INCENTIVISING BOTNET MITIGATION

TOWARDS A REPUTATION MEASUREMENT SYSTEM FOR INTERNET SERVICE PROVIDERS

Author: E.J.F. Spruit

Date: 4-12-2014

SPM5910 Master Thesis

Delft University of Technology

Incentivizing botnet mitigation:
Towards a reputation measurement system for Internet Service Providers

Author: E.J.F. Spruit

Graduation Committee
Prof. Dr. M.J.G. van Eeten (Chairman)
Dr. M.L.P. Groenleer (1st supervisor)
Dr. Ir. W. Pieters (2nd Supervisor)

Delft University of Technology

Delft/Rotterdam, December, 4 , 2014

ACKNOWLEDGEMENTS

This thesis combined literature about multiple research fields as Computer Science, Economics of information security and the field of reputation to determine possible reputation systems for incentivizing botnet mitigation by Internet Service Providers. It is the deliverable of the Master Thesis Project of Systems Engineering, Policy Analysis and Management master's program at Delft University of Technology. This report is a continued work of previous research done at the Economics of Information Security group at TPM. The outcome of the report is a set of knowledge about reputation and reputation systems in the field of cybersecurity.

Writing my thesis was at sometimes very hard, but it was a great experience. I have learned very much in writing it and it allowed me to develop myself. Many people helped me in writing my thesis, whom I have to thank. To start with, the people who have helped me by providing input for writing the actual thesis. I have to thank my supervisors, Martijn Groenleer, Wolter Pieters and Michel van Eeten for the given feedback and arranging interviews. On a personal note I would like to thank Martijn especially, he kept me on track and helped with motivating me, even when I had a difficult time at a personal level.

The people from Qnetlabs have provided much insight from another perspective than the scientific community. Therefore I would like to thank them for their provided insights and time. Furthermore all the other people who have provided me with information, as some of the Postdocs from the EconSec group. The last ones I would like to thank for their input are the organizations whom allowed me to interview them, as the ACM and one of the ISPs.

On a personal level I would like to thank family and friends whom have spent time either supporting me in the process of writing my thesis, or whom have contributed to the thesis by reviewing my work. I am very grateful.

Erik Spruit

December 2014

EXECUTIVE SUMMARY

In the introduction (Ch. 1) the topics of cyber-crime, Internet Service Providers (ISP) and botnets are introduced. A botnet is used to for cybercrime. An infected machine is infected with a bot, this bot communicates with its controller (Cooke, Jahanian, & McPherson, 2005). Other machines are infected with the same bot, which communicates with the same controller. These infected machines together form a zombie army of machines: the botnet. It is estimated that 10% of all computers is infected with malware at any point in time (van Eeten, Asghari, Bauer, & Tabatabaie, 2011).

A bot on a machine can be removed. Enforcing better security would help here. Security comes at a cost and requires specific knowledge. Generally there can be two groups: 1) People with knowledge and resources about security and 2) People without knowledge and resources about security. The first group (usually larger companies), with knowledge has to determine which security measures to invest in, usually this group is aware of botnets. The second group (smaller organizations and the home consumer), without knowledge; the issue with regard to botnets is that this group does not have the knowledge to determine if they are infected.

For many end users it is therefore difficult to determine 1) if they are infected and 2) how to mitigate the infection. The end user accesses the internet via the ISP from which they "buy their internet". This means that all the internet traffic from an end user is going via the ISP. The ISPs are in an ideal position to mitigate the problem of botnets; the ISP can determine *who* is sending out *what* kind of traffic. This knowledge makes ISPs are a good candidate to mitigate the botnet problem.

By law ISPs have to do something about this problem, but it is unspecified how much. These ISPs could mitigate botnet activity of their customers i.e. by informing end users or in a more radical situation shutting down connections until the user's bot is cleaned up (see Ch. 2.1). ISPs are aware of the problem and are working on it, but their incentives are not aligned towards botnet mitigation. Contacting customers is costly, the law is vague as it specifies only an effort has to be made: the negative incentives. ISPs do face costs if they have infected users, since botnet traffic increases the amount of bandwidth they have to provide. If an ISP hosts a lot of infected machines, other ISPs could be affected by it and force the ISP to "clean-up", in practice this does not happen often. These incentives in favor of mitigation are in place, but they are much lower than the incentives for not mitigating.

The result of the incentives for ISPs is that they are contacting some customers, but with many ISPs this only is a very small fraction of the actual infected machines they host. ISPs operate in a competitive market and are susceptible to brand damage. They fear that it becomes publicly known how they are doing number wise on infected machines. For this reason reputation based on the bot infected machines ISPs hosts was researched. Such a system also has an advantage for ISPs, it helps to create awareness among people. Such awareness could make their mitigation efforts easier. First by mapping the current situation regarding botnets, Internet Service Providers and the market of ISPs. Second, since a reputation system has to be developed, the concepts of reputation and reputation systems were researched.

There are already some methods which measure the botnet activities and in some cases already give a grade or ranking to the performance of a company based on spam output. However, there is a limitation to these initiatives. First the initiatives that result in a grade or ranking (i.e. Spamrankings or BGP ranking), focus mainly on spam and do not provide a comprehensive reputation metric. The second one is that some of these initiatives only provide metrics, so to use them as a method to convey a message to ISPs, governments or end users, they have to be transformed to a reputation (system).

In chapter 3 reputation and reputation systems were researched. In this research reputation is defined as:

"A reputation is the degree to which one party has confidence in another within the context of a given purpose" (from section 3.1)

A reputation can generally be classified onto two axis:

1. From general to specific
2. Based on human or machine feedback

A reputation system is an automated method that collects, distributes, and aggregates feedback about a participants' past behavior (Resnick, Zeckhauser, Friedman, & Kuwabara, 2000, p. 2).

The reputation system is the method that calculates and reports the reputation. To assess the quality of a reputation system there are four criteria (see section 3.2.1): Accuracy, Weighting towards current behavior, robustness against attacks, and smoothness.

A reputation system can be separated into five dimensions; the computation engine, the reputation assigner, the intended users, communication of the reputation system and cheating. The concepts of reputation and reputation systems merge together into a reputation metric. This happens in three stages; 1) data conversion; 2) calculation and 3) communication.

Chapter 4 identifies, describes and evaluates different existing reputation systems. This is done for four different existing reputation systems: PageRank, eBay's Feedback forum, corporate reputation and rating agencies from the financial world (S&P). The framework by S&P is identified to be a good reputation system. Other initiatives as eBay's feedback forum have issues with accuracy and cheating.

Chapter 5 is the bridge between theory and design. In this chapter the design requirements and design space is given. The design space shows the possibilities from which can be chosen to create a design, while the requirements show what a design has to adhere to. The requirements for the design are listed below. A reputation system for ISPs based on botnet activity should:

1. Decrease the information asymmetry in the market
2. Have a positive effect on botnet mitigation
3. Realign incentives for ISPs
4. Correspond to the right context
5. Only give a reputation to long living entities

6. Take into account previous behavior
7. show expected future behavior
8. Be a systematic algorithm
9. Be resistant to cheating from inside the system
10. Be resistant to attacks from outside of the system
11. Be well-distributed
12. Always work
13. Accessible for the intended user
14. Usable for the intended user
15. Deal with volatility of underlying metrics
16. Assign more value to new observations
17. Up-to-date
18. Maintain privacy of customers of ISPs
19. Based on automatically extracted data
20. Based on measurable concepts
21.
    a. Be able to deal with false positives
    b. Reliable
    c. Valid
22. Be able to deal with botnet takedowns
23. Account for industry trends
24. Differentiate between the size of the entities
25. Adaptable
26. Not represent a snapshot the situation
27. Not be opposed by the involved stakeholders

In section 5.2, the design space merges the dimensions from reputation and reputation systems with the context of ISPs and shows all the possibilities for design. It is listed what the possibilities for each would be in the context of designing a reputation system for ISPs. The criteria in the design space are:

- **Context**: should the reputation only be for ISPs based on botnet activity.
- **Information source:** should metrics as infected machines be used, or should also efforts of ISPs for mitigation be measured?
- **Governance model & supervision:** Who would govern the reputation?
- **Intended users:** who are the intended users of the reputation? This can be ISPs, Government, consumers or businesses.
- **Computation engine:** Which computation engine is possible in which context?
- **Communication:** How can intended users be reached?
- **Anti-Cheating measures:** How can cheating be mitigated?

The requirements and designs space are in turn used to determine possible designs for a reputation system to incentivize botnet mitigation by ISPs. Chapter 6 shows these designs.

Two designs have been provided. Alternative 1 is a design based on the rating agencies. The framework S&P uses is a red line in this design. It uses a two tiered methodology: first a base rating and with modifiers this base rating can be increase/decreased. The report is written with an open stakeholder, therefore a case is provided for each design. For alternative 1 a set of stakeholders is working together in developing a reputation system. This creates more resources, which will be required to develop a reputation system as alternative 1.

Metrics about botnet activity, so infections per subscriber and volumes per subscriber for different datasets can be used to determine the base rating. Such metrics would have different sizes and scales and have to be rescaled. Based on the confidence in the data, the size and impact of the metric a weight should be selected to calculate the base score. Other metrics as dealing with botnet takedowns, or mitigation efforts can be used as modifiers.

To create such a design would require a lot of negotiation rounds as the different stakeholders have many to agree on, ranging from technical details to the question who will provide what? For this reason in section 6.1.3 an approach to developing such a design is given.

Where the first alternative assumes cooperation between stakeholders and provides a design which will require much more resources, the second design shows a much simpler design. In this design is continued upon an existing initiative which ranks companies based on the spam they send, called spamrankings. Such a design will be a ranking of all the ISPs based on infections. Infections per subscriber and volumes per subscriber for different datasets can be used. For each metric a ranking is made, so the one with the highest amount of infections for metric A is ranked first. The rankings are than summed to determine the final score. This is a much simpler algorithm, but it is questionable if it is accurate.

Alternative 1 gives values to metrics based on confidence, size and impact. Not all botnets are equally important, while using the spamrankings method this is assumed. If a reputation system is always working, so people can use it, and it is resistant to false positive, valid and reliable it contributes to reducing information asymmetries, and incentivize botnet mitigation. For this reason such a reputation system, if it is accurate, it is likely to have a positive effect on botnet mitigation. Extra factors as some pressure from market regulators for underperforming ISPs would help to reinforce requirement 2 and 3.

These requirements are related to many other requirements, of which it is unknown if they will be met. Both reputation system can be developed, however at what cost? Creating a reputation system that is reliable, valid and low on false positives would take many resources. Alternative 1 is costly in its resources because many factors have to be decided on, and validation will be difficult. Alternative 2 does not take into account the differences between metrics and is therefore likely to be less accurate.

This is a limitation of the research; it is possible to create a reputation system, but the costs for development are not taken into account. It is only known that they will be high. Further research into this is therefore important. Another limitation is that evaluation of the reputation system is still unknown, this could also be an area for further research.

TABLE OF CONTENT

Acknowledgements 3
Executive Summary 4
1. Introduction 12
1.1 Research problem 13
1.2 Research questions 15
1.3 Research approach and main deliverable 16
1.4 Research methodology 17
1.4.1 Data 17
1.4.2 Methodology 18
1.5 Thesis outline 20
2. Describing the current situation: stakeholders, botnets and mitigation efforts 21
2.1 Botnets and Spam 21
2.1.1 Botnets 21
2.1.2 Spam 23
2.1.3 Detection 24
2.1.4 Botnet mitigation 26
2.2 Internet intermediaries 27
2.2.1 A separate category of intermediaries: Internet Service Providers 28
2.2.2 Dutch ISP market 29
2.2.3 ISPs, botnet mitigation and incentives 31
2.3 Metrics and performance indicators in cyber security 33
2.3.1 Botnet metrics research at TBM/TU Delft 33
2.3.2 Spamranking.net 35
2.3.3 Cisco SenderBase 36
2.3.4 BGP ranking and ASMATRA 37
2.3.5 CSRIC III working group botnet metrics 38
2.4 Conclusions 39
3. Reputation and reputation systems 42
3.1 Concept of Reputation 42
3.1.1 Context of a given purpose 43
3.1.2 Information source and extraction 44
3.1.3 Dimensions of reputation 46

3.2 Reputation systems 47
3.2.1. Concept of a reputation system 47
3.2.2 Reputation system computation engines 49
3.2.3 Reputation assigner 51
3.2.4 Intended users 52
3.2.5 Communication of a reputation system 53
3.2.6 Cheating and strategic behavior 53
3.3 From data to reputation 55
3.4 Conclusions on reputation and reputation systems 57
4. Existing reputation systems from other fields of study 59
4.1. Reputation systems in the digital world 59
4.1.1 PageRank 59
4.1.2 EBay 62
4.2 Corporate reputation systems 64
4.2.1 Introduction 64
MAC index 65
4.2.2 Characterization 66
4.2.3 Evaluation 67
4.3 Reputation systems in the financial world 67
4.3.1 Introduction 67
4.3.2 Characterization 71
4.3.3 Evaluation 72
4.4 Conclusions 73
5. Requirements and design space 75
5.1 Requirements 75
5.2 Design space 79
5.2.1 Context 80
5.2.2 Information sources 80
5.2.3 Governance model 82
5.2.4 Governance supervision 83
5.2.5 Computation engine 83
5.2.6 Intended users 83
5.2.7 Communication 84

5.2.8 Cheating 84
5.3 Conclusions 85
6. Design 87
6.1 Alternative 1: S&P framework 88
6.1.1 Technical design for alternative 1 90
6.1.2 Institutional design for alternative 1 96
6.1.3 Integrating the technical and institutional designs: the process design 98
6.2 Alternative 2: Borda Count as with Spamrankings 102
6.2.1 Technical design for alternative 2 103
6.2.2 Institutional design for alternative 2 105
6.2.3 Integrating the technical and institutional design: the process design 105
6.3 Conclusions 107
7. Conclusions, recommendations and future research 113
7.1 Conclusions 113
7.1.1 Conclusions from the ISP market, botnets and botnet mitigation 114
7.1.2 Conclusions from reputation and reputation systems 115
7.1.3 Conclusions on the requirements and design space 117
7.1.4 Design conclusions 119
7.2 Research limitations and further research 121
Bibliography 123
Appendix A Stakeholders and interests 128
Internet Service Providers and abuse information exchange 128
Data and knowledge providers 129
Authority Consumer and Market (ACM) 129
Uninformed end user 130
Informed end user 131
Criminals & malicious internet users 132
Appendix B Comparison of different normalisation techniques 133
Appendix C Requirements compared to existing reputation systems from other fields 135

List of figures

Figure 1 Generic engineering framework adapted from Herder & Stikkelman (2004)..... 19
Figure 2 Botnet architecture adapted from (Silva, Da Silva, Pinto, & Salles, 2013, p. 382)..... 22
Figure 3 Botnet lifecycle phases adapted from (Silva, Da Silva, Pinto, & Salles, 2013, p. 383)..... 23
Figure 4 ISPs, autonomous systems and internet users..... 29
Figure 5 Senderbase daily spam data per country adapted from (SenderBase.org, 2014)..... 37
Figure 6 Reputation dimensions..... 46
Figure 7 Objectives for a reputation system..... 48
Figure 8 From data to reputation..... 56
Figure 9 Simple representation of PageRank adapted from (Page, Brin, Motwani, & Winograd, 1999)..... 60
Figure 10 Corporate criteria framework adapted from (S&P, 2014b, p. 2)..... 69
Figure 11 S&P anchor list adapted from (S&P, 2014b, p. 4)..... 70
Figure 12 Technical architecture for alternative 1..... 91
Figure 13 Roadmap phases for development..... 99
Figure 14 Roadmap for alternative 2..... 106
Figure 15 Data to reputation; the reputation system..... 116
Figure 16 Technical architecture for alternative 1..... 120

List of tables

Table 1 Internet intermediaries and their purposes..... 27
Table 2 Comparison of computation engines..... 50
Table 3 Requirements..... 76
Table 4 Design Space..... 79
Table 5 Comparison of computation engines..... 83
Table 6 Example comparison table..... 94
Table 7 Simplified example of design alternative 2..... 104
Table 8 Alternatives compared to requirements..... 110
Table 9 Reputation requirements compared to existing initiatives..... 135

1. INTRODUCTION

Today, people, companies and governmental organizations rely on information and information systems. These systems and their data play an important role for everyone. Everybody has become very dependent on data.

There are two sides to having many information systems. On the one hand, they support the revenue generating process. A good system can create a lot of business potential (Johnson & Goetz, 2007). On the other hand, these systems also create risk. Every system and each piece of information can fail, leak, get lost or be stolen (Campbell, Gordon, Loeb, & Zhou, 2003), the risks of information systems. Having many systems and valuable information increases these risks. Digital criminals, the so-called cybercriminals, try to steal information and create disruptions. A method cybercriminals often use to do harm is botnets. With these botnets, they can burden society with high damages (Rao & Reiley, 2012).

Botnets use malware. When a machine becomes infected with malware, they can unwillingly and unknowingly join a botnet. Such a thing can happen by i.e. clicking on false advertisement, opening email packages or having an unsecured machine (Puri, 2003). An infected machine is infected with a bot; this bot communicates with its controller (Cooke, Jahanian, & McPherson, 2005). Other machines are infected with the same bot, which communicates with the same controller. These infected machines together form a zombie army of machines, usually called a botnet. Many of the end users' infected machines are part of botnets. A botnet is a network with a very large number of infected machines (sometimes even up to millions); these machines are controlled by a botnet herder (van Eeten, Asghari, Bauer, & Tabatabaie, 2011). The herder, or controller can use these machines, i.e. to send spam. These botnets are used in cybercrime.

The Dutch police force reports that for 2012 about 40% of businesses were a victim of cybercrime (Politie.nl, 2013). It is estimated that 10% of all computers is infected with malware at any point in time (van Eeten, Asghari, Bauer, & Tabatabaie, 2011). Other estimates are that around 30% of the end users machines are infected with some sort of virus, worm, Trojan or another form of malware (Net-security, 2013). Regardless of the actual percentage, if it is 5%, 10% or 30%, the number of infected machines would be in the millions for the Netherlands alone.

Botnet activity can be indicated (van Eeten, Asghari, Bauer, & Tabatabaie, 2011). A way to indicate botnets or botnet activity is by measuring outbound Spam email traffic. Spam emails are advertisement which was not asked for. Spam email is not advertisement sent by a company that you bought something from in the past, but from companies that you did not buy from. Often these companies are fraudulent, i.e. selling false Viagra. There is a strong correlation between spam and botnets, 80-90% of spam comes from botnets (van Eeten, Asghari, Bauer, & Tabatabaie, 2011, p. 8). Spammers often use infected machines, so machines that are part of a botnet, to send out their spam emails.

Increasing security is a solution to solve the botnet problem, but this is very difficult to implement. There are large differences in information levels in the market, this is called an information asymmetry (Anderson, 2001). Usually the uninformed end user has the lowest amount of information. A result of these information asymmetries is that for many it is difficult

to distinguish the good from the bad, i.e. a computer being infected or a computer being clean. The end user often misses the knowledge about security to keep their machines clean. For example: for companies it is difficult to determine what security measures to invest in. A company has a budget for IT. Security solutions are part of such a budget. If in the past the company did not have any knowledge about cybercrime, they likely underinvest in security measures, as they are costly (Campbell, Gordon, Loeb, & Zhou, 2003). In the case of the home user the lack of information and knowledge about security is especially large, resulting in under security.

The end users (home user, business or government) of internet buy an internet connection from an ISP (Internet service provider). All the internet traffic is going via an ISP. Such an ISP can determine which end users are infected with malware. They are therefore in an ideal position to mitigate the problem of botnets. Research has been done into the relationships between botnets, ISPs and end users (van Eeten M. , Bauer, Asghari, Tabataie, & Rand, 2010). A small proportion of the ISPs host most of the infected machines. These ISPs could mitigate botnet activity by shutting down connections until the user's bot is cleaned up, or get into contact with the people that own these infected machines and informing them of an infection. ISPs contacting customers would reduce the information asymmetry mentioned above. A problem here is that getting into contact with those end users, is expensive and would reduce the profit margin for these ISPs. But, ISPs could help mitigate infections; less infected machines means that the threat of botnets likely reduces.

The end users machines these ISPs host occur high costs for society. Companies, governmental organizations but also home users sometimes have high costs and problems because of these attacks by botnets. Organizations incur directly these costs; they invest in security measures and with a botnet attack their services cannot be used, i.e. attacks on banks, where people cannot use online banking. The end user or home user incurs these costs indirectly; they cannot use these services, i.e. because of a botnet attack a company or person cannot use online banking so they cannot sell products or buy products. The mal functioning of systems because of botnet attacks therefore has direct costs for a company, but there is also an element of societal costs. Because of those high societal costs there is a need for society to improve this situation; the threat of botnets has to be reduced.

The next section (1.1) introduces the research problem. By looking at the main research problem, the research questions, approach and methodology will be described in sections 1.2 to 1.4. Section 1.5 provides the thesis outline.

1.1 Research problem

Last section has identified that there is a serious problem regarding the number of infected machines. Botnets can cause serious problems for the (Dutch) economy. A big problem is that it is difficult for the end user to determine if their machines are infected with malware and correspondingly are part of a botnet. ISPs can help end users with removing infections. These ISPs can connect which infected address belongs to which end user. Giving them a unique position to solve the problem. Other stakeholders are in a less ideal position as they do not have such information. Many Dutch ISPs are already aware of the problem and have already agreed

to work together with most of the ISPs to stop this problem by starting a collective called the abuse information exchange (Techzine, 2013).

Although many seem to be working on the problem, are there differences. Some ISPs perform relatively well in terms of the infected machines they have on their network, while others perform much worse. This can be contributed to mixed incentives ISPs have. When an ISPs is aware of an infection, they can do different things. One option is to get into contact with a customer, i.e. by phone, email or an automated system that gives instructions to remove the bot. The problem is that contacting customers is very costly for these ISPs (van Eeten & Bauer, 2008); this gives an incentive to just ignore the problem. Another option is to ignore the infection. The costs of ignoring an infection are much lower than contacting a customer.

The incentives for ISPs are misaligned, therefore it is no surprise that for ISPS the number of customers they contact because they are infected, is much less that the actual infected users. The current situation is therefore ineffective. ISPs are working on botnet mitigation. Given the differences between ISPs, it appears that many do not work hard enough. This suggests that they only contact a small proportion of the infected customers.

In van Eeten & Bauer (2008) it is identified that ISPs are susceptible to brand damage. They do not want it publicly known if they have a bad performance in mitigating botnet activity in their networks. The threat of brand damage gives an incentive to mitigate the problem of botnets.

A reputation system is a method to show how well an ISP is doing in terms of botnet mitigation. For example: in the field of finance is the rating system used for financial products as the rating system by Standard and Poor's. Such a reputation system could provide ISPs a rating that shows how well they are performing in terms of botnet activity.

By having such a reputation system the ISPS with more relatively many infected machines, are pushed to be more proactive in mitigating botnets. Being a "clean" ISP, thus having a good rating, becomes an asset. The publicly known reputation could align the incentives for ISPs to mitigate the botnet problem.

This report continues on previous research done at the Economics of Information Security (Econsec) group at TPM TU Delft. In van Eeten et al (2010) the relationship between ISPs and botnets was researched. This report contains a set of knowledge about Botnet-metrics and issues with regard to measuring botnets. The previous research has identified how to measure infections on IP level per Internet Service Provider. In this report the focus is not on how botnets could be measured, but on how to use such information to turn it in to a unified reputation metric. This report continues on previously identified metrics and determines how to use such information. As will be shown in the next chapters, turning different metrics into a unified metric is not only a technical or mathematical task, but there are also societal aspects. Previous research from the Econsec group is the starting point of this research. It provides building blocks for design of a reputation metric.

