>
<td></td>
<td>ukrain</td>
</tr>
<tr>
<td>exabanque.net</td>
<td></td>
<td>management software</td>
</tr>
<tr>
<td>banque-chaix.fr</td>
<td>0,087</td>
<td></td>
</tr>
<tr>
<td>pass.fr</td>
<td>2,4</td>
<td></td>
</tr>
<tr>
<td>lcl.fr</td>
<td>6,35</td>
<td></td>
</tr>
<tr>
<td>ebay.fr</td>
<td></td>
<td>ebay</td>
</tr>
<tr>
<td>bnpparisbas.net</td>
<td>7,7</td>
<td>6,9</td>
</tr>
<tr>
<td>cdiscount.com</td>
<td></td>
<td>onlinestore</td>
</tr>
<tr>
<td>coopanet.com</td>
<td></td>
<td>bedrijven vereniging website</td>
</tr>
<tr>
<td>societegenerale.fr</td>
<td>8,5</td>
<td></td>
</tr>
<tr>
<td>numericable.fr</td>
<td></td>
<td>tv provider</td>
</tr>
<tr>
<td>amazon.fr</td>
<td></td>
<td>e-commerce company</td>
</tr>
<tr>
<td>URL</td>
<td>Description</td>
<td>Rating</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>sfr.fr</td>
<td>telecom</td>
<td></td>
</tr>
<tr>
<td>cmb.fr</td>
<td>credit mutuel</td>
<td></td>
</tr>
<tr>
<td>sghambros.com</td>
<td>uk bank</td>
<td></td>
</tr>
<tr>
<td>edibanque.com.fr</td>
<td>bestaat niet meer</td>
<td></td>
</tr>
<tr>
<td>secureactive.fr</td>
<td>security</td>
<td></td>
</tr>
<tr>
<td>bdp.fr</td>
<td>&lt;0.5</td>
<td></td>
</tr>
<tr>
<td>monabanq.com</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>bnpparisbas.com</td>
<td>bnpparisbas.net</td>
<td></td>
</tr>
<tr>
<td>sophosbenelux.com</td>
<td>malware security</td>
<td></td>
</tr>
<tr>
<td>gdbdirect.fr</td>
<td>groupbanque</td>
<td></td>
</tr>
<tr>
<td>ingdirect.fr</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>banque-courtois.fr</td>
<td>0.211</td>
<td></td>
</tr>
</tbody>
</table>

Norway

<table>
<thead>
<tr>
<th>Financial Inst.</th>
<th>Number of clients</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsb.no</td>
<td></td>
<td><a href="https://www.hsb.no/om-oss/organisasjon/bankens-historie/">https://www.hsb.no/om-oss/organisasjon/bankens-historie/</a></td>
</tr>
<tr>
<td>handelsbanken.no</td>
<td>0.08</td>
<td><a href="http://docplayer.no/1094578-Strukturendringer-i-">http://docplayer.no/1094578-Strukturendringer-i-</a></td>
</tr>
<tr>
<td>Bank</td>
<td>Number of clients</td>
<td>References</td>
</tr>
<tr>
<td>------</td>
<td>------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Bank Bank</td>
<td>0,38</td>
<td><a href="https://skandiabanken.no/IR/IR-english/stock-statements/?id=1963742">banknaeringen-synspunkter-pa-fremtidig-strukturutvikling.html</a></td>
</tr>
<tr>
<td>Storebrand</td>
<td>0,07</td>
<td><a href="http://docplayer.no/1094578-Strukturendringer-i-banknaeringen-synspunkter-pa-fremtidig-strukturutvikling.html">http://docplayer.no/1094578-Strukturendringer-i-banknaeringen-synspunkter-pa-fremtidig-strukturutvikling.html</a></td>
</tr>
<tr>
<td>EDB</td>
<td>site bestaat niet</td>
<td></td>
</tr>
<tr>
<td>Nordlandsbanken</td>
<td>is dnb.no</td>
<td></td>
</tr>
<tr>
<td>Vossabanken</td>
<td>klein spaarbank</td>
<td><a href="https://vossabanken.no/om-oss/">https://vossabanken.no/om-oss/</a></td>
</tr>
<tr>
<td>Portalbank</td>
<td>site bestaat niet</td>
<td></td>
</tr>
<tr>
<td>Flekkefjordsparebank</td>
<td>klein spaarbank</td>
<td><a href="https://flekkefjordsparebank.no/om-oss/">https://flekkefjordsparebank.no/om-oss/</a></td>
</tr>
<tr>
<td>SOR</td>
<td>klein spaarbank</td>
<td><a href="https://www.sor.no/felles/om-sparebanken-sor/">https://www.sor.no/felles/om-sparebanken-sor/</a></td>
</tr>
<tr>
<td>SBM</td>
<td>klein spaarbank</td>
<td><a href="https://www.sbm.no/om-oss/om-sparebanken-more/om-oss-292/0/">https://www.sbm.no/om-oss/om-sparebanken-more/om-oss-292/0/</a></td>
</tr>
<tr>
<td>BNB Bank</td>
<td>0,05</td>
<td><a href="https://en.wikipedia.org/wiki/Bolig-_og_N%C3%A6ringsbanken">https://en.wikipedia.org/wiki/Bolig-_og_N%C3%A6ringsbanken</a></td>
</tr>
<tr>
<td>Vekselbanken</td>
<td>kleine spaarbank</td>
<td><a href="https://vekselbanken.no/omoss/rapportar">https://vekselbanken.no/omoss/rapportar</a></td>
</tr>
<tr>
<td>Norman</td>
<td>anti virus</td>
<td></td>
</tr>
</tbody>
</table>