Botnets and spam are large burdens for society (Rao & Reiley, 2012). Cybercrime is a big issue in today's society. Although this research wouldn't solve the issue of cybercrime, it could contribute to mitigating it. A successful reputation system (after development) could decrease

the number of infected machines. Less infections lowers the societal burden of cybercrime as it increases the security. A successfully employed reputation system might also increase the awareness of the uninformed user with regard to cybersecurity.

From a scientific point of view, this research contributes to the narrow field of reputation systems in cybersecurity. A few initiatives are already in use. However, there is no initiative providing such a method in the context of ISPs. The combination of the different fields together provide the insight into reputation in the context of ISPs and botnet mitigation.

Reputation is an ambiguous concept. The knowledge from the chapters on reputation and reputation systems could also be seen as a scientific contribution. Combining different sets of literature to explore and define different dimensions of reputation and reputation systems show how these reputation systems work.

In the next section is explained how the research problem translates to research questions.

1.2 Research questions

This section shows the research question and sub questions. Also a brief approach is shown to answering these questions. The main research question is:

"How could a reputation system to incentivize botnet mitigation for ISPs be constructed?"

As mentioned in the previous section the reputation system focusses on botnet activity. Therefore it will not be a reputation system that evaluates i.e. pricing of subscriptions. The focus is on botnet activity.

First it will be determined what kind of efforts there are already available to provide a reputation for ISPs. From this, it can be determined what is not available (the research gap). The information that is missing should be determined in the rest of the research. To address the main question several sub questions have been identified. They follow the ordering of the next chapters. They are provided below.

Research questions (**Rq**):

1. What is the current situation regarding:
    a. ISP market? **(Rq 1a)**
    b. Botnet detection? **(Rq 1b)**
    c. Measuring botnet activity? **(Rq 1c)**
2. How should reputation be defined and what are its dimensions? **(Rq 2)**
3. What are reputation systems? **(Rq 3)**
    a. What are the dimensions of reputation systems? **(Rq 3a)**
    b. How do reputation systems work? **(Rq 3b)**
    c. What are the objectives for a reputation system? **(Rq 3c)**
4. How are reputation systems used in practice **(Rq 4)**?
    a. Which are effective? **(Rq 4a)**
    b. Which can be adapted to a reputation system for botnet mitigation? **(Rq 4b)**
5. What are the requirements for a reputation system based on botnet activity? **(Rq 5)**

6. What are the possibilities for designing a reputation system based on botnet activity measurements? **(Rq 6)**
7. What are the challenges in designing a reputation system based on botnet activity measurements? **(Rq 7)**

1.3 Research approach and main deliverable

The research consists of four phases; these are described below. The study is based on literature and expert opinions. Because the topic is a mix of different fields of literature, a deliverable is the set of knowledge about the integration of different fields of study. In the end of the research a design for a reputation system is also provided. This is the main deliverable.

Phase 1: Current state of research (Rq 1a, b, c)

In this phase the current situation is described. It contains the current state of research about botnet and spam metrics, describing how botnets work and how they can be measured. The incentives for ISPs are further explained and the ISP market is described. The outcome of this chapter will be used to create a design space for phase 3. In this phase a few initiatives to measure botnet activity or rank entities are also evaluated to set up the requirements for design.

Research questions 1a, b and c are covered in this section. It is important to identify the current situation. The knowledge from the current situation should help to determine what kind of efforts there are already made to relate botnet activity to ISPs. The knowledge from this chapter can be used in a design to determine what is possible (i.e. to measure)?

Phase 2: Reputation and reputation systems (Rq 2)

This phase focuses on the literature about and around reputation system and serves as input for phase 2 where several reputation systems for ISPs are selected for development. Before selecting reputation systems, it is important to determine what reputation exactly is and what its dimensions are. This research question tries to determine based on literature how to define the concept “reputation”. The next sub phases continue on this by doing research on reputation systems, and their practical applications.

Phase 2a: Research on reputation systems (Rq 3a, b, c)

In this sub-phase reputation systems in literature are researched. Many reputation systems are available in literature. These reputation systems are all different in some way, i.e. in the way the measure reputation or by the party that is assigning reputation (the owner of the reputation system). The goal of this sub-phase is to determine what kind of reputation system and what kind of governance (be it organizing by the market or government intervention) is effective in which situation. This can later be used to determine which reputation systems for ISPs should be used, for whom the reputation system is and who should assign a reputation.
Factors as the type of reputation system, the type of governance, the expected users of the reputation system, or the consequences for having a bad reputation can influence the effectiveness of a reputation system. The dimensions for reputation systems are identified in

this section. The knowledge from this will in turn be used to characterize and classify existing reputation systems.

Phase 2b: Research on existing reputation systems (Rq 4a, b)

In this sub phase other areas of research that use reputation systems are identified. For example the method for rating financial products or the methods by which physical objects are secured and rated could be used. Practical application of reputation systems can be used; first to identify if the theory corresponds to practice. Are there differences between the two? Second existing initiatives could be transformed and applied to the situation of ISPs and the number of infected machines they host.

Phase 3: Requirements and design space (Rq 5, 6)

The knowledge from reputation systems (phase 2) together with technical and institutional possibilities from phase 1, are joined to form requirements and design possibilities. Phase 3 is the bridge between previous phases and phase 4: Design. Connecting the possibilities from chapter 1 to reputation dimensions enables possibilities to design a reputation system for ISPs.

Phase 4: Designing reputation systems for ISPs (Rq 7)

In this phase some reputation systems will be designed. There are three types of design, a technical, an institutional and a process design. The input from the first phases will be used for this section. The technical design is about the underlying technical dimensions and operationalization of the dimensions for the reputation system. The institutional design covers issues as the governance of the reputation system and the context in which the reputation system will be used. Finally the process design merges the two designs together into a roadmap for development. The main design deliverable will be a roadmap towards design and designs, without a specific stakeholder in mind. In a specific situation these designs could serve as a guideline for development.

1.4 Research methodology

This section describes the data that is used in this research and the methodology for the rest of the chapters. It gives an overview of the rest of the research and how chapters follow each other. Section 1.4.1 describes the different types of data, and how it is found. Section 1.4.2 contains the methodology.

1.4.1 Data

Data from two types of sources are used in the research:

- Data from literature
- Expert interviews

For literature several database are used for finding articles. These are: Google Scholar, Scopus/Science direct, the TU Delft Library and the IEEE repository. The research consists of a mix of different topics as: Information security economics, Botnets/spam, reputation and

reputation systems, governance of reputation and botnet/spam mitigation. The data gathering process therefore uses these topics and variations of these topics as input for the database engines.

Several reports from governmental institutions as the US FCC, European Union Agency for Network and Information Security or the OECD have published reports where their preferences, objectives and goals are mentioned. From scientific research there is much information about information security/ economics of information security focusing on botnet/spam. With regard to reputation (systems) there are many articles, amongst which a lot are ambiguous in their concepts and explanation. Not much of this research is focused on botnets; instead the general concepts of reputation and reputation systems are described. The data from basically these three (governmental institutions, economics of information security and literature on reputation) are combined to find new insights.

Often the literature that was found initially, i.e. research on information security lead to new data from other articles, i.e. by looking at the sources they have used, or by searching the literature databases with new keywords leaned from these articles. Such a research method is called snowballing. The first articles found, help to identify new literature.

Expert information is also used as an information source. Information from several experts, from the TU Delft, Qnetlabs and experts from other fields is used as extra information. This is because literature alone gives a too narrow overview and does not cover everything. As already mentioned above, the combination of literature from multiple fields gives new insights; but for this also expert knowledge is required. From the TU Delft information is given by Giovane Moura. From Qnetlabs, an expert, is used as to gain knowledge field of botnet mitigation from the perspective of an ISP. As an expert from the field of finance, Ms. M. Pieterse Bloem is interviewed.

The Authority Consumer and Market (ACM) is also consulted about cybersecurity and their role. The information from the interview is mainly used to determine the point of view for the ACM with regard to cybersecurity, botnet mitigation (by ISPs) and a reputation system. One of the ways such an interview is used to determine possible effects of a reputation system. What would an ACM do with such information?

Representatives from an ISP were also interviewed. This is done to determine their interests in a reputation system. To maintain their privacy the ISP and people whom are interviewed are not mentioned.

In the next chapter the research methodology is described. This methodology covers the approach for the next chapters.

1.4.2 Methodology

For the methodology a general engineering design framework is used (Herder & Stikkelman, 2004, p. 3880), see the figure below. The proposed framework is a good framework for this report, since it focusses on design. Such a design oriented framework fits the research, as the main focus is to identify possibilities for reputation systems. To provide a design, requires analysis. This means that the conceptual stages as problem formulation, identification of

requirements and design possibilities are also important. The figure below shows the framework.

Generic engineering framework

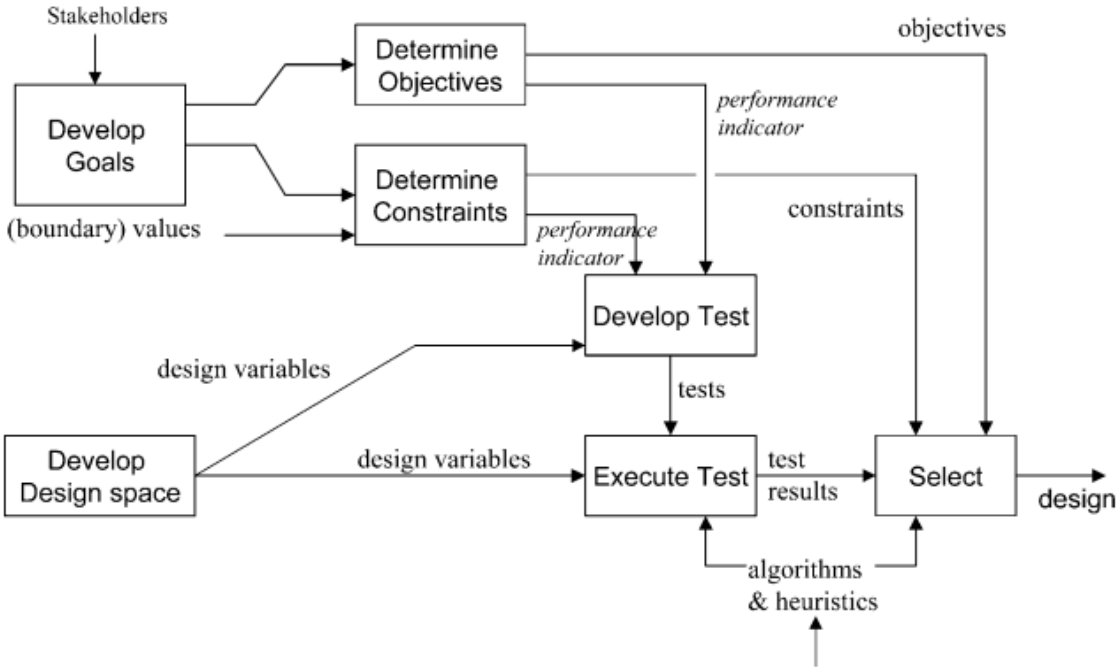

Figure 1 Generic engineering framework adapted from Herder & Stikkelman (2004)

This framework starts with the goals. These goals determine the objectives and constraints for the design. These are very important because these criteria are used to evaluate the different tests in a later stage. The goals, objectives and constraints are gathered by using literature and expert interviews. The knowledge from chapters 1-4 are merged into a design space and the requirements. The requirements translate to the constraints in the figure.

Concrete this means that design space shows design variables: the options for possible designs. Requirements show what the designs should oblige to: the Objectives and constraints.

The end result of this is a set of requirements for a reputation system that mitigates botnet activity. Looking at other fields is vital because it gives “virtual prototypes”. The field of reputation systems for botnet mitigation is still very limited, therefore other fields can be used as a comparison.

Chapter 6 gives possible designs and corresponds to the “selection” part of the figure above. In this chapter technical designs and institutional designs are described for the reputation system. In the technical design the metrics and the way they form a reputation is described. In the institutional design the intended users, communication of the reputation, cheating and cooperation between stakeholders are described. The section describes the institutional settings into which the technical design has to be placed.

1.5 Thesis outline

This thesis is structured in the following order. The next chapter, Ch.2, starts to further explore the problem of botnets. Knowing more about botnets, can help to further determine who can solve the issue of botnets. This is identified to be the ISPs. Therefore their situation is researched in ch2. Finally ch. 2 determines what is already being done to provide rankings for ISPs. With this information the research gap is introduced in the conclusions of ch.2.

It is identified that reputation can help, but there is not much knowledge available about reputation and botnet mitigation. For this reason ch. 3 researches the literature about reputation and reputation system. The end result of this chapter is a set of dimensions for these concepts. These concepts are tested in practice by looking at existing reputation systems from other fields.

Finally the knowledge from these chapters can be used to identify requirements and design possibilities in chapter 5. These design possibilities and requirements are used to identify possible designs in chapter 6. A technical and institutional design is made in chapter 6.

Chapter 6 also merges these two designs are described in a road map for development of a reputation system. The final chapter, Chapter 7, shows conclusions, recommendations and further research.

2. DESCRIBING THE CURRENT SITUATION: STAKEHOLDERS, BOTNETS AND MITIGATION EFFORTS

The previous chapter introduced the research problem. In society there is a large problem: botnets and cybercrime. In this chapter the current situation related to botnets, botnet mitigation, ISPs and other stakeholders to the problem is described. There are different topics in this chapter related to the current situation.

First: botnets and spam are defined and described. The section describes what botnets are, how they work, who is operating them, how spam and botnets can be detected and finally who can mitigate this problem. Second; the internet intermediaries are described. ISPs are part of a wider range of intermediaries; together they provide access and content to the internet.

Some scholars already incorporated some detection methods with initiatives to provide metrics and rankings based on spam and botnet activity. The existing initiatives to rank entities as ISPs, companies or countries based on technical metrics are described in section 2.3.

The information and knowledge from ISPs and the market in which they operate can in botnets and the metrics and current ranking initiatives all give information for later stages of design. The institutional settings about the ISP market and the technical challenges and possibilities are to be used in the design in order to determine what can technically be done and what is institutionally feasible.

This chapter describes first botnets and spam are described in section 2.1. In section 2.2 internet intermediaries are described. Finally in section 2.3 current state of research regarding metrics and ranking initiatives are described.

2.1 Botnets and Spam

The first chapter already saw a brief introduction to botnets, this section further explains botnets. Questions as: what they are, how they operate, what is the relationship between botnets and spam, are covered here. This section is structured in the following order: botnets, spam, detection, mitigation.

2.1.1 Botnets

When discussing botnets the following concepts are important: a bot, a controller or botmaster, the command and control (C&C) infrastructure and the botnet (Silva, Da Silva, Pinto, & Salles, 2013, p. 380). A bot is a piece of malware installed on someone's machine (usually without their knowing). This bot is installed on a vulnerable host. A vulnerable host is a machine with for example outdated security definitions. This bot opens the door for communication with the C&C infrastructure. Without instructions from the botmaster, the bot is simply "waiting for orders" (Silva, Da Silva, Pinto, & Salles, 2013, p. 381). In such an infrastructure, there is not only contact with one machine infected with a bot, there is communication with 1000 to several million machines infected with a bot. These machines together form the botnet. A botnet is controlled by the botmaster or botherder. A botnet is thus a network of enslaved host machines (the machines infected with a bot), they are controlled by botmasters to pursue criminal activities. The figure below shows these elements of a botnet.

Representation of a botnet

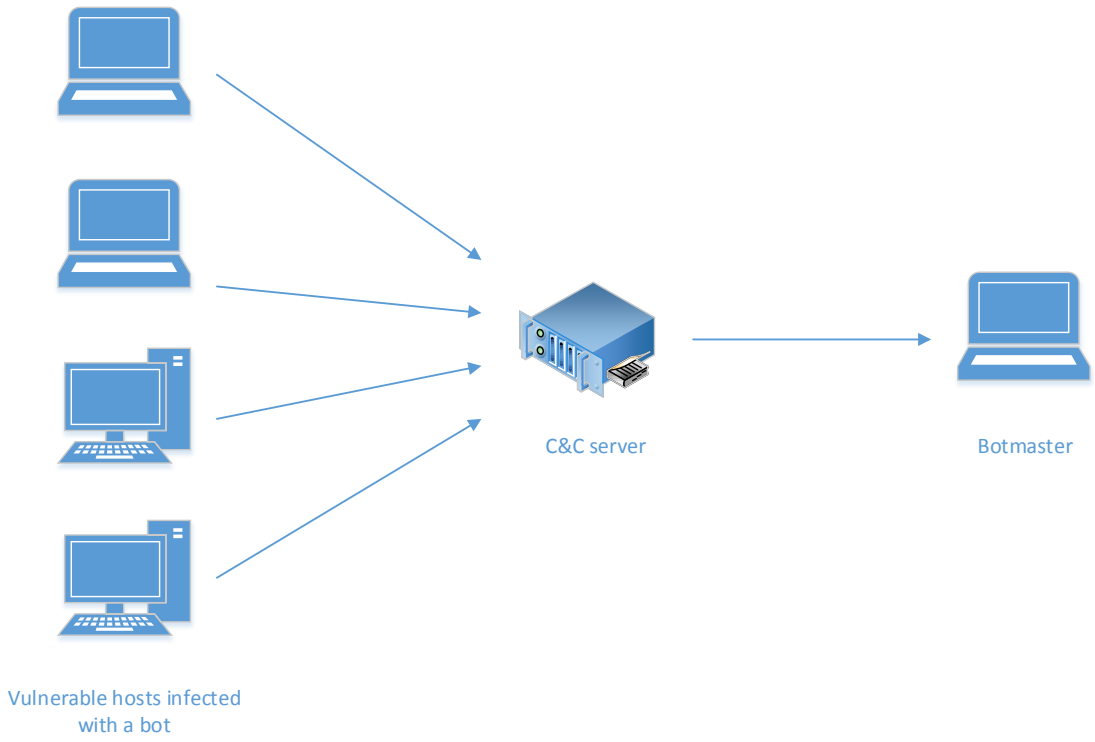

Figure 2 Botnet architecture adapted from (Silva, Da Silva, Pinto, & Salles, 2013, p. 382)

Such a botnet can be used for several criminal activities as: Spamming, Distribution of malware, DDos (Distributed Denial of service) attack, identity theft and attacks on financial systems. It is no surprise that botnets have large (financial) burdens for society (Elliot, 2010).

In botnets it is a numbers game. The botmaster wants to infiltrate as many machines as possible, and therefore botmasters look for victims that have the right features as: easy availability, low levels of security, low monitoring rates, and distant locations (Puri, 2003).

Next to these factors, the infected machine should (ideally) also have a fast internet connection. To avoid detection a botnet only uses a fraction of the capacity, thus having a fast internet connection would mean that such a machine is more usable for a DDos attack (Silva, Da Silva, Pinto, & Salles, 2013, p. 382).

Botnet Phases

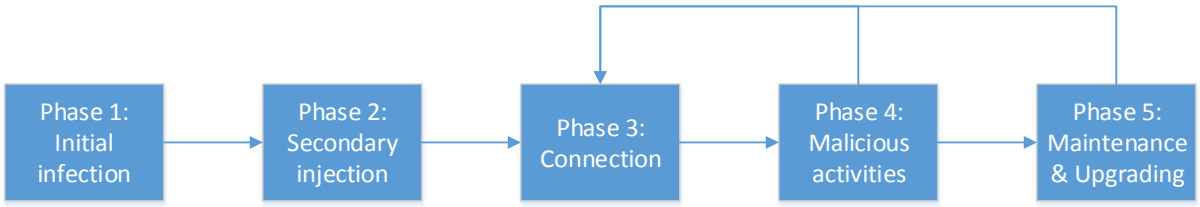

Figure 3 Botnet lifecycle phases adapted from (Silva, Da Silva, Pinto, & Salles, 2013, p. 383)

Once a machine is infected it travels through several phases. The figure above shows how these follow each other, below the phases are described:

1. Initial infection: a host is infected, it becomes a potential bot. Infection is usually due to an unwanted download, malware from a website, infected files attached in an email or an infected removable device (USB).
2. Secondary injection: if the first phase is successful, this phase starts. In this phase the infected host runs a program that searches and downloads the malware onto the host.
3. Connection: with the malware on the host, the host connects to the C&C server to see if it has instructions. The host in not always online, so every time the machine starts up this connection procedure starts.
4. Malicious activities: the bot is ready for an attack.
5. Maintenance and upgrading: to avoid detection, adding new features or changing C&C servers' bots sometime have to be upgraded.

Purpose of a botnets

Botnets can have very different purposes. For example, there are botmasters who are just doing it for fun. Below a list of different purposes for botnets are given:

- DDoS attack (Distributed Denial of Service attack)
- Sell information to adware companies
- Sell infected machines/services i.e. for sending out spam
- Phishing scams
- Identity fraud
- Hosting illegal content
- Political protest
- Terrorism
- Espionage (Elliot, 2010, pp. 82-85)

2.1.2 Spam

In the year 2010 100 billions emails were sent; of these 100 billion emails 88 percent was Spam (Rao & Reiley, 2012). Spam are unwanted emails (advertising) which someone did not give permission for and does not have an unsubscribe method. So getting a newsletter from a

company were you had some interactions with in the past usually is not a version of spam. A lot of spam is not received by the user, because it is stopped by the spam filter (otherwise users would receive 100 of spam messages per day) (Rao & Reiley, 2012).

It is estimated that the public cost of spam is around 20 billion dollars annually for the US, while the private benefits for spammers are relatively low: around 200 million dollars per year worldwide (Rao & Reiley, 2012, p. 88). Spam is a market externality. An externality is a cost spilled over to a third party. An externality occurs when the pubic costs differ largely from the private benefits The externality ratio of public costs to private benefits for spam are estimated to be 100: 1 (20 Billion: 200 Million dollar). Such a ratio means that for every dollar a spammer makes, he or she costs society 100 dollars (Rao & Reiley, 2012, pp. 88-89).

The people receiving spam occur costs for spam, but there are also people gaining from sending spam messages: the spammers. These spammers have complex mechanisms to send out spam. In the beginning they made automated scripts that generated email addresses or had their own servers. Every time a one of these servers was shut down because of blacklisting, they would just turn up in another place. In the end this method was evolved further by a significant innovation: Botnets. Currently 80-90% of all spam messages originate from botnets (van Eeten, Asghari, Bauer, & Tabatabaie, 2011, p. 8). This shows the relationship between spam and botnets: Most of the spam messages originate from botnets sending out spam. An IP address sending out spam is an indication of being in a botnet. As shown in the previous section, there are also other purposes for botnets. The relationship is therefore one way; spam indicates being infected with a bot. The other way around does not apply. If a machine is infected with a bot, it does not have to send out spam. The botnet can also be used for other purposes as stealing passwords. To conclude: sending out spam is a strong indicator of being infected by a bot.

2.1.3 Detection

There are basically two categories of detecting botnets: honeynets and Intrusion detection system (IDS) based. Honeynets are good for detecting information about botnets, from this information it is possible to learn and understand the technology in use. With such a honeynet the characteristics of a botnet could be characterized, signatures and C&C servers could be identified with it (Silva, Da Silva, Pinto, & Salles, 2013). A Honeypot is a trap, with the goal of attracting the attention of attackers. The goal of a honeypot is being exploited. By being exploited it gives information about a botnet (Shadowserver, 2014). The bot can do no harm, the underlying computer systems do not have any value, i.e. because they do not have any interaction with other systems; so exploitation is useless for the attacker. An example of a honeypot can be a spamtrap**,** these are accounts specifically set up to be exploited for receiving spam.

Honeypots are very effective for collecting and tracking botnets; it gives an overview over the botnet. Honeynets do have limitations because they can only track a limited scale of exploited activities and they cannot capture the botnet. So they cannot be used for botnet takedowns which a sinkhole can do. They only give reports about the infected machines based as trap.

An IDS looks either at signatures or anomalies. The basic idea is that information from packets of monitored traffic is extracted. So specific patterns are extracted and put into a database which

contains knowledge about existing bots. The basic difference between these two is that a honeynet operates outside of a network, where an IDS is within a network (Silva, Da Silva, Pinto, & Salles, 2013).

Although an IDS usually operates within a network, there are also datasets which aggregate data from multiple IDSes and firewalls. The DShield database is a set of data obtained from a network of sensors. The data is used by the Internet Storm Center to monitor levels of malicious activity. It basically contains data from firewalls and IDS. The result of the database is a set of offending IPs for each day (DShield.org, 2014).

In botnet detection it depends where the measurement point is. Is someone inside a network or outside of a network? Data internal to a network would (and should) not be accessible to people from outside of a network. Data about IDS would be such data. On the other hand, i.e. Spam could be measured outside of the network with a honeypot.

A mitigation/deactivation technique for botnets are sinkholes. A sinkhole gives data internal to the botnet. An entire botnet is taken over in such a case. The machines are still infected, but the C&C communication is taken over. Such a takeover would give an entire overview over all the information about all the machines infected with such a bot. An example is the Conficker dataset (Shadowserver.org, 2014a). Although the botnet C&C structure is taken down, many machines are still infected. I.e shadowserver.org shows information about Autonomous Systems (AS) and corresponding infected IP addresses. Although the botnet is taken down and the information about the botnet does not give insight into active botnets; it shows machines which are infected with bots. Machines infected with a bot from i.e. Conficker are insecure. An insecure machine is thus likely also part of or susceptible to other botnets.

Sometimes a botnet takedown occurs. In such a situation an entire botnet is taken over by infiltrating the C&C server and thus the botnet is eliminated.

In measuring botnet activity there is an issue of false positives (FP) (St. Sauver, 2012). A false positive could be compared to a false alarm. With a FP data suggests i.e. that a machine is infected with a bot, but in reality it actually isn't. In the case of Spam there can be false positives. Estimates are that 80-90% of spam originates from botnets. This means that if an IP tests positive for sending spam, there is a 10-20% change that this spam does not originate from a bot. For spam as indicator of being infected with a bot there is a 10-20% chance of false positives. The data suggests that an IP is infected with a bot, because it sends out spam. In practice it is sending spam because of another reason. To overcome the issue of false negatives, data should be cleansed and compared between sources. If multiple sources indicate infection, the possibility for false negatives reduces (Chess & McGraw, 2004).

In detecting botnets and spam there is also an issue of false negatives (FN). A false negative occurs in the situation where the detection algorithm neglects to identify a machine as infected. In such a situation a detection algorithm would identify i.e. an IP as being clean, while actually it is infected with a bot. False negatives give a false sense of security (Chess & McGraw, 2004), as it underestimates the number of infected machines.

2.1.4 Botnet mitigation

Previous sections have shown that botnets are a serious problem, which causes society large problems i.e. in monetary values. There is an upside; botnets can be mitigated. There are two approaches to such mitigation:

1. Finding the cause of the botnet: the side of the “hacker”
2. Removing malware from infected machines: “the hacked user”

One approach is to determine and prosecute the “owners of the botnet”. This is a painstakingly difficult procedure (Elliot, 2010, p. 98), as the hacker might be based from country A, using infected machines in country B to attack a system in country C. Such roundups are typically done by high tech police forces (Politie, 2013), often together with specialized companies. Finding and prosecuting the hacker is a difficult process, and often cannot be done at all (as many cybercriminals do not get caught). If the attackers are found, the machines they infected remain infected. This is the other side of botnet mitigation: removing the malware threats from the infected machines.

The infected machine at with the end user. Such an end user can employ security to defend their machines from malware. The introduction (ch.1) already identified these end users to be often uninformed. Therefore they either do not know about security, or they underinvest in security. If someone is infected with malware this can be very hard to detect. Often specialized software is required to trace such infections. This even increases the difficulty to remove the bots from the end users machine.

As shown in the previous sections, the botnets use the internet. For example, sending spam generates email messages. Therefore it is possible to measure botnet activity and to see where it originates from (see section 2.2.3). With this knowledge an infected IP address can be determined. It is for the outside world unknown which IP belongs to which customer. However there is one stakeholder which can determine this: the Internet Service Provider (ISP). This ISP assigns the end user with an IP, and has this information.

ISPs are in a unique position to solve a big part of the botnet problem, since they can address the home users that are infected (ENISA, 2011, pp. 127-128). The Dutch Telecom law, section 11.3, states that ISPs have to put in an effort in mitigating the botnet problem. This means that it is ambiguous for ISPs to determine what level of effort they should put in.

For other parties (being not an ISP) this is more difficult since they do not know which IP address belongs to whom. An ISP has a better view over which of his customers has which IP address. Someone outside of an ISP, does not have this information they can only determine which IP address belongs to which autonomous system. An ISP is not allowed to give this due to privacy of its customers, so people that are outside of the ISP cannot get into contact. ISPs are thus in a unique position because they have more information about which IP belongs to what customer, but also because they can block botnet communications, in fact disrupting botnets (Silva, Da Silva, Pinto, & Salles, 2013, p. 398)

An ISP could filter the botnet traffic, by blocking inbound and outbound malicious users connection (Spam, malicious code, attacks etc.). By doing this they disrupt the communication

with the infected user and the C&C server. As already described in the section about botnets, a machine infected with bot, becomes an enslaved machine. Such a machine waits for orders from the C&C server. If it receives no orders, it is i.e. not sending out spam. Thus by disrupting the communication, the botnet activity is also disrupted (Silva, Da Silva, Pinto, & Salles, 2013, p. 398).

Another method ISPs could use is to contact the customer. Informing them on that they are infected, what kind of infection it is, and how the end user can mitigate the infection. In such a case the ISP would actively support customers in removing their botnets. As identified above the customer usually is unaware of being infected. The ISP informing them would thus help to become aware.

Given the unique position of ISPs in relation to the botnet problem, are in the next sections the ISPs researched. What is an ISP, what is their relationship to the problem, what are ISPs already doing about the problem of botnets, is all described in section 2.2: the internet intermediaries.

2.2 Internet intermediaries

There are a multitude of involved actors which together are the "internet intermediaries". The role of these Internet intermediaries is to enable economic, social, and political interactions between third parties on the Internet (OECD, 2011, pp. 5-6). Internet intermediaries provide access to hosts and transmit or index content that comes from these third parties. The role of these Internet intermediaries is therefore critical.

There are many different types of Internet intermediaries as: Internet access and service providers, data processing and web hosting providers, Domain Name Registrars, internet search engines, E-commerce platforms & payment systems and Participative web platforms (OECD, 2011, p. 85). There is a difference between Internet access/service providers and the other Internet intermediaries; these Internet access/service providers provide the end users with access to the Internet, while the other intermediaries are more focused on providing the content on the Internet. The table below shows the different internet intermediaries and their purpose.

Table 1 Internet intermediaries and their purposes

Nr	Name	Purpose
1	Internet access and service providers	Provide users (home consumers, government, businesses) to the internet
2	Data Processing and Web hosting providers	Process and store data on the internet
3	Domain name registrars	Register domain names
4	Internet Search engines	Indexing and navigation on the internet
5	Internet commerce platforms	Provide platforms for online buying and selling
6	Payment services	Provide the link between the bank and the

		end user to support payments
7	Participative web platforms	Social media and news platforms

The rest of section 2.1 focusses on Internet Service Providers because they perform a vital & physical task: connecting users to the Internet. Section 2.2.1 explains what ISPs are and how the internet works. The next section (2.2.2) continues on this knowledge and explains the market for ISPs. Finally in section 2.2.3 a justification is given to the question: why an ISP should mitigate the botnet problem?

2.2.1 A separate category of intermediaries: Internet Service Providers

In this report often the word ISP, or Internet Service Provider is mentioned. An ISP is a company that offers internet access to the end user, so it is as the abbreviation ISP already says the provider for internet. Basically, consumers have a home router which allows them to connect to the ISP and these consumers get an IP number from this ISP. The ISP in turn, is connected to many other ISPs via the backbone of the internet (Economides & Tag, 2012, pp. 92-93). The internet consists of an uncountable number of small networks and routers. An ISP acts therefore as an intermediary between the internet's end user and the backbone of the internet.

Webhosting companies, are often also called an ISP since they also provide a service on the internet. Therefore the term ISP can be a bit ambiguous. In this report these parties are not seen as an ISP; ISPs are only the parties that function in the intermediary role between end user and the backbone of the internet.

An ISP is an autonomous system, or even (which is usually the case) a collection of many autonomous systems. On the internet, data is send between multiple Autonomous Systems (AS). An AS is: "a set of routers under a single technical administration, using an interior gateway protocol and common metrics to determine how to route packets within the AS, and using inter-AS Routing protocol to determine how to route packets to other ASes (Rekheter, Li, & Hares, 2006, p. 3)". Every autonomous system has a number (ASN), a unique integer for identification. These ASN are used for the identification of which IP address corresponds with which ISP. This is under the assumption that knowledge about which ASN corresponds to which ISP. The figure below shows a representation of ISPs and Autonomous systems. ISP 1 in the figure has several autonomous systems. These systems again connect to home users or companies. Thus, an ISP consists of one to several autonomous systems. The home user is connected to an autonomous system. Autonomous systems are linked to each other.

ISPs, autonomous systems and internet users

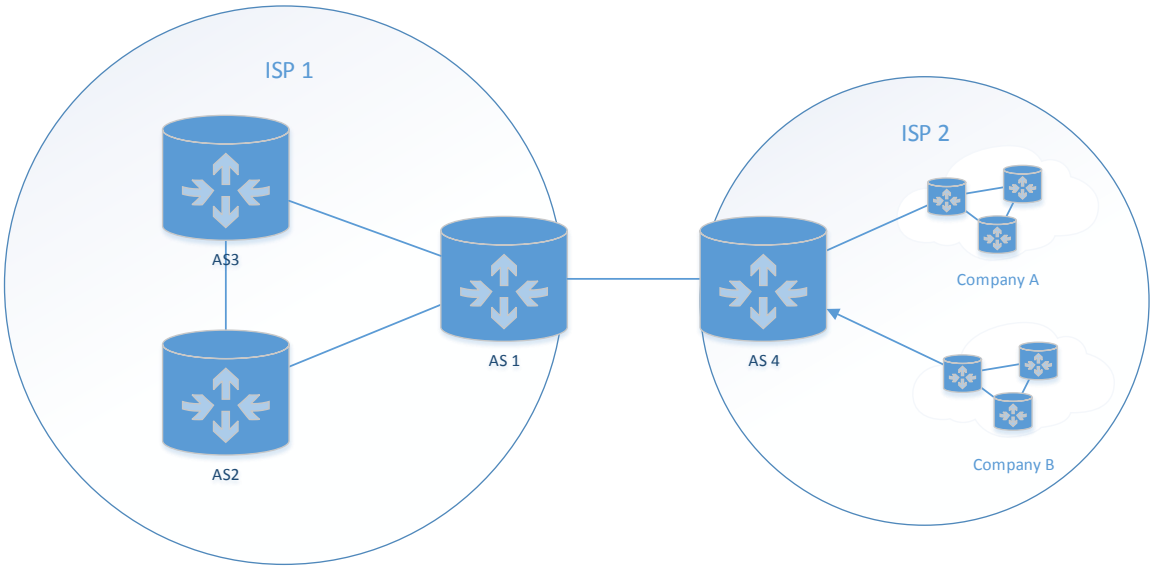

Figure 4 ISPs, autonomous systems and internet users

There are three kinds of autonomous systems: a stub, a multihomed and a transit AS. A stub AS is only connected to one other AS, where a multihomed AS has connections to many ASes but it does not forward traffic. The last type of AS is a transit AS, such an AS provides a transit service between ASes connected to it (Wagner, Francois, Dulaunoy, Engel, & Massen, 2013, p. 1).

So ASes can be an ISP: which facilitates the routing on the internet, between its customers and between customers and (customers of) other ISPs. AS routing uses the Border Gateway Protocol (BGP) (Wagner, Francois, Dulaunoy, Engel, & Massen, 2013, p. 260). BGP ensures the connectivity on the internet, because BGP is the method that allows different ASes to communicate with each other; in essence BGP is a very big list which keeps records about the different IP addresses, so that ASes can get into contact with them when a request is made (Rekheter, Li, & Hares, 2006). I.e. large hosting providers can also have their own autonomous system.

2.2.2 Dutch ISP market

In the Netherlands there are several ISPs. This research continues on the research and data collected by van Eeten et al. (2011), therefore the same ISPs for the Netherlands are used. These are:

- Bbned
- KPN
- Luna
- Online
- Scarlet

- Solcon
- Tele2
- UPC
- Xenonite
- XS4ALL
- Ziggo

These ISPs together obtain 90% of the market (as of 2011) (van Eeten, Asghari, Bauer, & Tabatabaie, 2011, p. 6), this means that there are also other ISPs. Although these 11 ISPs are all operating in the Dutch market, is it not possible for a consumer to buy their internet services from all of these parties. In many cases it is more likely that the consumers can only choose between a few of these; this is because the market for internet is not completely efficient. In many cases there are monopoly/duopoly in the market (Economides, 2007, p. 4).

To understand how the market for ISP works, it is important to understand how the relationship between end user, ISP and the backbone of the internet is. In a simplified version, an end user buys internet from an ISP. They usually pay a fee and get access to the whole internet in return. To access the internet an end user often is provided with a router at home (by the ISP). ISPs in return, pay for using the backbone of the internet, every month (Economides, 2007). These ISPs buy bandwidth according to their expected use. In return they can peer with other ISPs. These backbone networks provide transport and routing services and represent thus the "physical lines of the internet".

The list ISPs above can be separated into several sub lists which influence the market for ISPs. First of all there is a difference between the backbones which ISPs use for internet delivery. Some of these ISPs offer internet via the cable (TV cable/coax, UPC or Ziggo). The other method that is used in this backbone is the telephone cable and fibreglass, xDSL (i.e. ADSL,VDSL). Other parties as KPN or XS4ALL operate on this backbone.

A house in the Netherlands in general has only one incoming TV cable. This cable is owned by the cable company: which is either UPC or Ziggo. The total of the cables per company forms the network for this company. There is no agreement that these parties can use each other's network; this means that at a house there is either a cable in the ground owned by UPC or by Ziggo. In a house with a cable from UPC, only UPC internet can be delivered via this cable, and vice versa. This means that these cable companies are active in other regions and do not really compete with each other. They therefore have a monopoly position in terms of cable internet. Also it can be possible that a building cannot receive UPC or Ziggo; in that case they have to rely on internet via the telephone cable.

Internet via the telephone cable works differently. In this case also there is a backbone network (the cables going to buildings), this is owned by KPN. However, this network is open to other ISPs. KPN offers bandwidth on their cables to other ISPs. There are thus ISPs that sell internet to end users and use the network of KPN to deliver this. Such an ISP buys bandwidth from KPN. This means that KPN has a double role: of internet provider to ISPs, and to end users.

In the above list of ISPs there is also a difference in market share. Providers as KPN or Ziggo have a much larger number of connections than a provider as Luna or Scarlet.

The market for ISPs is complicated. There are not many ISPs; some are regional with the effect that not all the ISPs can offer their services to all the home consumers. Still there are possibilities for a consumer to choose/switch between ISPs. This is because the market is regulated (i.e. the Dutch Telecom Law and the regulator for the market: "Autoriteit Consument en Markt (ACM)". Without regulation the market would not be competitive because, i.e. KPN, might not have to open its network to other ISPs.

2.2.3 ISPs, botnet mitigation and incentives

Previous sections have identified ISPs as a party with a unique position to mitigate the botnet problem. What can and do such ISPs do about botnet mitigation? Do they have incentives for botnet mitigation and if so, what are they?

Formally, the problem of botnet mitigation is not for ISPs (see appendix A). In practice it can be seen as their responsibility. Many ISPs are working on the problem of botnet mitigation. However not every ISP has the same level of botnet activity in their network. There are ISPs with relatively high and low amounts of infected machines (see appendix A). This means that some ISPs are (relatively) clean, and some (relatively) infected. This suggests that either ISPs have different levels of effort with regard to informing their customers, or different ISPs use other methods of helping their customers.

There are some options an ISP can opt for to mitigate the problem (MAAWG, 2007). Examples are:

- Customer awareness campaigns: almost all ISPs have such campaigns
- Providing security solutions to their customers: many ISPs provide for free or reduced pricing security software as anti-virus software, or firewalls
- Active participation in anti-botnet initiatives as Abuse hub
- Warn customers: either by mail or with automatic solutions as walled gardens
- Improving the quality of routers for their customers
- ISO27001 certification

The most used methods an ISP can use are either to call the infected user with the message of being infected or to quarantine a customer that is measured to be infected. Such a user would not be able to use the internet until it is "clean". An ISP redirects the infected customer to a specific page, where instructions are described to clean-up the machine. After it is cleaned up, the user can mention this on the clean-up page and un-quarantine itself. Such a solution is called a "Walled garden" solution.

So it is possible for ISPs to mitigate the problem but do they have incentives to do so? To start: the cost for a customer support call is very high, estimates are that an incoming service call for an ISP costs about 8 euros and the costs for an outbound call to a customer are as high as 16 euros (van Eeten & Bauer, 2008). When an ISP starts mitigating botnets by quarantining infected machines of their customers; these customers will start to call the ISP and thus create costs for

these ISPs. Also customers might change ISP and ISPs might lose customers because of it. There are thus incentives for not doing anything at all, ignore customer calls and let the customers be infected. Such an ISP (often called a rogue ISP) makes typically more money because of the reduced costs for customer support. Moore (2010) proposed a solution were the costs are to be divided among many parties as government, ISPs, software vendors (Moore, 2010, p. 111). The solution to whom should pay for botnet mitigation is another topic of research and thus further out of the scope of this report.

These rogue ISPs face another problem. Such an ISP which facilitates that these externalities also affects other ISPs. If an ISP has many infected customers, it will not only be receiving a lot of malware (i.e. Spam), but it will also be sending out a lot of malware. The customers of these ISPs are (unwillingly and/or unknowingly) sending out malware. Other customers from the same ISP or another ISP are receiving this. There are two effects here: 1. Increased traffic for the "rogue ISP", 2. Peer pressure from other ISPs because other ISPs are affected by the rogue ISPs botnet traffic. Other ISPs could blacklist a rogue ISP; but this is a remote option which is not often exercised.

More infected users for an ISP increases traffic which can be an issue. Some types of malware will increase the traffic: if an ISP has many infected customers, their machines will generate more traffic. More traffic requires more capacity for an ISP. It will therefore increase the cost for an ISP incrementally. Although an ISP with enough capacity will not be influenced by this (van Eeten & Bauer, 2008, pp. 22-23).

Peer pressure can result because one ISP is very infected and other ISPs are affected by this. The more abuse there is on one ISPs network, the more other ISPs will ask for intervention. In an extreme scenario, the other ISPs can block such a rogue ISP. The rogue ISP will be unable to reach content hosted by the ISPs that blocked them.

Although many ISPs have agreed to participate in the Anti-Botnet Working Group, and definitely a must be aware of the problem of botnets (van Eeten, Asghari, Bauer, & Tabatabaie, 2011), they can generally be seen as negative towards external influence in the botnet mitigation problem. ISPs claim that they can solve the problem on their own, by self-regulation. In practice there are large differences between ISPs. Some ISPs put in very little effort and it appears that they are trying to maintain their independence/status quo (van Eeten & Bauer, 2008, p. 26). They have found indications that ISPs are only dealing with a small fraction of the infected machines in their network: an ISP with over 4 million customers, only contacted around 1000 per month while estimates are that they might have up to 200,000 infected machines in their network.

In interviews with ISPs, van Eeten & Bauer (2008, p. 29), identified an issue where ISPs are prone to: costs for brand damage. These ISPs operate in a competitive market, performing worse than competition is something they do not want (known). This means that being publicly mentioned as a bad ISP, be it in terms of botnet mitigation, is something many ISPs do not want. They do not want such knowledge to be in the media because they fear for their reputation. A bad reputation results in brand damage. Brand damage could potentially lead to higher costs and lower revenues. This suggests that publicly mentioning a ranking gives incentives to mitigate the botnet problem and contact infected users.

Publicly mentioning rankings, thus mentioning ISPs which are performing worse than other ISPs would result in intangible costs of reputation damage to such an ISP. Higher costs for underperformance could incentivize the ISP to invest in botnet mitigation measures, or to increase the botnet mitigation measures. On the other hand, the ISPs which are doing well would be able to show how well they are doing. For them it would boost reputation.

There is another way a reputation score based on botnet activity could help ISPs. If such a score is published well, it helps to create awareness at the end user. Many end users are unaware of the problem (app. A). The ISP which contacts an infected customer could be helped if the awareness of the customer increases. The helpdesk (responsible for contacting the infected customers) would need less time to convince a customer that he or she is actually infected, resulting in lower contacting costs for ISPs. Lower contacting costs reduces one of the issues for ISPs, botnet mitigation costs. A reputation system can help with this as it increases the information to end users.

The net effect of these incentives are that ISPs are mitigating the botnet issue, but only to a fraction of the amount of infected users they could mitigate. Since ISPs are mitigating only a fraction, the mitigation efforts are not up to the right level yet. The threat of brand damage and reputation effects should realign incentives for ISPs, so the amount of effort they would put in increases.

A reputation system can help to realign incentives for ISPs, reduce information asymmetries for the end users. By reducing the information asymmetry to the end user, the ISPs which are actively contacting customers are helped as the reputation increases can help to increase the awareness to end users. The next section (2.3) gives an overview over what kind of efforts there are already out there, in the field of cyber security, to measure or rank entities based on botnet activity.

2.3 Metrics and performance indicators in cyber security

This section covers the topic of reputation systems in cyber security. There are already some initiatives that grade i.e. companies according to their amount of spam traffic. Section 2.3.1 covers some botnet metrics from the Economics of information security group at TBM. Also other initiatives are described, section 2.3.2 covers an initiative Spamrankings.net, ranking companies based on spamtraps. Section 2.3.3 shows the senderbase from Cisco, showing global trends. Another ranking initiative is shown in section 2.3.4. Finally a proposal of the US government is described in section 2.3.5.

2.3.1 Botnet metrics research at TBM/TU Delft

In van Eeten et al. (2010) research was conducted into the link between ISPs, botnets and botnet mitigation. They have found empirical evidence that ISPs are indeed critical control points for botnets. One of the key differences between this research and i.e. Spamrankings (see next section) is that this research mapped the different autonomous systems to ISPs, where other research is only reporting spam/ per autonomous system (they do not know which AS belongs to which IP). Another main deliverable of the research is the botnet metrics. The

sections below give a brief overview over the used data sources and the identified/used metrics for the research.

Data sources

Spam traps: internet domains are set up with the goal of capturing as much spam as possible. The email accounts have not been used for something else, so all incoming email can be seen as spam.

DShield: a network of sensors as firewalls, IDS and home devices that log unwanted network traffic to the DShield database. The database gives a daily list of offending IP addresses, hosts, ports and attempted targets. Such an IP address points to an infected machine.