### Bulgaria

<table>
<thead>
<tr>
<th>Financial Inst.</th>
<th>Number of clients</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wiki</td>
<td>Bank</td>
<td>Different</td>
</tr>
</tbody>
</table>
### Union Bank

<table>
<thead>
<tr>
<th>Financial Inst.</th>
<th>Number of clients</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>epay.bg</td>
<td>financial services, but not for this research</td>
<td></td>
</tr>
</tbody>
</table>

### Ireland

<table>
<thead>
<tr>
<th>Financial Inst.</th>
<th>Number of clients</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>facebook.com</td>
<td>social media</td>
<td></td>
</tr>
<tr>
<td>365online.com</td>
<td>bank of ireland</td>
<td><a href="http://www.ihf.ie/content/bank-ireland">http://www.ihf.ie/content/bank-ireland</a></td>
</tr>
<tr>
<td>open24.ie</td>
<td>permanent tsb</td>
<td><a href="https://en.wikipedia.org/wiki/Permanent_TSB">https://en.wikipedia.org/wiki/Permanent_TSB</a></td>
</tr>
<tr>
<td>ulsterbank.ie</td>
<td>permanent tsb</td>
<td><a href="http://allcustomercarenumbers.net/Ue_No_/Ulster-Bank-Ireland-Phone-Number-Customer-Service-Contact-032184">http://allcustomercarenumbers.net/Ue_No_/Ulster-Bank-Ireland-Phone-Number-Customer-Service-Contact-032184</a></td>
</tr>
<tr>
<td>mcafeestore.com</td>
<td>security company</td>
<td></td>
</tr>
<tr>
<td>ulsterbankanytimebanking.ie</td>
<td>ulsterbank.ie</td>
<td></td>
</tr>
<tr>
<td>365business.com</td>
<td>not a fi</td>
<td></td>
</tr>
<tr>
<td>boi-bol.com</td>
<td>domain doesn't exist</td>
<td></td>
</tr>
<tr>
<td>ebs.ie</td>
<td>0,4</td>
<td><a href="https://en.wikipedia.org/wiki/EBS_Limited">https://en.wikipedia.org/wiki/EBS_Limited</a></td>
</tr>
</tbody>
</table>
Appendix V: Graphs to check the assumptions of the linear regression analysis after log transformation and deleting outliers

After Log transformation

The residuals in the residuals vs fitted graph and the scale-location graph don’t show a pattern. Besides the residuals are distributed randomly around the zero line (dotted line). Furthermore the normal Q-Q graph shows almost a straight line, which indicates that the data is almost normally distributed after the log transformation. However, there are still outliers, although they seem less extreme.