Sinkhole data: data from the Conficker sinkhole is used. A sinkhole tries to intercept the connection between machines infected with the Conficker bot and the C&C server. The benefit is that the data is free from false positives, since it only contains data from all the machines infected with the specific bot from that botnet, in this case the Conficker bot. However, this data is limited because it only represents machines infected with a Conficker bot. Section 2.2.3 gives more information about sinkholes in general.

An IP listed in one of these datasets is an indicator of being in a botnet. A machine does not have to appear in multiple datasets to be in a botnet; the correlation between IPs in different datasets is even very low (smaller than 10%).

Existing Metrics

The research conducted by van Eeten et al (2010), tested the relationship between several independent variables as: institutional incentives, organizational incentives, user behavior, technology and the national context and the dependent variable the botnet activity. Independent and dependent variables could be used as metrics either to describe an ISP or to describe the amount of infected machines and volumes of infection.

Several metrics have already been identified in (van Eeten, Asghari, Bauer, & Tabatabaie, 2011), (van Eeten M. , Bauer, Asghari, Tabataie, & Rand, 2010):

- Number of infected machines per subscriber (IP addresses per quarter)
- Number of infected machines per subscriber (Daily averages over each)
- Spam messages per subscriber
- Unique sources per subscriber
- Total number of unique sources
- Total spam volume
- ISP performance bandwidth
- ISP Size / Market share of an ISP
- Cable vs DSL

The research by van Eeten et al. (2010) confirms ISPs are critical control point for botnet mitigation. Just 50 ISPs worldwide consistently account for half of the infected (Spam) sources. This gives evidence that ISPs could be a better control point, instead of the end user, whom is

often unaware of the problem of cybersecurity. Institutional settings and incentives drive the behaviour of ISPs. Changing the incentives for ISPs, thus creating incentives for botnet mitigation, or disincentives for not mitigating should change the performance of ISPs in terms of botnet mitigation.

2.3.2 Spamranking.net

Spamrankings.com is an initiative by Quarterman et al (2013), it covers a method that measures spam traffic per company and assigns a reputation accordingly (Quarterman, Linden, Tang, Lee, & Whinston, 2013).

Quarterman et al (2013) have designed a randomized controlled trials to test the reputational effects of spam and botnet rankings as proxies for internet security on a firm level. They provide top 10 rankings per country by spam volume. They use spam traffic as the indicator of botnet activity, with the goal of testing whether actively disclosing information on the quantity of outbound spam emitted by an individual company will cause such a company to improve their information security (Quarterman, Linden, Tang, Lee, & Whinston, 2013, p. 2).

The study uses panel data from companies in 8 countries. Divided in to pairs of similar population and spam volume. The randomized controlled trials, mean that 4 of these are in the control group, and 4 are in the test group and are thus "treated", to test the effect of publishing information about their spam performance.

In developing a reputation system based on spam, some issues have been found (Quarterman, Linden, Tang, Lee, & Whinston, 2013, pp. 3-4), these include the Geographical heterogeneity, organizational classification, extraneous events such as takedowns and data incompleteness, patching spam. Each of these is explained below and their consequences for this project are explained accordingly:

- **Geographical heterogeneity:** in different countries, legal regimes have different characteristics, this could potentially skew results. Botnets and Spam are usually do not operate one country only, but crosses the territorial borders and are active in multiple countries. Different regimes indeed is an issue, however this study only focusses on Dutch ISPs (since ISPs usually are per country since they operate per region)
- **Organizational Characterization:** Quarterman et al (2013) are measuring traffic per company. Measuring spam traffic per company is more difficult than spam traffic per ISP. First of all companies (can) have multiple IPs. Second companies often consist of many companies and branches with other legal names. ISPs however assign IPs from their IP ranges, this means that the issue of organizational characterization is not an issue here.
- **Patching spam** (turning off the possibility to send spam, without removing the agent sending spam): When dealing with spam, be it a regular company or an ISP, there are (at least) two options of dealing with it. Either a company can fix this problem by addressing the underlying problem (as an installed bot on a machine) and thus fixing the cause of spam (thus treating the disease), or the effects of spam can be mitigated (thus treating the symptoms not the disease).

- **Extraneous events as takedowns:** Sometimes there are proactive takedowns of a botnet. In such a case i.e. a governmental organization can take town the bot herder: the operator of a botnet. If the herder stops using a botnet, the total botnet activity is at that moment lower. Given the strong correlation between botnet and spam, the spam traffic also decreases. However this effect is usually temporary since another botnet will just take over (Quarterman & Whinston, 2010). A botnet might be taken down, but the bot is still installed on some machine, thus the security flaw is still an issue.
- **Data limitations:** Not all spam traffic can be measured and it can only be estimated what amount of spam traffic is measured. It is unknown what the total amount of spam traffic is. Therefore data are indications which could paint the wrong picture occasionally. This is a limitation of the research, since it is impossible to determine if all spam traffic is measured. However, given a certain amount of data it could be stated that the measured data is representative for the population meaning that the measured spam traffic is a good representation of all spam traffic.

In their article, Quarterman et al (2013) are not completely transparent in the actual methods they use to calculate a reputation, however some metrics and calculations are given in their article and on their website. Their main metrics are spam volume and unique spam sources or spam count.

In calculating a reputation for a company these authors look at the following indicators:

- **Total spam volume:** if it goes down after a ranking of a company, it suggests that they have done something about the problem
- **Duration of infection:** If the time of an infection is shorter, suggests that malware is ejected quicker
- **Frequency of infection:** If an organization is infected less frequently, it could be that they have done something about the problem

The data used is obtained from various sources, they do not measure spam themselves. The different data sources and indicators are put together into one metric (the reputation) using a Borda Count (Spamrankings.net, 2014). The Borda count gives a weighted composite ranking based on spam volume and spamming address count. The spam volume and count is obtained from two blacklists (CPL and PSBL). Page 8 of Quarterman et al (2013) and the frequently asked question page from spamrankings.net give some more insight into the actual weighing in the Borda count.

In the Borda score, they basically sum the ranking for each metric for each dataset. A high score indicates sending out more spam.

2.3.3 Cisco SenderBase

Cisco provides a website that shows data about spam. It does not really give a rating to it; it measures the total amount. A user of Senderbase.org can distinguish between top senders by IP, Network owner or by country. Figure 1 shows an example of senderbase's interface. It shows daily spam data per country (SenderBase.org, 2014).

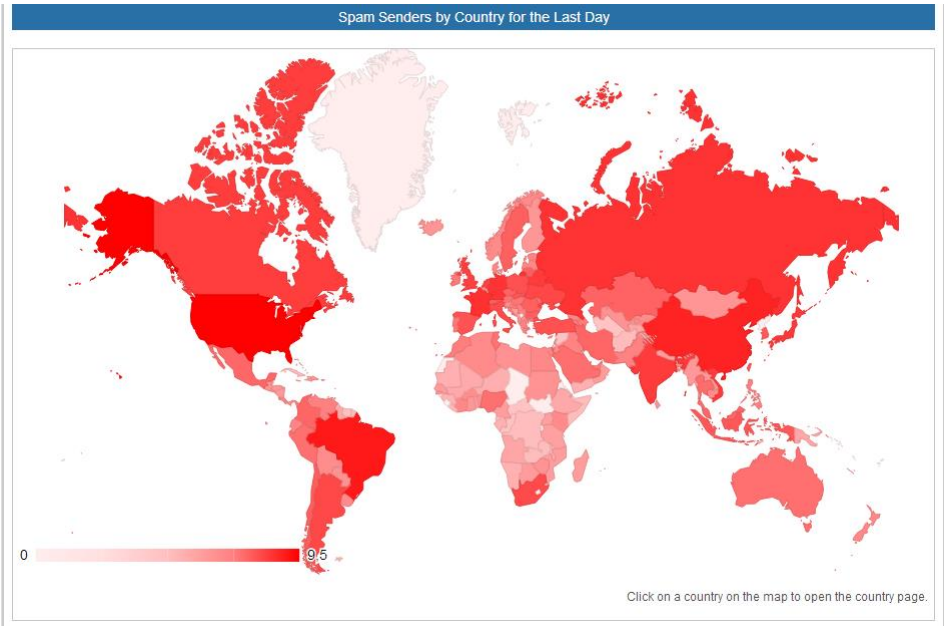

Figure 5 Senderbase daily spam data per country adapted from (SenderBase.org, 2014)

Such an initiative as Cisco SenderBase shows thus performances based on summations of spam volumes.

2.3.4 BGP ranking and ASMATRA

Wagner et al (2013) have identified a method to rank Autonomous Systems. Using the BGP lists there is a method to rank ASes based on the malware they host. An AS (called an AS) is ranked on a scale from 1 to unlimited, with one being the top score/default score. Scoring is based on a set of lists, called blocklists (BL). These blocklists hold (publicly known) malicious IP addresses (so IP addresses that transmit malware) (Wagner, Francois, Dulaunoy, Engel, & Massen, 2013, p. 262). Examples of blocklists can be found at www.blocklist.de, or www.dshield.org. For every day a list is collected, since the list is variable (and incomplete), a factor ”b” for impact is added.

For every occurrence of a malicious IP within the range of an ASx, a counter sums one extra. For all different blocklists (BL) the sum of all the occurrences (OCC) is computed (with the impact factor b) the sum is divided by the size of the autonomous system (ASx size). In formula (Eq1) this is (Wagner, Francois, Dulaunoy, Engel, & Massen, 2013, pp. 1, formula 1):

$$ASrank(ASx) = 1 + \frac{(\sum_{b \in BL} occ(b, ASx) * impact)}{ASx\ size} \quad \text{(Eq 1)}$$

Wagner et al (2013) concludes that this method does provide a method to rank ISPs, but there are 3 types of AS and thus ISPs. One of these is a transit AS, which only provides a routing service from another AS. Imagine a bad AS, with a high ASrank, such an AS could route via another transit AS. Wagner et al (2013) conclude that in such a case the ASrank for the transit

AS would not be so high. Although (in an extreme case) the bad AS would be disconnected from other ASes, the transit AS would still ensure the connectivity for this bad AS.

2.3.5 CSRIC III working group botnet metrics

For the US FCC a report on botnet metrics has been drafted in 2012 (St. Sauver, 2012). It looks at botnet performances also per ISP for the US. Although there is not reputation system developed by them, they do give some guidelines in what kind of factors should be measured. The report is a bit coloured, because it is written in a negative way. The authors do not like the idea of a reputation system for ISPs, but had to evaluate the option. They are negative as their incentives are related to those of the US ISPs, because they are also in the council (FCC, 2013). However the concepts to be used could be selected for this research.

The basic argument of the report why rating ISPs for botnet activity will not work is because researchers will not be able to have all the information. Not all botnet infections can be measured. There will be cases where people are indicated to be infected, but in practice they are not infected, so called false positives. Where will also be infected machines that are unknown; so called false negatives. Researchers are not able to manually check every machine, so not every infection can be measured.

In the report some factors to count have been identified, these are (St. Sauver, 2012, p. 9):

- Individual infections
- Botted systems
- Botted IP addresses
- Botted subscribers
- Type of infected device (PC, Laptop, Tablet, phone)

The above factors are all in a way counting machines, or the amount of Spam/botnet activity per user, but these factors alone do not have much value because there is no reference into what kind of timeframe it is measured and also the impact is not mentioned. This means that the when and how long (the measurement window) is also important (St. Sauver, 2012, p. 10). It is important that the timeframe of measurement includes the time that the infected machines are on. For example, if botnet measurements are only running during the day and if an infected machine is turned off during the day, the measurement will not give an infection although the actual machine is infected. On the other hand, continuous monitoring would result in seeing the same botted host more than once, so duplications in data. A botted host is another word for a machine that is infected with a bot. The underlying issue here is the activity of the botnet that infected the machine, how active is it, what is its impact, when is it active. There is a difference in a botnet that only sends out spam or a bot that is attacking the financial system. The question of monitoring does also show that a botnet does not have to be active the whole day, week, month or year, so it might not be active at the moment of measurement.

Another factor to hold into account is the potency of the botted hosts. The report compares two different infected machines. The first one is an ancient consumer system connected via a very slow internet line and the other one is a high end server, with high capacity and a fast internet

line. If both the old system with the slow internet line and the high performance system with fast internet line are infected with the same bot, the results could differ. The bot could use more internet speed, and computer power with the second machine, and could potentially do more harm. The differences between computer capacity and internet speed should therefore (preferably) be taken into account. This could be measured as i.e. an average spam throughput or as average DDos output (DDOS = a distributed denial of service attack, usually an attempt to make a service unreachable or forcing a service to go offline).

2.4 Conclusions

This chapter discussed three topics. Botnets (section 2.1), Internet service providers (section 2.2) and performance metrics for Botnets and ISPs (section 2.3). In this section some references to **(Req number)** are made. They will be used in a later stage to form requirements, the reference in the conclusion is purely for traceability purposes (see footnote 1).

Botnets cause our society large problems. Many machines are infected with a bot, a piece of malware which allows a botherder to control the infected machine. The end user becomes infected with a bot, but this end user is not the only one. A botnet is a herd of bot infected machines up to many thousands.

Less infected machines is preferred over more, therefore the amount of infected machines should be as low as possible. Botnet mitigation is therefore very important because botnets and spam are a serious problem in society (par 2.1). For example the cost of spam is very high, especially compared to the revenues spammers make. Spam usually originates from botnets. It is estimated that about 10% of all machines are infected with a bot (see par 2.1). Botnets can be used for a variety of purposes ranging from sending out spam, to potentially be used for terrorism or espionage. The threat of botnets is very serious and increasing. A bot on an infected machine has a purpose: for example sending out spam. Because a machine is sending spam, the IP address of this machine can be traced.

Botnets can either be taken down (i.e. by a police force), or botnet infections can be mitigated at the end user. When the end user becomes infected with this bot, it can be mitigated by increasing security. The machines which are infected with bots often have bad or outdated security (definitions). They are under secured. Often the end user is unaware of such problems, or does not have the funds, resources or knowledge to mitigate the problem. There are large differences in information which people know. The end user is often uninformed, resulting in the end user not being able to determine if they are infected and how to mitigate this infection **(Req 1[1], see footnote, to be used in section 5.1 later on)**.

Since the end users do not have the knowledge and lack information to determine if they are infected, other stakeholders can help these end users, for example by informing them or helping them to mitigate the infection.

[1] The "**Req 1"** refers to a specific requirement. Such requirements will be used in section 5.1 do determine what a design should fulfill to. It is not hierarchical so the number does not refer to any importance of the requirement compared to other requirements. The reference in this text is to provide traceability of later requirements.

The Internet Service Provider is identified to be the stakeholder to help the end users. The ISP is the link between the end user and the internet. The end user receives their IP address from the ISPs. Botnet infections can be measured on IP basis. Since ISPs are the only one who can know which IP belong to which customer, they can get into contact with the customers. For this reason the market for ISP was researched.

The ISP is the bridge between an end user and the internet. Other stakeholders (i.e. a government) do not have this information, they do not operate in the network of the ISP. There are many different ISPs in the Netherlands and it is a competitive market. Some have small networks, while others have a large market share **(Req 24)**. Section 2.2 identifies ISPs as the party to mitigate botnet problems.

ISPs are aware of the problem of botnets, and claim to be working on it, i.e. many Dutch ISPs together form an anti-botnet working group. They could mitigate the botnet problem, by contacting infected users. However, for many ISPs the number of customers they contact, compared to the number of customers they could be contacting because of infections, is very low (see par 2.2.3). By law ISPs should put in an effort for mitigation, but also maintain customer privacy **(Req 18)**.

Apparently the effort (some) ISPs put in is not enough yet. This is because the incentives for ISPs are not aligned enough to increase the botnet mitigation efforts. ISPs have high costs when contacting customers with infected machines, giving an incentive to just ignore threats of botnets. The consequences for not intervening are also very low; for other ISPs it is also hard to determining how well their competitors are performing in mitigating botnet activity. There should thus be an incentive for these ISPs to start mitigating botnet activity. Assigning a reputation to these ISPs from their botnet activity should give a strong incentive for botnet mitigation by ISPs.

ISPs operate in a competitive market. A bad reputation could yield brand damage for an ISP. The fear of brand damage, and the costs for brand damage incentives them to mitigate the botnet problem at their customers. If all ISPs do this, the botnet problem decreases. For this reason a reputation system for ISPs should be constructed. Assigning a reputation to ISP on how well they perform in terms of botnet mitigation could realign the incentives for ISPs to increase their mitigation efforts. The reputation system could therefore have an effect on botnet mitigation **(Req 2)** by realigning the incentives for ISPs to increase the mitigation efforts **(Req 3)**.

Reputation helps to increase information, as reputation can be seen as an information signal, decreasing an information asymmetry. The reputation system can help the ISP also. If such a score is published well, it helps to create awareness at the end user. Many end users are unaware of the problem (app. A), a well published score and explanation might help mitigate this. The ISP which contacts an infected customer could be helped if the awareness of the customer increases. The helpdesk would need less time to convince a customer that he or she is actually infected, resulting in lower contacting costs for ISPs. Lower contacting costs reduces one of the issues for ISPs, botnet mitigation costs.

A reputation system can align the incentives for ISPs in such a way that the botnet mitigation efforts would increase. For this reason, in section 2.3 research was conducted to determine if

there are already reputation systems in cyber security. There are already some methods which measure the botnet activities and in some cases already give a grade or ranking to the performance of a company based on spam output. However, there is a limitation to these initiatives. First the initiatives that result in a grade or ranking (i.e. Spamrankings or BGP ranking), focus mainly on spam and do not provide a comprehensive reputation metric. The second one is that some of these initiatives only provide metrics, so to use them as a method to convey a message to ISPs, governments or end users, they have to be transformed to a reputation (system).

Section 2.3 provides a set of possibilities and shows what kind of information about botnets can be measured, it provides metrics about botnets. These metrics show what can be measured about botnet activities **(Req 20)**. The metrics measure activity and volumes for different botnets and spam. Botnet traffic is volatile **(Req 15)**, therefore these metrics are changing from day to day. These metrics are based on automatically measured data **(Req 19)** from sinkholes, honeypots, blocklists and Dshield data. Some of these data sources are susceptible to false positives. With such a false positive the data suggests an infection, but actually there is no infection. This is an issue with measuring botnet traffic **(Req 21)**. Botnets can be taken down **(Req 22)**, in such a case a police force removes the botnet altogether. Many metrics would show lower values with a takedown. From such initiatives trends about botnets become known. Usually botnets follow a trend, i.e. the trend for spam is that it is decreasing **(Req 23)**. The trend suggests that old botnets become less active, and new ones arise **(Req 25)**.

Research gap

The situation is thus that ISPs can mitigate the problem, but have the wrong incentives to undertake enough mitigation efforts to mitigate botnets. Increasing ISPs botnet mitigation efforts requires to publish a reputation based on the infected machines they host. There are already initiatives to rank systems to the level of spam they host, however these initiatives are too limited still. They only provide a ranking based on spam, while spam is an indicator of botnets, but certainly not the only one. The initiatives do not specifically focus on ISPs, they either rank autonomous systems or they rank companies.

The initiatives are not sufficient yet, but do give good metrics to indicate botnet activity. There is not a reputation system in practice which incentivizes ISPs to increase their mitigation efforts. A reputation system based on the amount of infected machines an ISP hosts could be designed to incentivize botnet mitigation by ISPs. Therefore, in the next chapter the concepts of reputation and reputation systems are explored. Exploring reputation and reputation system should give a set of knowledge about how such systems work and what they are based on.

3. REPUTATION AND REPUTATION SYSTEMS

In the previous chapter it has been identified that botnets are a serious problem for society and Internet Service Providers are the stakeholder best suited to mitigate botnets. These ISPs do not have the proper incentives to undertake enough effort to mitigate botnets. Reputation, based on the number of infected machines they host, may realign this incentive. At this moment there are no reputation systems to show a reputation for ISPs to incentivize botnet mitigation. The literature about reputation and reputation systems in cyber security is incomplete, as no reputation systems for ISPs exist yet. This chapter tries to identify the general concepts of reputation and reputation systems based on literature. In a later stage this knowledge can be used to define new literature about reputation systems for cybersecurity (specifically for ISPs).

In section 3.1 the concept of reputation is described. Questions are: What is reputation, what does it consist of, how does one get a reputation, are explained. There is a strong relation between reputation and reputation systems. This is the method in which a reputation is calculated and reported. The dimensions for reputation and reputation systems are therefore not unrelated. In literature these concepts are used ambiguously, some articles use concepts in describing reputation (e.g. in (Alperovich, Judge, & Krasser, 2007)), while other articles use them in describing reputation systems (e.g. in (Josang, Ismail, & Boyd, 2007)).

Section 3.2 describes the dimensions important to reputation systems. The section describes what a reputation system is, how its quality can be assessed and what kind of dimensions together form a reputation system. The process from data to a reputation, is shown in section 3.3. Section 3.4 concludes this chapter. The information from sections 3.1 to 3.3. together should help to gain a set of knowledge about reputation and reputation systems. This set of knowledge is identified to be lacking in chapter 2. In a design this knowledge should help to determine what kind of reputation system should be designed for ISPs.

3.1 Concept of Reputation

The concept of reputation is difficult. Reputation is a something many talk about, without knowing what it exactly is. People often say that someone, or something has a good or bad reputation, but what does this mean and how did they derive this conclusion?

Determining what reputation exactly is, is difficult because reputation is not an exact concept. Reputation is ambiguous and intangible (Fombrun, 1996, pp. 11-12), this makes reputation hard to define and to measure on one scale. Although reputation is hard to define, it is important to develop a good reputation (Fombrun & Shanley, 1990). Many people form opinions about something because of its reputation (Saxton, 1998, p. 397). This suggests that companies rely on their reputation to compete (Fombrun, 1996).

For example: between the management of a company and a company's stakeholders, there are often differences in the information both have (Fombrun & Shanley, 1990, p. 235); management knows about a company's private information, while stakeholders only know the public information. Public information is much less. Stakeholders have therefore different and less information than management about the company. This information difference often drives

stakeholders to search for more information about the company, a reputation can decrease the difference in information for stakeholders. A reputation can thus be a positive or negative information signal.

Firms with good reputations, will protect their reputations. This limits managers in the firm from doing activities that can be seen as unacceptable, since it might damage their reputation (Fombrun, 1996). A firms' reputation is often publicized, i.e. by Forbes. These listings can influence the decisions management have to make, for example by determining which threats and which opportunities should be investigated.

A reputation is not straightforward, but it is an ambiguous concept. It focusses on past actions and the expected future actions to form a reputation. Many know about it without knowing a definition.

In many cases (e.g. Fombrun (1996)) reputation is based on perceptions. This is already very specific the context of perception based reputation; thus not usable in this research. Literature has defined dimensions into which reputation can be separated. Reputation is dependent on context (Sabatier & Sierra, 2005), the purpose (Josang, Ismail, & Boyd, 2007). Hoffman et al (2009) defined context and purpose further as:

"A reputation is the degree to which one party has confidence in another within the context of a given purpose (Hoffman, Zage, & Nita-Rotaru, 2009, p. 3)*"*

A reputation can generally be classified onto two axis: (Josang, Ismail, & Boyd, 2007)

1. From general to specific
2. Based on human feedback or machine feedback.

General and specific are derived from the context of the reputation. The context shows the scene were the reputation is for, while the purpose indicates what the reputation is used for. They are fairly similar and are described together in section 3.1.1.

The confidence in the reputation can be based on two information types (Section 3.1.2). The information type contains the information used to derive a reputation. Such information originates either from human feedback (section 3.1.2.1), or from machine feedback (section 3.1.2.2). The type of feedback influences how subjective the information is. It is important to note that a reputation does not have to be based on one type of feedback only, it is possible to use both (i.e. see section 4.3). After discussing reputation the topic shifts from reputation to reputation systems in section 3.2.

3.1.1 Context of a given purpose

The purpose of the reputation is very important, because it drives the question about the level of specificity (Josang, Ismail, & Boyd, 2007). A reputation to be used for many different contexts has to be more general, where a reputation used for one scenario can be more specific (Josang, Ismail, & Boyd, 2007, p. 628).

Therefore a reputation is dependent on the context (Sabatier & Sierra, 2005, p. 39). Take as an example a doctor describing a medicine. If a doctor recommends a medicine, usually people trust this doctor to be right, because he has studied for it. However, if this doctor suggests a bottle of wine to be very good, he or she doesn't have to be right (Sabatier & Sierra, 2005). Being a good doctor does not make someone a good wine connoisseur. Having a good reputation for something, does not mean that it applies to everything. This suggests that factors as reputation are dependent on the context and every reputation has an intended purpose and is associated with a specific context.

However it is also possible to have a reputation which is multi context, so usable for more than one situation. Such a reputation increases the complexity, since data about different contexts have to be merged to one reputation. In a multi- context situation therefore a reputation would be more general than in a single context situation. In practice not many reputation models are multi context, because the models focus on a specific scenario. The context is therefore an important dimension of reputation. A reputation can only be valid in the right context. Also the contexts in which it operates eventually will influence the complexity of the way the reputation is derived. The range for the context is from one context to multiple contexts.

3.1.2 Information source and extraction

A reputation can result from both direct and indirect information (Mui, Halberstadt, & Mohtashemi, 2002). This means that a reputation can be based on direct encounters (first-hand information), or indirect encounters. With indirect reputation, the information is gathered indirectly (i.e. by word-of-mouth), where with direct reputation data is gathered based on direct information or observations (so data is measured).

Sabatier & Sierra (2005, p.36) identified the direct experiences to be the most relevant and reliable information sources. Indirect information, also called witness information, is however more abundantly available. The problem with this information is that it is subjective. This indirect information therefore increases the complexity for reputation models; a "witness can manipulate or change information for their own good.

The information sources determine how subjective the reputation is. Is a reputation based on Machine feedback, or is it based on Human feedback? Often human feedback is more subjective as it contains human perceptions about a concept, were machine feedback is less subjective as it measures the concept itself. The machine feedback can become subjective as somebody has to determine what kind of feedback should be gathered by the machine. Sections 3.1.2.1 & 3.1.2.2 discuss human feedback and machine feedback further.

With reputation the rule of thumb usually is: the more different data sources, the more reliable the information is. To generate a reputation, information is required. A reputation can rely on two types of information: Human based feedback and/or machine based feedback (Alperovich, Judge, & Krasser, 2007, p. 11). These two types are described below.

3.1.2.1 Human feedback based reputation

Information gathering can be based on human or user experiences, in such a case a reputation system is based on human feedback. Such reputation systems are often seen in fields of

corporate reputation/ marketing. Institutes as Forbes rate companies according to several dimensions (see section 4.2).

These companies often rely on opinions because information is largely incomplete or ambiguous. This means that the firms' activities and known information over time drives people's judgments about a company. Firms activities as diversification, profit ratings, risks, advertising and social responsiveness drive market risks and performance, media exposure, dividends and institutional ownership (Saxton, 1998). These factors are for the "outside" stakeholders' information signals, allowing for an assessment of the firm's reputation.

In human based feedback, indirect experiences as word to mouth can be influential. People form opinions based on these second hand experiences. Not only the field of corporate reputation uses human feedback; the field of computer science also uses this concept. In this field messages are often examined by hand (instead automated examination). Examples are users submitting spam reports or voting systems (Alperovich, Judge, & Krasser, 2007, p. 11).

Human feedback thus measures the perception about something or someone. Instead of measuring the actual performance of i.e. a company, the perception of people about a performance is measured.

3.1.2.2 Machine *based feedback reputation*

Section 3.3.1 described reputation based on human feedback. In such cases reputation consists of the perceptions about someone or something. A human feedback based reputation system is subject to change if perceptions change. In practice this means that i.e. a company could perform very well, but still have a bad reputation because of perceptions about this company.

Another form of reputation is machine learned/based reputation. In such a case reputation consists of machine feedback (data that is measured or generated by machines). The difference between human feedback based reputation and machine feedback therefore is data extracted through automated means. There is an underlying mechanism that automatically extracts data and generates a reputation from this data (Alperovich, Judge, & Krasser, 2007, p. 11).

Reputation based on data extracted through these automated means comes often from the field of Computer Science (CS). In this field reputation is synonymous or at least very close related to trust (Mui, Mohtashemi, & Halberstadt, 2002).

Examples of machine based feedback are financial data or in the case of botnet mitigation data from spam traps or honeypots. There is a fundamental difference between machine based feedback and human based feedback (Alperovich, Judge, & Krasser, 2007). Human feedback contains opinions, were machine feedback contains measured data about something. The human feedback gives an opinion of someone about something, were machine data describes how something is performing; the object to be measured differs. Human feedback definitely contains perceptions and is subjective, but machine feedback also isn't objective. Machine based feedback can still be subjective, this is because the selection of information is still done by human intervention. In other words, somebody has to determine which data the machine should select.

The relationship between the information type and information extraction are linked. A human feedback based reputation system, will often rely on indirect information, because it measures perceptions. In a machine feedback based reputation, the data would be direct. Such (machine measured) data is more objective since it does not rely on opinions.

There are exemptions to be made here. A human feedback based reputation is susceptible to lying and untrustworthy behavior by the agents, for example if the agent has something to gain by misrepresenting the facts. A machine feedback based reputation can also be susceptible to the same problem. Instead of giving the wrong perceptions, an agent could still manipulate the data, i.e. by adjusting machine settings in a way that it appears as if the performance is very good/bad.

3.1.3 Dimensions of reputation

In the previous sections several characteristics of reputation have been described. The context, information types and information extraction influence the way a reputation is constructed. The context influences the complexity of a reputation. A reputation that is usable in multi context would be more general, were a one context reputation would me more specific. The other way around there are also differences. The information type is determined if information is based on human or machine feedback. Is it direct information (i.e. measured by a machine), or is it indirect/witness information? Indirect information is often easier to obtain (i.e. by expert findings or questionnaires or reports), but such data can easily be manipulated. Figure 6 shows the dimensions of a reputation.

Reputations dimensions

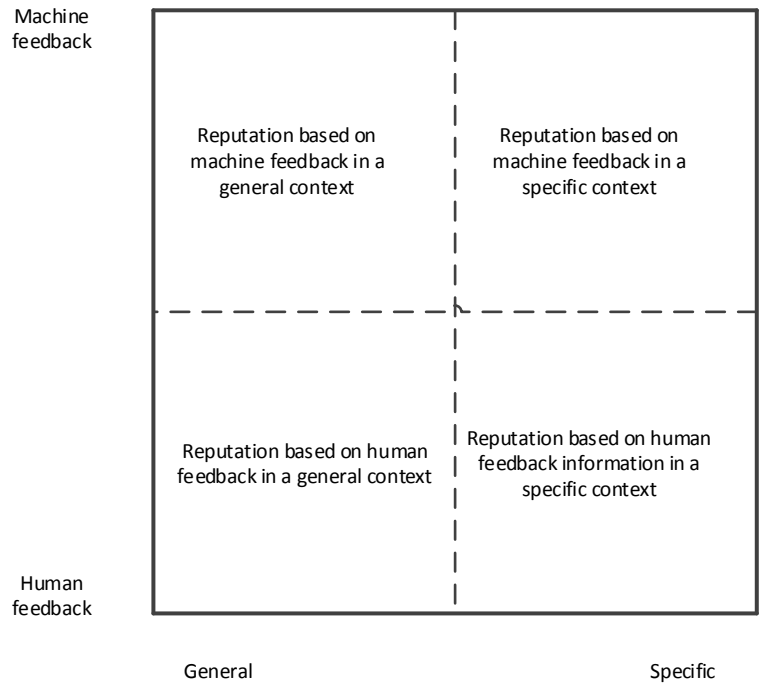

Figure 6 Reputation dimensions

Machine generated feedback, and direct information are more reliable since they are more objective. Such data is more objective as it the machine does not lie. Only the interaction between man and machine can affect its objectivity. Someone has to select or interpret information, this means that although the extraction of data can be machine based, the interpreting of data is still subjective. Human generated feedback, and indirect information is less reliable, because it is definitely more subjective, the focus shifts from hard measured data, to ambiguous data, were instead of concepts the perceptions about the context are measured. In the case of indirect data, there is an upside, it reduces the complexity of deriving a reputation. The relationship between information type and information extraction is very direct, therefore it is not necessary to include it as a dimension.

In the next section the focus shifts from reputation to reputation systems. Such a system calculates and reports the reputation. It is the method and setting into which a reputation is placed.

3.2 Reputation systems

Many fields study or use reputation systems; examples are the fields of economics & finance, business and the field of computer science. In these fields the concept of a reputation system is used in many different ways. Economists study reputation (systems) in game theoretic settings, i.e. with a game as the Prisoners dilemma (Mui, Halberstadt, & Mohtashemi, 2002, p. 281). The field of finance uses ratings to determine financial risks, examples here are the financial product rating systems as S&P uses.

Business reputation systems try to determine a corporate reputation based on the perceptions of people about a company and the company's performance (it thus includes perceptions and machine feedback data as financial fitness).

Finally in the field of computer science, there are many different types of reputation. The most well-known examples are rating system as used by eBay, but also in the field of cyber security there are already initiatives which report reputations or performances of entities. This chapter describes different aspects of reputation systems.

3.2.1. Concept of a reputation system

This section covers the concept of reputation systems. It is covered into two sections. Definition (3.2.1.1), here reputation systems are defined and the differences between a reputation systems and performance indicators are described. Objectives (3.2.1.2), here the objectives for a good reputation system are covered.

3.2.1.1 Definition

A reputation system, or reputation system, is an automated method that collects, distributes, and aggregates feedback about a participants' past behavior (Resnick, Zeckhauser, Friedman, & Kuwabara, 2000, p. 2). In other words, it is the underlying mechanism that determines the reputation from the collected feedback. A reputation is distributed into some kind of "grade" or ranking.

Resnick et al. (2000, p.7) described internet reputation systems to have the following properties:

- Entities are long lived
- feedback about current interactions is captured and distributed, and
- past feedback guides future decisions.

Herein lies the difference between a reputation system and a performance indicator or blacklist. A reputation system has also a predictive factor. Predictive feedback is used to classify a group of identifier prior to observing the behavior of these identifiers (since the behavior still has to occur, because it is in the future) (Alperovich, Judge, & Krasser, 2007) For example, Spam filters often use blacklists. Some initiatives, see section 2.3, count the number of occurrences on a blacklist and average this. Such a method reports the current performance, but there is little to no predictive aspect. A reputation system is broader because it would include other identifiers to also contain a predictive aspect. Different performance indicators together are part of a reputation. The relationship between performance and reputation as follows: a performance indicator can be an element of a reputation (system), or a reputation (system) can, partially, consist of performance indicators with a predictive element.

3.2.1.2 Objectives for assessing a reputation systems quality

Dingledine et al. (2000) have set 4 objectives to assess the quality of a reputation system. These are: Accuracy, Weighting towards current behavior, robustness against attacks, and smoothness. The figure below shows these objectives, they are explained below. In this research they are interpreted as general criteria for quality assessment. In turn they do translate in a later stage to requirements. This is because i.e. the quality of a reputation system cannot be good as it is inaccurate.

Objectives for a reputation system

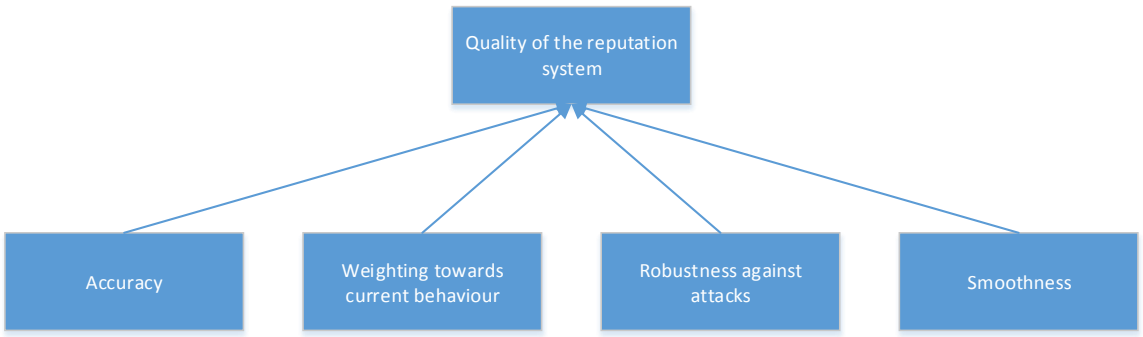

Figure 7 Objectives for a reputation system

Accuracy

Accuracy for long term performance is an important objective for a reputation system. It means that over a long time the reputation should represent the actual performance of the underlying entity, in other words the reputation should be correct. Also it must have the capability to distinguish between a new entity, thus with little data, and an entity which has a poor performance over time (Josang, Ismail, & Boyd, 2007, p. 640).

Weighting toward current behavior

An entity can be the best performer for the past years, but if it is recently performing badly, then a reputation system should also depict that. The system should recognize and represent the recent trends in performances (Dingledine, Freedman, & Molnar, 2000). For example, there are two entities: an entity with a bad reputation for the last years, but which is doing better at the moment and second an entity which a good performance, but which is recently doing very badly. It can be argued that the second one is doing worse than the first one.

Robustness against attacks

People will try to manipulate the system. Therefore the robustness against these manipulations or attacks should be important (Dingledine, Freedman, & Molnar, 2000). If a system is easily manipulated by the entities, it is worthless. It can be noted that the robustness against attacks also affects the accuracy, since a system with a low robustness, also has a low accuracy.

Smoothness

If a new observation is added to the data and the rating changes very much it becomes a very volatile system. An entity cannot have a good reputation and the next minute have a bad reputation. This means that a new observation in the data should not change the rating significantly (Josang, Ismail, & Boyd, 2007, p. 640). In the next section different computation engines are described.

These criteria can be seen as criteria for quality assessment. They will be used in later chapters to determine the quality of existing reputation systems. Although these criteria are for quality assessment, they are related to the dimensions in the next subsections. For example, the first criteria shows accuracy: if a reputation system uses a calculation method which is not accurate it affects the quality. Another example is the robustness against attacks. If a reputation system is susceptible to cheating, anti-measures have to be taken to be robust. These criteria above thus show how it can be assessed, while the dimensions below show how what elements of a reputation system there are.

3.2.2 Reputation system computation engines

Josang et al (2007, p.628) have defined different archetypes of reputation systems from literature. These are explained briefly described below. They differ mainly in the way a reputation is calculated (Josang, Ismail, & Boyd, 2007). In practice there will be many variations and expansions of these reputation engines, but often the core of such an engine has an origin that is derived from the engines below. Table 2 shows how the different computation engines score on the below described criteria.

The benefits and concerns for a computation engine are described in Josang et al (2007, p 628-629). The accuracy, understandability and required computational power are related. Josang et al (2007) give benefits and concerns for every computation engine based on these three factors (see table 2).

1. Accuracy: how well their computation scores correspond to actual performances.

2. Understandable: Is everyone able to understand the engine, or is it very complex
3. Computational power: The engine automatically calculates a score, a more complex algorithm would use more computational power.

Table 2 Comparison of computation engines

	SUMMATION	BAYESIAN	DISCRETE	BELIEF	FLOW
ACCURACY	-	+	+	++	0
UNDERSTANDABLE	++	-	+	--	-
COMPUTATIONAL	++	+	-	0	+

These factors are used because an ideal algorithm is accurate, understandable and is economical in the amount of computational power it uses. Accuracy is required since a score should correspond to actual performances, so an algorithm shouldn't give the wrong score. Understandable should be maintained so that everybody could understand the way the score is derived (Josang, Ismail, & Boyd, 2007, p. 628). The computational power required would influence the amount of reputations which can be calculated on a machine. An algorithm which uses more computational power, would decrease the amount of calculations a machine can handle. The next five subsections explain a type of computation engine, and how it scores on these three factors.

3.2.2.1 Summation/average

This is the simplest form of computing reputation scores. Positive and negative feedback are simply summed (separately) and used to form a total score. The average score than would be obtained by dividing the total score by the number of instances. A well-known use of this reputation type is used by eBay.

The advantage of such a type of reputation is that it is easily understood and explained. Its simplicity makes sure that it also requires little computational power. The disadvantage is that such a reputation system is very simple. Its simplicity can cause it not give the proper values to all the data, therefore it can be less accurate than a more complex algorithm.

3.2.2.2 Bayesian systems

A Bayesian system uses a statistical method to determine a reputation. It is scored by using the previous reputation score together with the new rating. The input of such a system is often a binary, a positive or negative value. The reputation score is computed by combining the previous reputation score with the new rating.

An advantage of a Bayesian system is that it gives a theoretical sound basis for computing reputation scores (in comparison to a summation score), it is much more accurate. Its disadvantage however is that is much more complex and therefore more difficult for the average person to understand, and requires more computational steps and thus also more computational power than summation.

3.2.2.3 Discrete model

In a discrete trust model measures are split into verbal segments as “good – neutral – bad”. This is because humans are better to rate a performance in these verbal statements, continuous values can also be discretized. In the case of human feedback in such a form there is a lot of subjectivity since the perception is actually measured instead of the underlying concept.

The advantage here is that is easily explained, it can be very accurate, the disadvantage is that lookup tables have to be used. In a lookup table values corresponding to the levels can be found, this gives some computational issues. Instead of calculations, values have to be found in the lookup tables. This requires more computational power.

3.2.2.4 Belief models

A belief model is also a model based on probability theory. In such a model, the sum of the probabilities does not have to add up to 1 (or 100%). The remaining probability is further modelled as uncertainty. A belief model is focused on measuring the amount of confidence there is in i.e. a statement from someone. It gives a value to the opinion of someone. The probabilities are similar to the Bayesian model constructed, with a beta probability density function.

It has an advantage, because it gives values to statements, so it can rate subjective feedback so the accuracy is increased. The downside of such a system is that it rates subjective values as people’s opinions, so it can be used on a limited set of problems. Also it is difficult to understand since it continues on the probability density functions and incorporates it in multiple nodes.

3.2.2.5 Flow models

A system that computes reputation by iteration through looped or long chains can be called a flow model. Such a system is often considered as a zero sum game. This means that one’s reputation can only become higher, at the cost of another’s reputation. The total value in the model is assumed to be constant. An example of a reputation model based on a flow model is googles ranking algorithm PageRank.

The flow model can be very accurate, but also very inaccurate. In such a system there can only be one “winner” and for one to win, others have to lose. For this reason it is moderately accurate. A vector bases system is also more difficult to understand, i.e. compared to a summation score. It is a process which can be computer relatively easy, but it requires to keep a score for every node in the system.

3.2.3 Reputation assigner

In a reputation system there is an entity which calculates the reputation. However, the location as to where a reputation is calculated, can be categorized into two types: Centralized or decentralized (Hoffman, Zage, & Nita-Rotaru, 2009, p. 8). A centralized system uses a centralized authority, this is often a straightforward solution and easier to implement with less possibilities for manipulation (Hoffman, Zage, & Nita-Rotaru, 2009). It is used in many e-commerce

platforms. In such a situation all data is gathered by or send to this authority, they store it and calculate a reputation.

Thus, a centralized system is easier to implement and less susceptible to manipulation, but there are also issues. Having a centralized authority, introduces a single point of failure. If the systems at the authority do not work, there is also no reputation system. A centralized system only works on the assumption that all participants to the system trust this authority.

Decentralized reputation means that entities themselves keep data, and work together to calculate a reputation. If one entity falls out, the others take over. Therefore there is no single point of failure. In such a situation the availability is much higher. The entities however can still have their own agenda, and with every entity there is a new possible point for manipulation (either by the entity or by an attacker of the entity). A de-centralized reputation also is more technically advanced. There are issues: which entity is responsible for what, how are decisions made, Duplication issues, Data synchronizing and dealing with manipulating. In other words a de-centralized system is more complex. In a design of a reputation system such concepts should be elaborated on for the specific context.

In case of a centralized authority, the participants have to trust this party. In such a situation the focus is more towards creating a body of trust and overcoming the issue of availability. De-centralized calculating of a reputation has other issues, as overcoming the complexity of de-centralization. In de-centralized system trust is less important. Data is often kept with the owner, so there is less need for trust. There is much more duplication and often data stays with the entities, in other words the shift here is towards overcoming the complexity of de-centralization; how to keep data synchronized, exchangeable and the corresponding responsibilities (Hoffman, Zage, & Nita-Rotaru, 2009). The end result should be that disregarding the type of system the reputation system should work.

3.2.4 Intended users

The intended users of a reputation system influence the design of a reputation system. This is because the intended user of a reputation system and the characteristics of a reputation system are correlated. A user has a level of knowledge, and a purpose to use a reputation system (If the reputation system has no purpose for the intended user, there is no need to have it).

The intended user can be distinguished into four types:

- **The reputation receiver**: A reputation can be used to see how well the one or thing is doing. The reputation therefore can be a benchmark for the one receiving the reputation. If one receives a bad reputation it is apparent things have to be improved (although it is also possible to completely ignore it). These users can compare their private information with the reputation to see what to improve.
- **Competitors and stakeholders of the reputation receiver**: These parties are interested in how the receiver is doing. This is because either their business is influenced by it, or to benchmark themselves. As with the reputation receiver, they do have a high level of knowledge about the reputation receiver, but less than the reputation receiver itself.

- **A consumer**: A reputation about a company, product, service etc. is often conveyed to the consumer. The reputation is used to determine which company, product or service to select. These consumers often have a lower level of knowledge, as they are not an expert on such a topic.
- **Government and governmental organizations**: A reputation gives government a good indicator about actions the reputation receiver does.

3.2.5 Communication of a reputation system

It is important to have a reputation system which assigns a reputation to actual and expected performances. However, if nobody is aware of such a reputation system, what is the point of developing a reputation system? The calculated values have to be efficiently disseminated to others, or be available upon request. In practice the calculation and communication are often intertwined, the system that calculates a reputation automatically communicates it to the word. Nonetheless, when analyzing reputation systems, it is still important to discuss this topic.

For example, in corporate reputation, most people look for reputation from reading print (i.e. on a website). Estimations are that 83 percent of a company's customers, and 100 percent of the companies distributors look for the reputation of the corresponding company (Saxton, 1998, p. 394).

Platforms to report the reputation can be very different:

- Reading & printed text
- Social media
- Word-to-mouth

3.2.6 Cheating and strategic behavior

An important difference between reputation systems is how they deal with cheating, manipulation and strategic behavior. Sabatier & Sierra (2005) have defined three levels of cheating, so there are basically three options into which a reputation system can be distinguished when talking about cheating:

1. Cheating is not considered
2. It is assumed that agents can hide vital information, but they do not lie
3. Cheating is considered, but there are mechanisms to deal with liars (Sabatier & Sierra, 2005, p. 40)

These three levels correspond to how reputation systems deal with cheating, but do not describe how a reputation system can be "attacked". Basically, there are two parties whom can attack (i.e. cheat, manipulate or strategic behavior) the system. First there is the one onto which a reputation is assigned, they do not want a bad reputation, giving incentives to cheat. The second one are outside parties, someone or a group which is affected by the reputation. They could also attack the system to make sure that there is no reputation at all. Basically the difference is if the attacker is inside or outside of the system.