Figure 19: Graphs to check the assumptions of normality and homoscedasticity of the linear regression analysis after log transformation
Appendix VI: Linear regression without 2013

Table 14: Linear regression information after log transformation without the year 2013

| Call: lm(formula = All$n_botnetweeks ~ All$NOC, data = All) |
| Residuals: Min 1Q Median 3Q Max |
| -2.6699 -1.1500 -0.3913 0.6431 4.9581 |
| Coefficients: Estimate Std. Error t value Pr(>|t|) |
| (Intercept) 1.28154 0.17036 7.522 2.23e-12 *** |
| All$NOC 0.18511 0.03737 4.953 1.63e-06 *** |
| --- Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 |
| Residual standard error: 1.613 on 186 degrees of freedom |
| Multiple R-squared: 0.1165, Adjusted R-squared: 0.1118 |
| F-statistic: 24.53 on 1 and 186 DF, p-value: 1.633e-06 |
Appendix VII: Getting insight into the influence of creating groups and the influence of the outliers into the found pattern

The differences in attack intensity between clusters with the size

Table 15 together with the scatter chart in figure 20, show that groups with a higher mean of the number of clients also encounter a higher attack intensity.

Table 15: Overview number of clients and attack intensity per group, based on groups of the same size

<table>
<thead>
<tr>
<th>Group number (# of domain years)</th>
<th>Average number of clients (in millions)</th>
<th>Average attack intensity (botnet weeks/ year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (44)</td>
<td>0,33</td>
<td>3,05</td>
</tr>
<tr>
<td>2 (44)</td>
<td>1,11</td>
<td>21,5</td>
</tr>
<tr>
<td>3 (44)</td>
<td>2,34</td>
<td>64,5</td>
</tr>
<tr>
<td>4 (44)</td>
<td>4,49</td>
<td>67,9</td>
</tr>
<tr>
<td>5 (44)</td>
<td>8,83</td>
<td>80,8</td>
</tr>
</tbody>
</table>

Figure 20: The relation between the number of clients and the attack intensity, based on groups of the same size

Since the groups are created, based on the number of domain years per group, it is possible that some domains in different groups are related to a financial institution with the same amount of clients, or even related to the same financial institution. To deal with this ambiguity, the next analysis creates groups based on the number of clients.

The differences in attack intensity between the groups of domain years with the same number of clients

There needs to be a perspective in place when looking to the data and searching for patterns. This perspective can be created by making assumptions from which you look the data; “the way you look to the data determines what you see” (Van den Berg, 2015). For the previous analysis, the choice was made to create five groups of the same size. Regarding the following analysis another perspective
has been chosen. Instead of creating groups of the same size, groups are created based on the same range of number of clients. Table 16 provides an overview of the groups, the limits based on which the groups are created, and the average number of clients and the attack intensity of the groups. The attack intensity is plotted against the groups, see scatter chart in figure 21.

Table 16 together with the scatter chart provided by figure 21, show a whole different pattern then the precious analysis, where groups were created of the same size. Although the group with the highest number of clients clearly has the highest attack intensity, the distribution of the other groups doesn’t make sense.

Table 16: Overview number of clients and attack intensity per group, with groups of 2.2 million clients

<table>
<thead>
<tr>
<th>Group number (# of domain years)</th>
<th>Min. # of clients (in millions)</th>
<th>Max. # of clients (in millions)</th>
<th>Average number of clients (in millions)</th>
<th>Average attack intensity (botnet weeks/year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (109)</td>
<td>0</td>
<td>2,2</td>
<td>0,94</td>
<td>11,7</td>
</tr>
<tr>
<td>2 (50)</td>
<td>2,2</td>
<td>4,4</td>
<td>3,39</td>
<td>109</td>
</tr>
<tr>
<td>3 (18)</td>
<td>4,4</td>
<td>6,6</td>
<td>5,41</td>
<td>9,06</td>
</tr>
<tr>
<td>4 (24)</td>
<td>6,6</td>
<td>8,8</td>
<td>7,74</td>
<td>36,9</td>
</tr>
<tr>
<td>5 (19)</td>
<td>8,8</td>
<td>11</td>
<td>10,3</td>
<td>140</td>
</tr>
</tbody>
</table>

Figure 21: The relation between the number of clients and the attack intensity, with groups of 2.2 million clients