The next three subsections describe different aspects of cheating. First, there are different incentives to cheat, this is described in section 3.2.6.1. The cheaters, can use multiple attack types, these are described in section 3.2.6.2, and finally section 3.2.6.3 describes the mitigation or defense strategies against cheating.

3.2.6.1 Incentives to cheat

A reputation system should give the receiver of the reputation an incentive to achieve the highest possible reputation. The party that assigns reputation, is important because it can influence the power of this incentive. For example a strong incentive can be large media coverage or high financial costs for a bad performance (De Bruijn, 2006). Where a light incentive only leads to (for example) having some explaining to do.

A fine balance is required between having a strong incentive and a light incentive. If an incentive to cheat becomes very high a reputation system could lose its effectiveness. The receivers of the reputation face higher consequences for bad reputation, they will have more resistance to or do not accept a reputation system. Where if the incentive is too light, they simply might not care about it, since there are no consequences.

Another problem with a strong incentive because of high consequences is that it generates a secondary effect. If consequences are very high, the affected parties might be willing to put in a higher cheating effort. So the incentives become wrong. Where goal of a reputation system should be to provide incentives to the receivers of the reputation to improve their reputation. In a situation where consequences are too high, or wrong, the incentive shifts from improving to cheating, thus undermining the purpose of the reputation system.

3.2.6.2 Attack types on reputation systems

Hoffman et al (2009) identified five possible attack types on reputation systems:

1. Self-promoting: An attacker manipulates the reputation systems to falsely increase their reputation. They could do this by falsely representing input data.
2. Whitewashing: Attackers abuse the system. They find some vulnerability of the reputation system to restore their bad reputation to a good reputation. The reputation shows the wrong values and the one behaving bad can just go on.
3. Slandering: An attacker manipulates other data. So instead of increasing their own reputation, others are decreased.
4. Orchestrated: This strategy incorporates one or multiple of the above mentioned attacks, in a situation where there is not one attacker but many attackers working together. Such an attack is usually done by attackers outside of the system, where the previous ones are usually done by an attacker inside of the system (i.e. the one who is rated).
5. Denial of service: Attackers try to prevent the reputation system from calculating reputation values and spreading them. In other words, in such an attack the reputation system is forced to be shut off.

A system which uses feedback data from humans, or where new accounts can easily be made is much more susceptible to i.e. self-promoting. A more technical system where someone cannot

just create a new account is much less susceptible to this. This means that different reputation systems have different problems regarding these attacks. For every situation it has to be determined to which types and how reputation systems are susceptible to these attack types.

3.2.6.3 Defense and mitigation strategies

Hoffman et al (2009) have also identified some defense strategies, these are:

- Preventing multiple identities: In many reputation systems, people have the option of creating their own profile. I.e. in a buyer seller relationship, buyers with a bad reputation on a platform, can just simply create a new profile. This can be limited by introducing unique identifiers (i.e. a cell number, bank account number or social service number) or asking for a (small) monetary introduction when creating a profile. With a unique identifier, the identity is integrated with this identifier. When someone wants to use a new account their unique identifier is already been used in another account, so it becomes more difficult to obtain a new identity.
- Mitigating false rumors (spreading and generation): A slandering attack can be started by introducing false rumors or gossip. In data this means that entities in the system start to introduce false values about each other. It can be mitigated at two stages, in generation and in spreading. In generation it can be mitigated by introducing integrate accountability, digital signatures and irrefutable proofs. These cryptographic methods are important in mitigating this problem. Mitigating the spreading can be done by actively assuming such cheating and to develop probability functions to incorporate the level of (dis)honesty of entities. In such a function it is assumed that x percent lies. In such a situation they only look at the 100-x % of the values. In such a system the outliers are removed.
- Mitigating denial of service attacks: Denial of service attacks can be mitigated by introducing duplication and randomization. In such an attack the system is attacked at i.e. the machine where the reputation is calculated. But if there are multiple machines which can do this and it is either unknown which one does this or where it is, the attack becomes more difficult. Security mechanisms and duplication are thus important to mitigate such attacks.

As with the attack types, the defense types are again dependent on the type of reputation system. There are so many different variables that these strategies can be used as a starting point for mitigation. For every reputation system it has to be manually identified how mitigation can be done. Cheating in the context of ISPs is described in section 5.2.

3.3 From data to reputation

The previous sections have discussed the different aspects of reputation and reputation measuring systems. Reputation and reputation systems are very much related. The figure below shows the whole process, starting with data gathering to the end, where the reputation metric is formed.

A reputation (metric) is formed in three stages, as visualized in the figure below. It starts with the transformation from data to information. Data can originate from multiple sources. The

information from this data that is used to calculate a reputation is transformed (section 3.1.2). In the second stage, the reputation is calculated (section 3.2.2). The different factors from the previous stage are, through an algorithm, integrated into a single value, the actual reputation (section 3.1.1). In the next stage, the reputation is reported (section 3.2.5). It can be argued that the value that is calculated in the second stage, only becomes a reputation metric if the value is known by the relevant parties, as the intended users (3.2.4).

An important feature in the figure is the governance section (section 3.2.3). Someone or somebody has to determine what settings to select. The governance of a reputation measurement system is therefore very important, and drives the reputation system from data selection to the reputation metric. Factors as what (type of) data to select, which factors should be formed and how the reputation should be calculated and disseminated are all driven by the governance. For example, the simple question of when values can be considered good or bad, can be interpreted very differently by different parties. The ones that influence the governance of a reputation system are therefore important. The characteristics set in the governance therefore determine the success or failure of a reputation system.

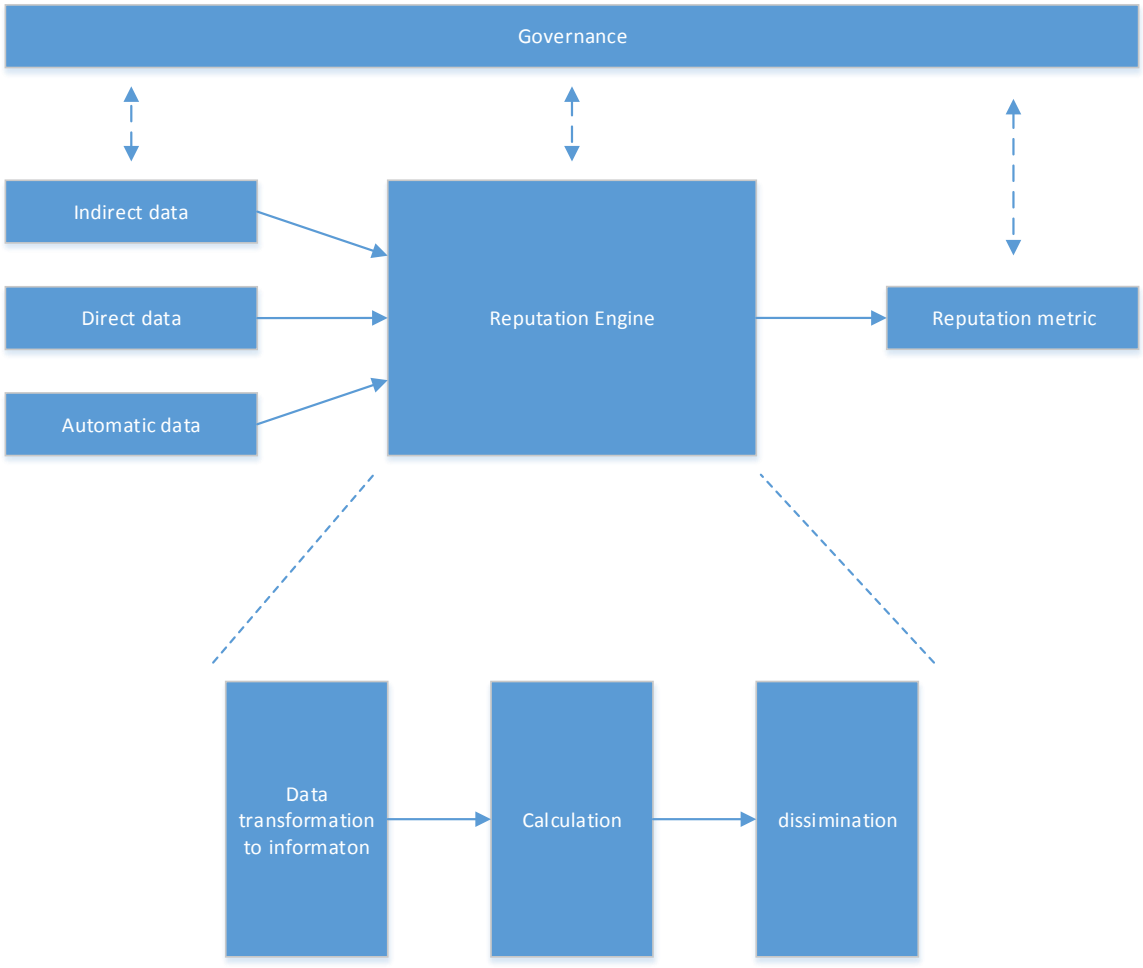

Figure 8 From data to reputation

3.4 Conclusions on reputation and reputation systems

This section covered the topics of reputation and reputation systems. In section 3.1 the concept of reputation is described. It showed that the concept of reputation is not a simple concept. Everybody has heard of it, and knows what it is, without knowing what it exactly is. So people know about it, but cannot clearly tell what it is, but many agree that reputation builds up over time, can be destroyed instantly and needs to be managed. But how a reputation exactly should be created is often unknown. In this research reputation is defined as:

"A reputation is the degree to which one party has confidence in another within the context of a given purpose" (from section 3.1)

A reputation can generally be classified onto two axis:

3. From general to specific
4. Based on human or machine feedback

In this section some references to **(Req number**) are made. They will be used in a later stage to form requirements, the reference in the conclusion is purely for traceability purposes (see footnote 2).

Reputation is set in a context **(Req 4[2], see footnote, to be used in section 5.1 later on)**. This context drives the level of specificity or generality of a reputation. The degree of confidence is based on information. Such information can be obtained from two types of data sources: Human feedback or machine feedback. The first one is definitely subjective, while the second can be more objective (sections 3.1.2.1-2). However, since the machine data also has to be selected it also involves human interaction.

There are different types of reputation, i.e. from a corporate reputation to a technical reputation which can be on different locations on these dimensions. Reputation and reputation systems are very much related, the two concepts are both described into one chapter. In section 3.2 the concept of reputation systems is therefore described.

A reputation system is an automated method **(Req 8)** that collects, distributes, and aggregates feedback about a participants' past behavior (Resnick, Zeckhauser, Friedman, & Kuwabara, 2000, p. 2). There are three basic requirements for a reputation system:

- Entities are long lived **(Req 5)**
- feedback about current interactions is captured and distributed **(Req 6)**, and
- past feedback guides future decisions **(Req 7)**

[2] The "**Req 1"** refers to a specific requirement. Such requirements will be used in section 5.1 do determine what a design should fulfill to. It is not hierarchical so the number does not refer to any importance of the requirement compared to other requirements. The reference in this text is to provide traceability of later requirements.

The reputation system is thus the method that calculates and reports the reputation. To assess the quality of a reputation system there are four criteria (see section 3.2.1): Accuracy **(Req 21)**, Weighting towards current behavior, robustness against attacks, and smoothness **(Req 15)**. To maintain these criteria; a reputation system should prioritize new information **(Req 16)**, be up-to-date **(Req 17)**.

A reputation system can be separated into five dimensions; the computation engine, the reputation assigner, the intended users, communication of the reputation system and cheating. The concepts of reputation and reputation systems merge together into a reputation metric. This happens in three stages; 1) data conversion; 2) calculation and 3) communication.

Data conversion is the process of turning data into information. This is the conversion to be done before a reputation can be calculated. In these first two steps it is critical to determine possible attack and corresponding defense techniques to account for cheating. Cheating can be done from inside of the system (i.e. insider information manipulation) or outside (i.e. by attacking the system). Ideally a reputation system should be resistant to both **(Req 9, 10)**. However some models can also assume cheating cancels itself out over time (sec. 3.2.6). Communication of a reputation is important. There can be intended users (3.2.4), but if they do not match the communication method, the reputation system remains unknown **(Req 11)**. The intended users should be able to access and use the reputation system **(Req 13, 14)**. This means that the machines and algorithms calculating the reputation should work **(Req 12)**.

This chapter has shown what a reputations and reputation systems consist of and what criteria assures the quality of a reputation system. The chapter gives some insight from theory of how a reputation and reputation system is formed (the transition from data to reputation, see section 3.3). However, how reputation systems work in practice is still a bit ambiguous. This can be because the theory has specified dimensions where reputations (systems) consist of, but the theory does not show how these dimensions in practice precisely operate. For this reason are in the next chapter existing reputations from other fields in practice described.

4. EXISTING REPUTATION SYSTEMS FROM OTHER FIELDS OF STUDY

In the previous chapters the current situation (Ch.2) and reputation systems (Ch.3) have been introduced. From chapter 2 it became apparent that there is a serious problem with botnets; ISPs can mitigate these botnets, but the incentives to do so are misaligned. There is a need to have a reputation system where ISPs are rated based on botnet activity, but this is missing. Since such a system is still missing, the concepts of reputation and reputation systems have been researched in chapter 3. In chapter 3 the dimensions for reputation (systems) are described, but how the reputation is obtained in practice is still relatively unclear. For this reason chapter 4 researches existing reputation systems from other fields of study.

These sections introduce a reputation system from other fields of study, other sectors or other industries. In those fields there is already more experience in assigning reputations. This way they can contribute to mitigating the botnet problem. Every described reputation system is an existing initiative. They are systematically described in the next sections, first a reputation system is introduced. In the introduction is described how they work. After the introduction they are categorized according to the identification of dimensions reputation systems from chapter 3, the characteristics are described. Finally their quality is assessed based on the criteria set in par. 3.2.1. There are four criteria which assess the quality of a reputation system. These criteria together should basically answer the question: does the reputation system work in practice? The evaluation criteria are: accuracy, weighting towards current behavior, smoothness and resistance to cheating.

This chapter is structured in the following order. First other reputation systems in the digital world are evaluated; these are PageRank and EBay's feedback forum. After this corporate reputation is evaluated, since an ISP is also a company might there be usable content from that field. Finally a well-known rating system is described; this is the rating system from the financial world (by S&P). This is done in respectively section 4.1 to 4.3. The final section shows the conclusions.

4.1. Reputation systems in the digital world

This section covers two initiatives widely being used in the digital world, these are PageRank and eBay's feedback forum (resp. in section 4.1.1 & 4.1.2). Each initiative is introduced, characterized and evaluated.

4.1.1 PageRank

The next subsections respectively cover description of the reputation system in the introduction, the characterization and the evaluation.

4.1.1.1 Introduction

PageRank is the name of probably the most commonly known rating system, Google. In its essence, Google ranks pages in an order of importance (Page, Brin, Motwani, & Winograd, 1999). PageRank is one of the algorithms they use for ranking. Below is simplified version of the algorithm is given.

Simple representation of PageRank

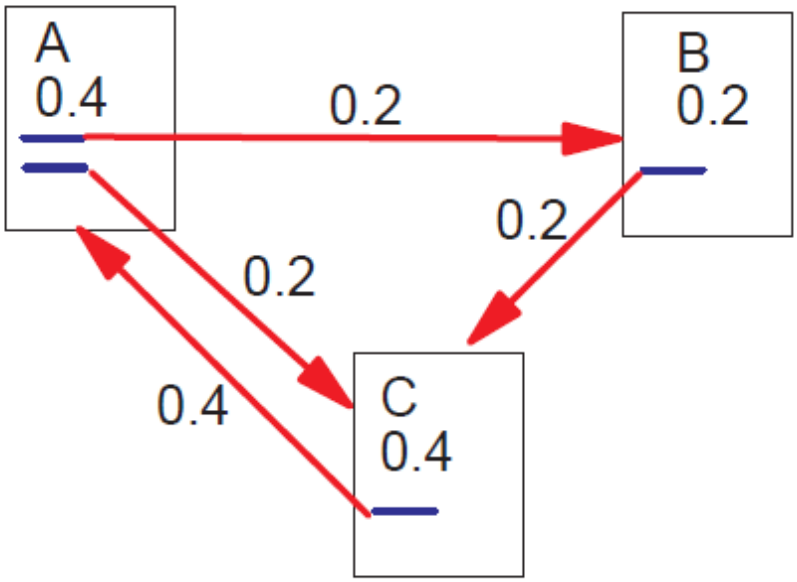

Figure 9 Simple representation of PageRank adapted from (Page, Brin, Motwani, & Winograd, 1999)

Every page has the same starting value (see Figure 9). For every web page the number of hyperlinks to this page (inbound links) is measured. On every page the number of outbound links is measured. The rank of a page X, is than the sum of the value of the page with an outbound link to page X divided by the total number of links on that page. This process is iterative so the page values change and the starting values per page are updated. More links to a page gives a higher score, although the actual algorithm also incorporates a damping effect.

That it is measured objectively does not mean that it is not possible to manipulate the system. Higher ratings are sold by the owners of pages that have high ratings. A page with a high PageRank, page Y can offer to put a link on the page X. Since the starting value is higher, the value given to the page X, is higher because of the high value of the page Y.

4.1.1.2 Characterization

Where the previous section explains the initiative, does this section characterize it. The dimensions identified in the previous chapter is used for this. PageRank uses objective information to determine the ranking of a webpage. It has a high level of complexity because of

the multitude of vectors that have to be used to calculate the score. The level of detail is very specific, it only ranks webpages, and is set in one context.

PageRank uses vectors and weights to determine ranking, and therefore is a flow model. There are possibilities to cheat, but over time these are likely to be insignificant. The model can be attacked by using false sites with multiple links to the page which has to be ranked higher. With such a system, the rank can be increased because these links increase the value or importance of a specific page. The algorithm is controlled by Google, they set policies and make decisions about the algorithm. They have a lot of independence in this regard, but they do have to abide by government rules. It has become publicly known that the large degree of independence has given Google the opportunity to favor their own services in search results. The European Commission have warned Google about honestly representing search results or face a fine (BBC.com, 2014).

In section 3.1 a definition for reputation is given. According to this definition, the degree to which a party has confidence in another for a specific context determines the reputation. The question here is: Is this initiative a reputation according to this definition?

Google assigns a ranking based on the number of referrals from other pages. Someone types a set of search criteria. In the context of this set of criteria a ranking is provided. The context of a purpose part is therefore met. The ranking is only valid for the search criteria. The ranking is based on those referrals. It can be argued that another site, posting such a referral, determines a page to be important. If in this case confidence is interpreted with the meaning of importance, it can be concluded that this algorithm fits the definition of reputation as set in section 3.1.

4.1.1.3 Evaluation

This method is probably the most used and known initiative, since almost everybody on the internet uses Google. PageRank is an accurate algorithm, making it possible to rank webpages, and provide the navigation on the web. The version employed by Google is very accurate because of the many iterations. The algorithm continuously working on updates, by crawling and indexing the web. It is very much weighted towards current behavior, because the process of ranking the web mainly uses the information as is, instead how pages were. The PageRank does not look back, it is continuously updated and if a link is removed from a site, the score does not take the older value into account after the iteration. It is not very smooth since the rankings can change suddenly and do not look at older values.

Given the anti-measure, punishment for falsely increasing rankings, which are enforced, it can be argued that the algorithm is very robust against attacks in the long run. It is possible to cheat by using fake sites. Sites with the purpose of falsely increasing ranking or sites that use false ranking are punished (BBC.com, 2014). There is some sort of domination, bigger sites get more importance. This makes is more difficult to increase in the ranking as a smaller site. Although Google is externally resistant to cheating, it is not resistant to cheating from inside the algorithm. Google is accused of favoring their own services, and putting them on top of the list. It is hard to prove this, since outsiders cannot look inside the system, but the European

Commission is accusing Google of false favoring their own services and the conclusions about this are still pending (BBC.com, 2014).

To conclude: this methodology works well in practice. It is accurate, robust to attacks and favors new information over old, only it is not very smooth. It can be used to provide a reputation for ISPs. The algorithm is already used in a performance measurement initiative by Wagner et al (2013). However, this initiative is still very limited as it only connects Spam from blocklists to autonomous systems, instead of ISPs. It is possible to incorporate the mapping AS to ISP to such a system. An issue with the solution proposed by Wagner et al (2013) is that it only uses Spam and leaves the topic of botnets aside.

4.1.2 EBay

4.1.2.1 Introduction

Nowadays, many people make transactions using the internet, i.e. using eBay. Such transactions are associated with risk, i.e. will the seller on eBay really ship the product, or would the product be in the same condition as mentioned online? EBay solved this problem by introducing a reputation system, where the buyer and sellers can be rated (either positive, negative or neutral). Buyers can leave their feedback about the seller. Obviously eBay is not the only one, which introduced a reputation system.

EBay's reputation system, the Feedback Forum, gives buyer and seller a chance to rate each other (with a -1, 0 or 1, for negative, neutral and positive), buyer and seller can also leave a comment about the transaction (Resnick, Zeckhauser, Friedman, & Kuwabara, 2000). People using eBay have totals of these scores attached to their screen name. Users can see this score and the feedback that is given to them; the idea is that positive feedback should help to trust a seller. When someone has the choice between two identical products and price; one sold by a seller with positive feedback and one with negative feedback: they would choose the one with positive feedback. The sellers with negative feedback are in a way "forced to improve" since they have to compete with sellers with positive feedback.

In general, this system would work perfectly fine, however there is quite some space here for strategic behavior by eBay users with bad reputation. On the internet, people often use screen names, which are easy to change. For someone with a "Bad reputation" it takes almost no effort to just create a new account and start over. Also users can manipulate the system by introducing false feedback, i.e. two (or more) users pretend to buy from each other and consistently give a positive rating (Josang & Golbeck, 2009).

There are different versions of eBay's algorithm, i.e. another variant is the beta reputation system. This is another variant of the above described eBay version, proposed by Josang et al (2002). The Beta Reputation system is an answer to known issues with rating systems as that of eBay. It uses a probability density function, to overcome some of the known issues, discounting (see above for more), with rating systems (Josang & Ismail, 2002, p. 3). Using the positive and negative ratings as binary numbers (so positive is 1 and negative is 0) a reputation function can

be built. It gives a value between 0 and 1, so 0 is a bad rating, 0.5 is neutral and 1 is a positive rating (Josang & Ismail, 2002, p. 5).

The reputation rating for entity A can be calculated by the subtracting the amount of collective negative information from the positive, divided by the sum of the positive, negative information and two, as shown in the formula below.

$$Rep\ A = \frac{(Positive\ information - negative\ information)}{(Positive + negative\ information + 2)}$$

The end result is a range between [-1;+1]. To account for forgetting or discounting the method of deriving the positive information is changed, for further information about this is referred to page 6-7 of Josang et al (2002).

However, different agents can provide feedback; these agents itself can also have a positive or negative rating. Agents with positive ratings can be seen as more reliable, thus their feedback should also be more reliable (and vice versa). These differences should be discounted in the reputation function.

Josang and Ismail (2002) have also determined that older feedback is less relevant. So a forget function should be used to give newer feedback more value than older feedback (Josang & Ismail, 2002, p. 7).

4.1.2.2 Characterization

EBay's voting system and the beta reputation system are variants of each other and are therefore described together. Both reputation systems get their data from user feedback: data thus might be biased. Users do not have to give the right information, so information is or can be subjective. The systems have data from one source and a general level of detail. It is a general level of detail because it applies to multiple context (i.e. both buyer and seller are rated). Both systems have mechanisms to mitigate cheating, but users are always able to just create a new account. The difference between the two is the way it is calculated. EBay uses simple summation where the other proposal uses a Bayesian probability density function.

The intended users of the system are the people on eBay. Such a system is a classic example of a reputation system that is very susceptible to be attacked. People with a bad rating, can very easy set up a new account with a new email address, and start again with a neutral rating.

The policies set for the reputation system are all set by eBay. They assign the reputations, and have the power to make decisions about the platform.

In section 3.1 a definition for reputation was given. According to this definition, the degree to which a party has confidence in another for a specific context determines the reputation. The question here is: Is this initiative a reputation according to this definition?

In the system others can assign feedback about other buyers and sellers. This feedback can be positive: showing confidence or negative: showing no confidence. The confidence is set in a

context: the confidence in the buyer or seller. Therefore the definition of reputation, as set in section 3.1, is met eBay's algorithm.

4.1.2.3 Evaluation

The reputation on the eBay platform only uses positive, neutral and negative information (i.e. -1, 0, +1). With an observation it is not specified how positive or negative somebody performed. Another issue in the reputation is that many users do not provide feedback, they only provide feedback if they have something to say, i.e. if something was disliked. A reputation therefore can be negatively biased, and less accurate. There are also buyers and sellers that only buy or sell occasionally on the eBay platform. Occasional sales and none mandatory feedback makes it that a reputation is often based on a few observations. The two issues, negative bias and a few observations can result in a less accurate rating. For example: somebody has three feedbacks, but has sold much more items than that. Out of the many sales he or she made, two have gone wrong resulting in negative feedback. The positive sales are often not rated, thus there is not much positive feedback to compensate this. In this example the seller would have a negative reputation, while he or she provides most of the transactions with a good result. The accuracy is thus in this example wrong.

The initiative favors new information over old, thus current behavior is more important than old and it is given more weight.

There are many measures taken to prevent cheating, but cheating remains a big problem, since some user can easily change accounts. The combination of a low robustness against attacks and accuracy gives problems. Cheating forces the accuracy to be lower, so this system has problems with both. Since it is not obligated to provide feedback, the system is not always used but every buyer/seller can see a score. In such a form this method cannot be applied to a reputation system for ISPs, but the general idea could be used. Potentially data could be transformed into a similar ordering with positive, neutral and negative assigned by the party executing the reputation system. This way such a reputation system could be applied. However the issues regarding cheating make it questionable if the system would ever be resistant to cheating.

4.2 Corporate reputation systems

This section covers corporate reputation systems. There are two leading corporate reputation systems, the reputation quotient, and the MAC index. They are very much related and are introduced together in 4.2.1. The two reputation systems are characterized and evaluated together in sections 4.2.2 and 4.2.3.

4.2.1 Introduction

Fombrun, Gardberg & Sever (2000) have developed a reputation quotient model in which 6 dimensions have been defined which contribute to a reputation. These dimension again are operationalized in several indicators. These dimension are (Fombrun, Gardberg, & Sever, 2000):

- **Financial performance:** the perceptions of the company's profitability, prospects and risk

- **Emotional appeal**: how much the company is liked, admired and respected
- **Products and services:** perceptions of the quality, innovation, value and the reliability of the company's products and services
- **Vision and leadership**: how much the company demonstrates a clear vision and strong leadership
- **Workplace environment:** perceptions of how well the company is managed, how it is to work for, and the quality of its employees
- **Social and environmental responsibility:** perceptions of the company as a good citizen in its dealings with communities, employees and the environment

Together these dimensions form a reputation. Each dimension again is measured by several indicators which are often measured by questionnaires. The reputation quotient can be measured on a 10 point scale. The focus in this reputation quotient is only on one stakeholder (the general public) (Wartick, 2002, pp. 385-386).

Fombrun & Gardberg (2000) have defined some principles that correspond with his Reputation Quotients' dimensions; these could be interpreted as general criteria when measuring reputation. These principles are conclusions from the research and are mentioned below (Fombrun & Gardberg, 2000, pp. 15-16):

- **The principle of transparency**: if companies are transparent in their operations the reputation grows
- **The principle of consistency:** if companies are consistent in their actions the reputation becomes stronger
- **The principle of focus:** focused actions around a core theme has a positive influence on reputation
- **The principle of distinctiveness:** reputations grow if a company has a distinctive position in the minds of resource holders
- **The principle of identity:** Strong reputations result when companies act in ways that are consistent with the principles of identity

These five principles could potentially be applied to other reputation system types. There are two other variants, these are described in the next sub sections 5.2.1 and 5.2.2.

MAC index

MAC stands for most admired companies. It is an index that ranks companies. This MAC is constructed in a similar way as the Reputation quotient; however, it gives a ranked list of companies where the Reputation Quotient gives a kind of grade to a company's reputation. It is measured similar to the Reputation Quotient, it uses the following dimensions:

- Ability to attract and retain talented people
- Quality of management
- Social responsibility to the community and the environment
- Innovativeness

- Quality of products or services
- Wise use of corporate assets
- Financial soundness
- Long-term investment value
- Effectiveness in doing business globally

As with the Reputation Quotient, most of these indicators are measured by using questionnaires. From the literature it becomes clear that with corporate reputation, the measurements heavily rely on soft variables measured by people's opinion.

4.2.2 Characterization

Such a reputation system uses two types of information: direct and indirect information, thus information based on human feedback and data based on machine measured values. The reputation system has a specific goal: rank a company and is thus specific. The reputation is used in one context: corporate reputation.

Such a reputation system is calculated by a summation over the different scores on the dimensions, therefore it is also of this archetype. However, many concepts are also measured by questionnaires: using ordinal scales as good-neutral-bad. Therefore it is also a discrete model.

A corporate reputation can be used by many different stakeholder, it has many intended users: the company itself, other companies and customers. The reputation can be used by the company itself, to see how it is doing in someone else's view. Other companies can use the reputation as a benchmark, and customers could use the reputation to determine if they should do business with them. There are many intended users of such a reputation.

In such a reputation system there are possibilities for cheating. Questionnaires are often biased. In calculating such a reputation, many sources of data are used. It is likely that cheating cancels itself out, because many respondents are used.

The reputation is assigned by a reputation institute, for example Forbes. They influence decisions about the reputation system and thus can be seen as a central authority. Such an institute is the main body of power when it comes to decisions. The companies which are rated have little influence in setting the guidelines of the reputation system. The authority is one sided, only by the reputation institute.

In section 3.1 a definition for reputation is given. According to this definition, the degree to which a party has confidence in another for a specific context determines the reputation. The question here is: Is this initiative a reputation according to this definition?

In general these initiatives correspond with the definition set in section 3.1. The reputation here shows how much confidence a reputation institute has in a specific company. The corporate reputation. It can be argued however that there are better definitions to suit these initiatives. I.e. Fombrun (1996) can be referred to in order to find more information about a corporate reputation.

4.2.3 Evaluation

Both the reputation Quotient and the MAC Index assign a corporate reputation to a company based on perceptions about companies and corporate values. These reputation systems use for a large part indirect information. As already described in chapter 2, indirect information is usually subjective. The reputation measures the perceptions of a company, instead of the actual company reputation. For this reason it is less accurate, since people can have the wrong perception. In determining the reputation, the weight is with the current values, but older values do also still hold weight. This means that the objective "weighting towards current behavior" is met.

These reputation systems are very vulnerable to cheating. Since for a large part they rely on questionnaires, there is also much room to provide inaccurate information. People do not have to provide the correct information on a survey. These reputations are also assigned by powerful institutes, i.e. Forbes. As with the PageRank example, it is possible that such an institute might favor their own services or companies were they do business with. There is minimal supervision over these reputations, therefore there is room to cheat the system from the inside, although this is a hypothetical situation.

The reputations are often widely known and publicized. Therefore these reputation systems are successful in this regard, but it remains questionable if they are accurate. In assigning a reputation to ISPs, these reputation systems in this way are unusable, but there are lessons to be learned from these.

The previous examples: i.e. the rating agencies or PageRank, only use technical values, extracted though automated means. They assign a value based on hard measured values: the output of a company's efforts (i.e. with the case of S&P this output is the companies (annual) figures). This raises the question: is a reputation only to be based on output, or should efforts to improve the output also be accounted for? ISPs could also be given a reputation which is (partly) based on their mitigation efforts.

A reputation where also the efforts are measured would introduce a more complete overview of the mitigation efforts. It is based on what is valued higher: efforts or results.

4.3 Reputation systems in the financial world

The financial world uses reputation in another form. It is called a rating. The corporate credit framework S&P uses is described in the sections below. Also an interview was done with Ms. M. Pieterse -Bloem, an expert in the field of finance about the topic of rating agencies.

4.3.1 Introduction

The financial world is full of rating systems for financial products, companies and even countries. Several institutes assign ratings, examples are Standard and Poor's (S&P), Moody's or Fitch. Most of them work in a similar fashion, therefore for this project (the rating system S&P) will be researched, but S&P assigns ratings to many products. Since ISPs are also companies, they will be compared to the rating system for companies S&P offers, instead of the ratings S&P offers for countries or financial products.

S&P are fairly transparent in the methods they use for issuing ratings. They provide the general methodology: only the specifics how they weigh factors or how they derive values they leave aside. On their website, after you request an account, the concepts and methodology they use are explained and operated. In this section two of these are used, the S&P Criteria for the corporate methodology in America and the S&P General Criteria: the principles of credit ratings.

Companies can be assigned a rating ranging from AAA to CC (D is also possible but in that case people speak of bankruptcy). A triple A rating is the most positive for a company, where CC is bad. Such a rating is important, especially for bigger companies (small companies usually are not rated), because first of all it shows to customers and distributors that a company is doing well, second an S&P rating is basically just a risk profile for a company (S&P, 2014a)., A triple A rating reduces the interest rate the company has to pay when issuing debt. A bad rating might even prevent a company from issuing debt, or would give a much higher interest rate. S&P assigns a rating, and corrects this semi-annually, so twice a year.

Companies usually hire S&P to rate them, this is an important factor to remember here. Companies as S&P are commercial companies, which are paid to give a rating to a company. The company uses this rating in turn to attract entering the debt market. The incentives for rating agencies can be wrong, these agencies want to sell as many ratings as possible (since they want high profits), therefore it is difficult for them to give a rating that is too low. This means that such rating agencies are more likely to give a rating that is higher than it should be. In the financial crisis of 2008 it showed that these rating agencies were not innocent here (Smith, 2008). In the subprime financial crisis there ratings agencies have had a large role by systematically assigning (prime) AAA ratings to subprime products as CDOs. Nowadays these rating agencies are more strictly regulated, especially in the EU.

In the assessment of this risk S&P uses factors as industry risk, country risk and the competitive position of a company to determine the business risk (S&P, 2014b). Other than business risk, S&P also determines financial risk, based on cash flows and leverage ratios. The business risk and financial risk, together form an anchor: a kind of base rating. This anchor can be modified by diversification (also called the portfolio effect), the capital structure of a company, the financial policy, the liquidity, management styles and comparable rating analysis. After these modifiers, a standalone credit profile is determined.

The rating analyst can opt to in or decrease the rating up to two notches in the end. This is based on his or her own interpretation and it such a decision has to be argued why to a board of analysis's (Pieterse-Bloem, 2014). If there are no changes by the analyst, this standalone credit profile (SACP) is also the rating. Otherwise it can be adjusted pending approval (Pieterse-Bloem, 2014).

There are two types of ratings to be used. S&P can give a solicited rating: in this case the rating agency is invited to assign a rating, and they get full cooperation from the company. It is also possible the rating agency gives an unsolicited rating, in this case S&P can only use public records as annual figures (Pieterse-Bloem, 2014).

The figure below shows this process.

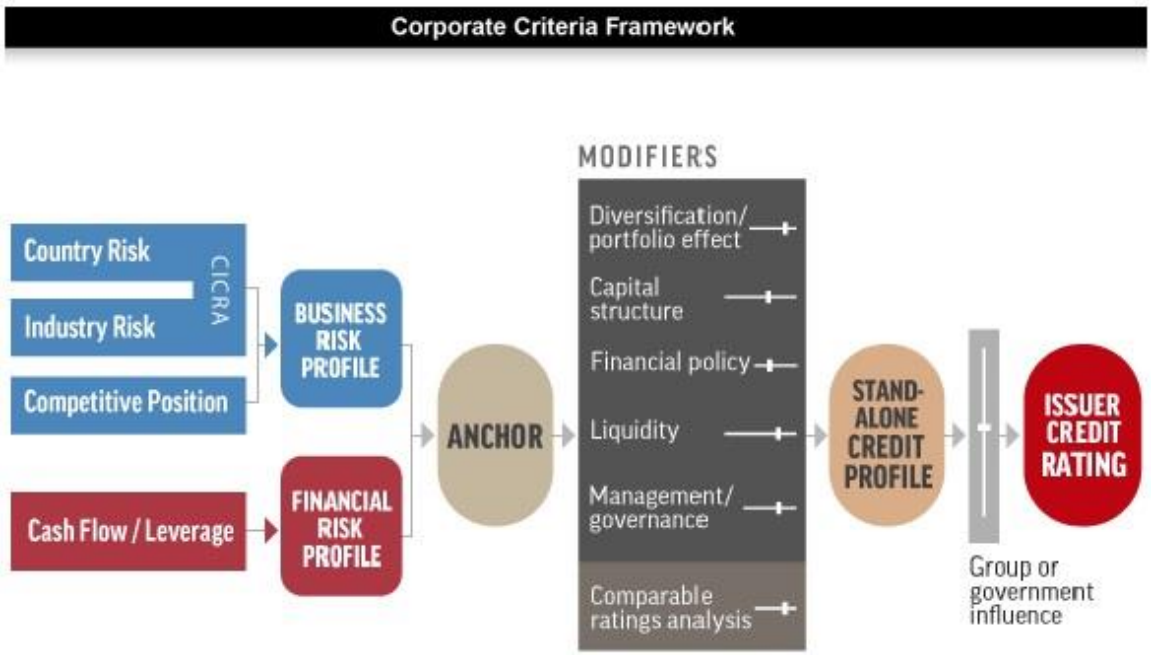

Figure 10 Corporate criteria framework adapted from (S&P, 2014b, p. 2)

Business risk profile

The business risk profile covers the issue of risk and return for a company and the markets it is active in. The idea is that higher risks should yield higher (potential) returns, since the risk of losing valuable investments is also higher. The risk of financial loss should be compensated with higher potential return.

In the business risk profile, the competitive climate is also investigated (as the industry risk), the same holds for country level risk. The country risk and industry risk together with the competitive advantages and disadvantages form a business risk profile. This business risk profile affects the financial risk profile (the two concepts are interrelated). It is the foundation for determining the potential success of a company (S&P, 2014b, p. 1).

Industry risk and country risk are estimated on an ordinal range from 1 to 6, (so very low risk =1, low risk = 2, intermediate risk =3, moderately high risk =4, high risk =5 and very high risk =6). The competitive position is rated also on a scale from 1 to 6, with 1 being excellent en 6 being vulnerable.

Financial risk profile

The Financial risk profile is influenced by management decisions that evaluates the business risk profile with the financial risk profiles. The relationship between business risks and financial risk tolerances are important here. The methods used in funding the company are important here (i.e. is it funded with equity or by a bank, the debt-to equity ratio).

Similar to the business risk, financial risk is also rated on a scale from 1 to 6, with 1 being minimal and 6 being highly leveraged (so a lot of debt) (S&P, 2014b, pp. 3-4). These ranges are calculated from i.e. values of the balance sheet and ratios as debt/EBITDA (S&P, 2014b, pp. 17-19).

Anchor

The business risk and financial risk are put together into a matrix, with financial risk horizontally and business risk profile vertical. Both are rated on a scale from 1 to 6, so a company with an excellent (value 1) business risk profile and minimal financial risk profile (also value 1) is rated aaa, and a business and financial risk profiles of 6 give a b- rating. This initial rating is called the anchor.

The matrix used is shown in figure below. This anchor is used in the next stage, were it is adjusted according to the modifiers.

Combining The Business And Financial Risk Profiles To Determine The Anchor

	--Financial risk profile--					
Business risk profile	1 (minimal)	2 (modest)	3 (intermediate)	4 (significant)	5 (aggressive)	6 (highly leveraged)
1 (excellent)	aaa/aa+	aa	a+/a	a-	bbb	bbb-/bb+
2 (strong)	aa/aa-	a+/a	a-/bbb+	bbb	bb+	bb
3 (satisfactory)	a/a-	bbb+	bbb/bbb-	bbb-/bb+	bb	b+
4 (fair)	bbb/bbb-	bbb-	bb+	bb	bb-	b
5 (weak)	bb+	bb+	bb	bb-	b+	b/b-
6 (vulnerable)	bb-	bb-	bb-/b+	b+	b	b-

Figure 11 S&P anchor list adapted from (S&P, 2014b, p. 4)

Modifiers

Modifiers do not change the financial or business risk profiles, but they can change anchor by changing up or down a notches.

For example diversification/portfolio effect can have a positive effect on the ratings. I.e. a company that is well diversified: a company as Unilever that sells a multitude of very different products, can get a rating up to two notches higher because of diversification. Although bad diversification cannot lead to a lower rating in this case.

Factors as financial policy, liquidity and management governance are modifiers with a more “punishing” in character. They manly lead to lower ratings in case of bad scores, sometimes up to 3 notches lower. The capital structure can have an effect from 2 notches up to 2 notches down.

After these modifiers, the anchor can still change, i.e. by the governmental influence on a company.

Governance

S&P assigns a rating and changes it (if there are changes) two times a year. Probably because they have to use company records, which are published (for larger companies) usually quarterly or per half a year.

For the governance, it is very important to note that S&P is paid by the company to which it assigns a rating; so their incentives have not always been in the right place and in the past S&P has assigned to high ratings, because it keeps their clients and customers satisfied.

4.3.2 Characterization

In section 3.1 a definition for reputation is given. According to this definition, the degree to which a party has confidence in another for a specific context determines the reputation. The question here is: Is this initiative a reputation according to this definition?

This can simply be answered as yes. The rating from S&P and alike rating agencies simply show the confidence a rating agency has in a company to pay its debts over the coming years (Pieterse-Bloem, 2014). So the confidence criterion and context criterion is met.

S&P uses data that is extracted trough automated means. They extract company data from financial databases, (annual) figures about a company's performance are used to determine the rating, therefore the information type is initially direct and relatively objective as it is mostly based on financial data. However in a later stage, they also use some professional judgments to modify the rating. The goal of such a rating is also specific. S&P uses different types of ratings for different types of purposes. In other words, for different scenarios they have different methods, making the rating specific and for one context.

The way the score is compiled is by a discrete model. Numerical values are transformed into ordinal levels. The intended users of the reputation, are ranging from the company that is evaluated itself to financial institutes to determine if they should assign a loan to such a company. The reputation communicated by the receiver of the rating itself, it is used as benchmarking and it is a vital to attract foreign capital (debt).

There are incentives for cheating. Misrepresenting numbers could give a higher rating, the company which is rated, can attract more capital. In the past this also has gone wrong, because the party that receives the reputation, pays S&P to determine it. For S&P there is an incentive to provide a good rating, to keep them coming back. It can be assumed that companies and S&P can cheat, but S&P again is also monitored by governmental agencies and still has a reputation to uphold. Therefore there are cheaters here, but there are also mechanisms to deal with it.

S&P is paid to by the one which is receives a rating. The policies in the reputation system are set by S&P, but they are not completely free to set these policies. S&P also has to uphold a reputation as trustable rating agency. As mentioned this has gone wrong in the past. Nowadays government intervenes and supervises S&P. S&P also is not the only party which assigns ratings, for example Moody's offers a similar service. Although the rating does not have to be identical, it should be in the same ballpark. This means that when a company is rated by two rating agencies, the difference for a company cannot be too large. Such a situation would suggest that

at least one of the rating agencies is wrong. Since the differences between rating agencies cannot be too large, rating agencies are also bound in setting policies by their competitors. The reputation is assigned by a central authority, but this authority does not have all the power, it is bound by competitors and government.

4.3.3 Evaluation

Generally it can be concluded that a rating is accurate (Pieterse-Bloem, 2014). The multi-tiered model, with a base rating, continuing with modifying it in later layer helps to increase the modifiers. Specifically, the modifiers (Figure 10), help to increase the accuracy. The accuracy also originates from competitors in the market. S&P has competitors, so they can benchmark their methodologies to other rating agencies.

S&P assigns such a rating only twice a year, because of the availability of data. This means that although a rating can be accurate, it is lagged. This lag is also due to taking older information into account. Previous ratings and/or previous values as company specific values are also taken into account.

A rating is fairly resistant towards attack from outside of the rating system. A company which has to be rated, could provide false information. However much of the information for the ratings is obtained from company figures, which also have to be handed over to revenue services or for the stock market. Cheating the rating in such a way would mean that companies also have to commit i.e. tax fraud. Therefore cheating by the company to be rated is less likely in such a way. As mentioned in 4.3.1, S&P is not exactly innocent in the recent financial crisis of ‘07/’08, where they purposely have rated financial products higher to achieve higher profits.

The argument can be made that rating agencies have (had) too much freedom: the control over these agencies was insufficient. The last years the governmental influence over rating agencies have increased significantly (Pieterse-Bloem, 2014).

S&P does not use a continuous scale, but a layered (ordinal) scale. This is partly because a rating is only given two times a year. Therefore in several months the situation could change, reducing the smoothness. The system however also looks at older values, so it is both unsmooth and smooth. The layered scale makes is less smooth, since it is in steps. The rate of assigning a value twice a year means that it potentially can change significantly, making it less unsmooth. When a new rating is determined, the older values are also taken into account, making it smoother.

The ratings are widely used. Not only business are rated, but countries, stocks, bonds and other financial products are also widely rated by these agencies. Many know and use these rating. Larger companies have to obtain such a rating. It can therefore be concluded that these ratings and rating agencies are widely known.

Overall this methodology seems to work well in practice. It is accurate, favors new information, fairly resistant to attacks and smooth. Also the ratings and rating agencies are well known and used, these agencies make high profits by calculating these ratings for various stakeholders.

4.4 Conclusions

In this chapter different reputation initiatives already used in practice have been identified, explained, characterized and their functioning is evaluated. Reputation systems in the digital world (PageRank, eBay), in corporate reputations (Reputation quotient) and in the financial world (rating agencies) have been used as examples.

Some facts have been identified or confirmed. One of these is the question: should a reputation system measure efforts or results? A reputation, where also the efforts are measured would introduce a more complete overview of the mitigation efforts. It is based on what is valued higher: efforts or results. If efforts are to be used as a reputation system, an initiative as the reputation quotient could be used for determining a reputation for ISPs. A problem with this framework is that it is very unclear how it is exactly measured, probably because it contains company secrets. The criteria are known, but the actual operationalization is not given.

The framework by S&P could be seen as a general framework. The initiative, as by S&P, could be applied to develop a rating system for ISPs, where they are rated according to their botnet mitigation activity. The Business and Financial profiles could be adapted, where instead of financial or company values, botnet metrics are used. The two tiered methodology of assigning a rating could be used as a red line for design. The content of this methodology needs to be adapted to another situation. The botnet metrics could be used where daily and monthly averages are compared to industry wide trends.

In a way, the framework proposed by S&P (see Figure 11), could be seen as a general framework which could be used in many situations. The current state of metrics measures data about specific botnet and or spam. Botnets can be taken down, or new botnets originate, an adaptable framework is a good solution because it is usable in a new situation.

A reputation initiative as by eBay has been described. In practice it could be applied to ISPs, but there are a lot of issues with such a rating system related to cheating. For this reason such a framework is not really usable as guideline for design. Adapting such a system as with eBay would mean that either ISPs start giving each other ratings, or customers of ISPs can assign feedback to how well an ISP is performing in terms of botnet mitigation. Both are not really feasible. Wrong incentives for ISPs and rating each other would not work since they could just simply give the others wrong ratings, or give each other high ratings such that everyone has a high rating. Customers rating ISPs is in this stage also unfeasible, as it is identified that most of the customers do not have the knowledge to determine this.

The PageRank algorithm is already used in practice as a performance initiative for autonomous systems, only based on spam. PageRank uses vectors, such a vector could indicate how much spam is given on to the next ISP.

This chapter has also shown some requirements literature from chapter 3 did not identify. Most reputation initiatives in this chapter account or correct for the size of the entity and industry trends **(Req 23, 24)**.

In this chapter most of the initiatives obliged by identified requirements from section 3.4. However not all requirements were met. This means that there are differences in theory and

practice. For example: not all systems are resistant to cheating (PageRank, eBay), previous behavior is always not taken into account (PageRank) and some also give a reputation to a short living entity (eBay). Therefore requirements 5, 6, 9 and 10 are not met in practice. S&P gives ratings twice a year, therefore the rating is not up to date (req 17). The field of corporate finance uses questionnaire data, it is questionable if such a system is always reliable (req 21). In appendix C a table is provided which shows how the requirements correspond to the existing alternatives.

The previous chapters have introduced three topics so far: the current situation (Ch.2), literature on reputation (systems) in chapter 3, and in this chapter existing reputation systems from practice. The knowledge from these three chapters together is merged in the next chapter (ch. 5) to form the design requirements and the design space. These two originate from all the chapters which have been described so far.

5. REQUIREMENTS AND DESIGN SPACE

This chapter shows the requirements and the design space. The requirements are derived from literature about reputation systems, institutional settings as the ISP market and the technical state of research regarding botnets. The requirements are criteria for design (next chapter).

The knowledge obtained in the previous chapters (Ch. 2, the current situation, Ch.3, literature on reputation and Ch. 4, existing reputation is joined together to design requirements and design space. The requirements and design space are the bridge between theory (Ch. 2-4) and design (Ch. 6). The requirements, section 5.1, show what a design has to oblige to. The design space, section 5.2, shows the possible design features. The difference between requirements and design space is that the design space shows what is possible, while the requirements determine what has to be in the reputation system.

5.1 Requirements

In this section the requirements for design are described. Throughout the previous chapters requirements for designing a reputation system for mitigating botnets have been identified by literature research and expert knowledge. This section joins these requirements and describes them.

The requirements originate from different sources in previous chapters. In the conclusions of chapters 2, 3 and 4 requirements have been identified. They are shown as **(Req X)**, with x representing the numbered requirement in Table 3, below. This table has three categories.

Requirements can originate from three categories (Table 3):

1. Requirements from reputation and reputation systems (Chapters 3 and 4)
2. Requirements from institutional settings (Chapter 2)
3. Requirements from technical possibilities (Chapter 2)

These three categories correspond to the table below. The table below shows the requirements which have been identified throughout the previous chapters. In the table an “X” is placed after each requirement in one or multiple columns representing one of the categories above. The X” shows that the requirement belongs to that category and therefore is identified in the section describing that category. The requirements are not hierarchical, meaning that requirement 4 does not have to be more important than requirement 5 or 15. They are ordered so they correspond with the text below the table which explains what these requirements mean for the designs.

Category 1 shows requirements from chapters about reputation and reputation systems, described in chapters 3 and 4. Chosen is to show the requirements obtained from these two chapters in one category since most of them have been identified in both: the lists would therefore be very similar. In the previous chapter, section 4.4 the differences between the requirements in chapter 3 and 4 are already shown. It can be argued that some requirements are theoretical and in practice would be less existing. Vice versa is also possible; from practice it could be learned that good reputation systems have something in common which is not

described in theory. For the designs it matters less is the requirements are obtained from theory or practice, as long as the requirement applicable to the design.

Categories 2 and 3 originate from the knowledge in chapter 2. This chapter describes the current situation regarding the stakeholders, incentives, botnets, spam, mitigation, detection and initiatives to map botnets and their activity. There is information about stakeholder', incentives, and the ISP market. This describes the institutional settings (category 2 in Table 3). On the other side technical information about botnets, spam and measurement efforts describes the technical landscape (category 3). Although the institutional and technical landscape is described in one chapter, for the requirements it is separated in two categories.

In the table below the requirements are shown. They are numbered. Throughout conclusions in previous chapters, these numbers are also shown. The numbers in the conclusions of chapters 2, 3 and 4 show the where requirements are obtained from and how they relate to the numbers in the table.

Table 3 Requirements

A REPUTATION SYSTEM SHOULD	REPUTATION & REPUTATON SYSTEMS	INSTITUTIONAL SETTINGS	TECHNICAL POSSIBLITIES
1. DECREASE THE INFORMATION ASYMMETRY IN THE MARKET	X	X	
2. HAVE A POSITIVE EFFECT ON BOTNET MITIGATION		X	
3. REALLIGN INCENTIVES FOR ISPS		X	
4. CORRESPOND TO THE RIGHT CONTEXT	X		
5. ONLY GIVE A REPUTATION TO LONG LIVING ENTITIES	X		
6. TAKE INTO ACCOUNT PREVIOUS BEHAVIOR	X		
7. SHOW EXPECTED FUTURE BEHAVIOR	X		
8. BE A SYSTEMATIC ALGORITHM	X		
9. BE RESISTANT TO CHEATING FROM INSIDE THE SYSTEM	X		
10. BE RESISTANT TO ATTACKS FROM OUTSIDE OF THE SYSTEM	X		
11. BE WELL-DISTRIBUTED	X		
12. ALWAYS WORK	X		
13. ACCESSIBLE BY THE INTENDED USER	X		
14. USABLE FOR THE INTENDED USER	X		
15. DEAL WITH VOLATILITY OF UNDERLYING METRICS	X		X
16. ASSIGN MORE VALUE TO NEW	X		

	OBSERVATIONS			
17.	**UP-TO-DATE**	X		
18.	**MAINTAIN PRIVACY OF CUSTOMERS OF ISPS**		X	
19.	**BASED ON AUTOMATICALLY EXTRACTED DATA**			X
20.	**BASED ON MEASURABLE CONCEPTS**			X
21.	**A) BE ABLE TO DEAL WITH FALSE POSITIVES**	X		X
	B) RELIABLE	X		X
	C) VALID	X		X
22.	**BE ABLE TO DEAL WITH BOTNET TAKEDOWNS**			X
23.	**ACCOUNT FOR INDUSTRY TRENDS**	X		X
24.	**DIFFERENTIATE BETWEEN THE SIZE OF THE ENTITIES**	X	X	X
25.	**ADAPTABLE**		X	X
26.	**NOT REPRESENT A SNAPSHOT THE SITUATION**		X	X
27.	**NOT BE OPPOSED BY THE INVOLVED STAKEHOLDERS**		X	

The text below explains what these requirements mean for the future designs. As mentioned above the requirements do not have some hierarchical ordering in the table. They are ordered to be in line with the explanation of the table below.

One of the most important requirements for design is that a reputation system has a positive effect on botnet mitigation (2). The reputation system, botnet mitigation should increase. The whole purpose of developing a reputation system is that in the end botnet mitigation by ISPs increases. For this reason should the reputation system have a positive effect on botnet mitigation (especially for underperforming ISPs). The reputation system should re-align the incentives for ISPs to increase their botnet mitigation efforts (3) and it decreases the information asymmetry in the market (1). From the ISP market it became apparent that there is often a mismatch in information, see Ch. 2 for more background info. The end user has difficulties to determine if they are infected with a botnet. ISPs can solve this problem because they can measure botnet activity and mitigate the problem by warning their customers. However, the incentive structure for these ISPs is in such a way that they are discouraged from doing this, resulting in requirements 1,2 & 3. Designing a reputation system should be the solution to this problem. If a reputation system does not satisfy these requirements it loses its purpose and value.

From the theoretical knowledge on reputation and reputation systems, requirements have been formulated. In Ch. 3 it became apparent that reputation and reputation systems have to

correspond to a context (4). The basis features of a reputation system are that they only give a reputation to long living entities (5). With the reputation previous behavior (6) is used to show expected future behavior (7). The expected future behavior means that the reputation should give some information about the performance of the entity in the near future.

A reputation system has to be systematic (8). Given a certain input, an output has to be expected. If the same information is put in the system, the same output should be generated, in other words it should give a consistent output, given a certain input. This means that in an ideal situation the reputation is resistant to cheating and attacks from inside of the system (9) and outside attacks (10).

A reputation system, calculates a reputation. A reputation should be calculated so it corresponds to the actual situation, but it should also be distributed in the right way (11). A reputation can be very well constructed; but if no one knows about it or uses it, it has no purpose or value. The same holds for the functioning of a reputation system. If it does not work for whatever (for example if it isn't accessible or usable) reason it is has no purpose. Therefore a reputation system should always work meaning that the algorithm is running (12), be accessible (13), and it should be usable (14). A reputation system is usable if people are able to understand how to use it.

The current state of research shows that Spam and botnets are very volatile, the activity changes continuously. A reputation system' quality depends (partially) on being smooth. Both from the theory over reputation systems, and the technical knowledge about botnets it can be derived that a reputation system should be able to deal with the underlying volatility of the metrics (15). New information should also be valued higher than older information (16), and the reputation system should be up to date (17).

An issue with a reputation system, for ISPs based on botnet activity, is that ISPs have to maintain the privacy of their customers. A reputation system shows in a way, although very anonymous, information about the customers of ISPs. A reputation system should maintain the anonymity of the end user (18).

From the technical state of research (see par 2.2, 2.3) a few requirements also have been identified. An issue with reputation systems and the relationship to measuring botnet activity is false positives. Metrics indicating infection do not have to occur from botnets, giving false positives. A reputation system should therefore take into account that there are issues with these false positives (21a). Although there are false positives, a reputation system should still be valid and reliable (21bc). Otherwise it cannot be accurate. Data on botnets are automatically measured and extracted. The reputation system should also be based on automated means for this reason (19). In chapter 2 a range of metrics has been identified. Not everything can be measured about botnets, therefore a reputation system should be based on measurable concepts (20). Botnets can be taken down. In the past, in large scale operations, entire botnets have been taken down (i.e. Zeus). In such a situation all the ISPs would have less infected users, a reputation system should be able to deal with these takedowns (22). Botnet and spam data have trends, i.e. the trend for spam is that it is decreasing slowly, a reputation system should therefore correct for the corresponding industry trend (23). ISPs are different is size and number

of subscribers. A reputation system should therefore take these sizes into account and correct for it (24).

New botnets arise, while others decline. New situations arise quite often. The reputation system should be adaptable to account for these changes. A reputation system for ISPs based on botnet activity must therefore be adaptable (25).

There are also two requirements about what effects a reputation system shouldn't do. There is a risk that a reputation depicts a snapshot of a situation. For example an ISP could have a bad week in terms of number of infected users. A reputation system should therefore have a mechanism to avoid it representing a snapshot (26). Participation between the stakeholders (mainly ISPs and ACM) are also important. Therefore in designing a reputation system, the settings should be in such a way that it promotes participation, the system should thus not be opposed (or has to be accepted) by the involved stakeholders (27).

5.2 Design space

The design space describes the different settings which can be used for a design. In this section the information is obtained from the previous chapters. It connects the design options about reputation systems, with the technically possible metrics from section 2.2-2.3 and the institutional settings (par 2.1) into which the technical solution has to function. The design space consists of two things. First the design criteria. Second the values in these design criteria. The design criteria are derived from chapter 3, showing all the dimensions of a reputation system. In chapter 2 current states of research and institutional settings are described. These settings form the possible values in the design criteria.

Table 4 Design Space

DESIGN CRITERIA	POSSIBLE VALUES				
CONTEXT	General	Specific			
INFORMATION SOURCE	Objective	Subjective			
GOVERNANCE MODEL	Central	Self-governance (decentral)	Coalition (decentral)		
GOVERNANCE SUPERVISION	Direct	Indirect	No		
COMPUTATION ENGINE	Simple summation	Bayesian	Discrete	Flow	Belief
INTENDED USER	Self/stake-holders	Consumers	Govern-ment	Competi-tion	
COMMUNICATION	Digital	print	Word-to-mouth	Social media	
CHEATING	Not considered (Lvl1)	Agents hide info, do not lie (lvl2)	Cheating & dealing with cheating		

In chapter 3 a range of design criteria is identified. These criteria are described in the table above. On the left side of the table, the design criterion is mentioned, on the right side of this criterion the possible values are described. For details about these criteria and values, see Ch. 3. These design criteria are derived from theory. The theoretical knowledge about reputation and reputation systems can be used to select a basic architecture of a reputation system. However the specific content also has to be supplied. For example, a reputation system requires input data: in the table can be seen that the information source can be objective or subjective, the specific metrics are not specified. The table only provides a high level design space. The specification of these design criteria is derived from the information about the current situation (Ch. 2) and expert interviews. This in turn shows the relation between the chapters, where the chapters about reputation (systems) provide a general framework, the knowledge about the current situation/current state of research provides the interpretation of these design criteria in the specific context.

In chapter 2 the different metrics are described. Some of these metrics are already used in performance initiatives as Spamrankings. These metrics can be used for the designs.

5.2.1 Context

The values for the context range from general to specific. It is dependent on the purpose of the reputation system. In designing a reputation system for ISPs, there are some factors that are affected by the context. Should it be a general reputation about botnets? Or should it be specifically for a type of botnet. Also should a reputation system only be used for assigning reputations to ISPs, or should it also be able to assign reputations to other intermediaries?

A very specific context would only be able to assign a reputation to ISPs, about a specific set of botnets. On the other hand, if the reputation would be for all intermediaries, it would be very general and would have less value as a reputation system for ISPs. The context thus sets the scene for some other design space criteria.

5.2.2 Information sources

Information sources can either be objective or subjective. In section 3.1.1-2 is shown that information is either based on machine feedback/automated feedback or information is based on human feedback (i.e. questionnaires).

The current state of research in section 2.3 shows what kind of metrics there are. These metrics are described below. A limitation of this research is that there are also false negatives. Infections which cannot be measured, as there is no measurement scale for it or the measurements are incomplete and missed an infected machine. There is no way of determining how big the amount of the false negatives are, as this is unknown information.

Metrics based on machine feedback

These objective metrics often measure some kind of output. The metrics show objective information about botnet activity without looking at efforts. The list below shows what kind of metrics there are to indicate botnet activity. In other words, the list below shows what is

possible to measure the output of ISP performances about botnet mitigation. The list a short overview over what kind of metrics available.

- Number of infected machines per subscriber (IP addresses per quarter) for Spam and other data sources as DShield or sinkholes
- Number of infected machines per subscriber (Daily averages over each) for Spam and other data sources as DShield or sinkholes
- Total spam/botnet volume
- Total number of infected machines
- Duration of infection
- Frequency of infection
- Botted systems
- Botted IP addresses
- Botted subscribers
- Type of infected device
- ISP Characterization
    - Nr of subscribers
    - Market share
    - Type of ISP (Cable/DSL)
    - Revenue per subscriber
    - Bandwidth

DShield, sinkholes and spamtraps operate on the internet, outside of the ISP's networks. The metrics are strong indicators of an IP being part of a botnet. ISPs have even more information. An ISP knows all the traffic of their users. They could identify infected machines on a more accurate level.

Measuring efforts based on human feedback

In the designs some decisions have to be made about measuring output or measuring effort. From the ISP market it becomes clear that ISPs are working on the problem of botnets (Ch.2). They could be scored based on the above metrics about how well they are doing on metrics as Spam count and volume. ISPs could also be ranked based on how much effort they put in into this problem. Measuring effort is subjective because it would largely be based on human feedback (See section 3.1.2). To quantify mitigation effort questionnaires could be used. The use of questionnaires makes quantifying effort subjective. In chapter 2 some (mitigation) efforts already have been described, these are:

- **Customer awareness campaigns:** almost all ISPs have such campaigns
- **Providing security solutions to their customers:** many ISPs provide for free or reduced pricing for security software as anti-virus software, or firewalls
- **Active participation**: in anti-botnet initiatives as Abuse hub
- **Warn customers**: either by mail or with automatic solutions as walled gardens

- **Improving the quality of routers for their customers**
- **ISO27001 certification**

Where the previous section (objective metrics) shows outputs, these efforts of ISPs could also be used as effort measures. The factors above only measure the output of these efforts. Attempts could also be made to quantify these efforts. ISPs could be surveyed on what kind of mitigation efforts they implement, this could be checked by how active their abuse handling is, how well they deal with incidents, how fast ISPs react to external complaints to their abuse email address. Although they do not give a complete view of the ISPs efforts, such indicators do give an indication about the ISPs mitigation efforts.

Efforts or outputs?

The consideration here is what should be measured. On the one hand what in the end matters is botnet activity? The goal is get this activity down. It can therefore be argued that the results of efforts, the outputs are most important. On the other hand an ISP can also have bad luck and although it puts in more mitigation efforts than other ISPs, they could still have a lower output than an ISP which puts in much less effort.

The question here is, is this fair? In selecting the types of information to use for design, this is an issue to take into account. The subjective and objective information sources are not mutually exclusive so it would also be possible to select a mix.

5.2.3 Governance model

In the design, some stakeholders have to execute the design. Stakeholders have different relationships to the problem, different options, different levels of power related to the ISP market and different levels of knowledge. In chapter 2, some attention was already given to possible stakeholders in the governance. The list below some stakeholders are described which can contribute to the problem:

- Academics as TU Delft
- Internet Service Provider (association)
- Governmental agencies
- Security vendors, other IT companies with knowledge or data.
- Consumer organizations

The different relationships to the problem will influence the technical design of the reputation system, but also the way the reputation system will be governed. Different stakeholders have different goals, but also different powers (see section 2.1 or appendix A).

Can a stakeholder govern the initiative on its own (as a central authority), as a market initiative (self-governance) or as a coalition of different stakeholders? The stakeholder that starts the initiative is limited by their situation related to the ISP market.

5.2.4 Governance supervision

Depending on the type of governance model, there is also supervision on the governance. Be it a legal context into which the governance have to apply to, or a direct supervision. The type of supervision on the governance is dependent on the governance model and settings.

5.2.5 Computation engine

In chapter 3 different types of computation engines have been identified. These are: Summation/averaging, Bayesian, Discrete, Flow and Belief. These are computation types based on literature. Chapter 2 already describes some ranking initiatives for Spam. In Chapter 4 examples of working reputation systems are described. These initiatives from chapter 2 and 4 use different types of computation engines.

The next chapter will use some of these initiatives as a red line for design. The core idea of an already working reputation system will be adapted to the problem of botnet mitigation and ISPs. A proven and existing reputation system's computation engine can be used as a start for determining the computation engine. This does not mean that this computation engine is the only solution and other computation engines wouldn't work. In section 3.2.2 benefits and issues with these computation engines have been identified. The table below gives an overview over these benefits and issues. The values in this table are derived from section 3.2.2.

Table 5 Comparison of computation engines

	SUMMATION	BAYESIAN	DISCRETE	BELIEF	FLOW
ACCURACY	-	+	+	++	0
UNDERSTANDABLE	++	-	+	--	-
COMPUTATIONAL	++	+	-	0	+

Although in an existing initiative a computation engine is already proven to work, its performance can be evaluated for the context of design. If the context of the reputation system fits the engine of the existing initiative, this engine would be a good engine to start the design with. If not, some adaptions could be made to integrate the engine with the content of the table above.

5.2.6 Intended users

The intended users for a reputation system based on botnet mitigation can be a range of the following stakeholders:

- **ISPs themselves:** i.e. for benchmarking and marketing their reputation if they perform well.
- **Governmental agencies:** How well are ISPs in general performing, which are doing well and which aren't.
- **Businesses:** giving transparency, the reputation could help to select with which ISP they should do business with.

- **Uninformed end user:** giving more information over their ISP. They are generally the stakeholder with the lowest knowledge about botnets. A reputation system could increase their awareness.

5.2.7 Communication

The communication of the reputation system would depend on the intended users. How can they be reached? Many could be reached digitally. I.e. in the form of a webpage, social media etc. The problem here is not how to develop a solution for communication, but how to get people to look at the solution. In other words the challenge is not to develop a platform where people could see reputation rankings for ISPs, but how to get people to look at the rankings.

This would depend on the stakeholder taking the initiatives for development, the involved stakeholders in the process altogether and the users to communicate to.

5.2.8 Cheating

The above dimensions and the content together determines how susceptible a reputation system would be for cheating. There are three ways a model could look at cheating (as identified in chapter 3):

1. Not considered (Level1)
2. Agents hide info, do not lie (level2)
3. Cheating & mitigating cheating (level 3)

The differences between the three levels are thus basically to not consider cheating (1); Acknowledge cheating but assume it averages itself out in the end (2); and (3) assume there is cheating and use anti-measures.

ISPs have methods to misrepresent data. There are a few categories identified from literature which ISPs could use and can be classified as some kind cheating effort.

- Filtering traffic: i.e. by port blocking. An ISP could misrepresent the number of infected machines by filtering traffic. An example is blocking port 25. Port 25 blocking would reduce the amount of outbound spam (van Eeten, Asghari, Bauer, & Tabatabaie, 2011). In the short term metrics about spam traffic would show decreases in the number infected machines with spam. Filtering techniques have in common that they treat the symptoms of botnets, instead of mitigating the bots at the end users machines (Quarterman, Linden, Tang, Lee, & Whinston, 2013). By filtering is appears that the number of infected machines becomes lower. Actually the number of infected machines is the same (assuming that all other things have remained equal).
- Increasing IP release speeds: A consumer usually does not have a fixed IP, this IP is dynamic. ISPs can, but do not change customers IPs often. Since measurements are based on IP level, changing the IP release speed would increase the complexity of measurements (van Eeten M. , Bauer, Asghari, Tabataie, & Rand, 2010). If IP addresses change often, an infected machine would have many IPs. In observations this IP might be represented by many IPs, while it is only infected once. The total number of unique IPs is dependent on these lease times. On the other hand, by increasing the lease times,

the duration and frequency of infection cannot be measured, since it is unknown which IPs a specific infected machine would have.

- Using NAT: In such a situation an ISP would put many of their customers in a subnet. These customers all would have the same outbound IP address. If many users are infected, they would be represented by one IP address. Such an ISP would increase its rankings based on the number of unique infected IPs they host (van Eeten, Asghari, Bauer, & Tabatabaie, 2011). However, since there are many more machines represented by one IP, the volume of infection per IP would go up drastically, indication of using NAT. In practice not many ISPs use this NAT, but it could be used to decrease the number of infected machines.

The information sources, stakeholders (incentives) and the type of reputation system are main factors that contribute to how much cheating is possible. If a design assumes that there is cheating and mitigation of cheating weak areas in the solutions should be determined and anti-measures could be taken.

An ISP can misrepresent infections (see above). However if they are really mitigating botnet activity their infection rates would go down over multiple metrics, and remain lower. Where misrepresenting infections would give results only for some metrics and over time it becomes more difficult to overcome this. An example would be blocking of port 25. If this is blocked, suddenly no spam traffic can be measured. On other metrics an ISP would still show infections, so other botnets are still generating traffic. This shows the difference between misrepresenting infections and mitigating infections. Misrepresenting would see sudden shocks (i.e. suddenly no spam at all), mitigation would be over time reductions in infections on multiple metrics.

If an ISP is really working in on the problem of botnet mitigation, both volumes and the number of infected machines should go down, over a range of metrics. Sudden abnormal changes in the infections of an ISP are indications of cheating. For example if an ISP suddenly has no spam anymore, or has abnormal (compared to other ISPs or previous days) volumes, it can be argued that it is likely that they are doing one of the above mentioned things.

On the other hand, if for a range of metrics, over time for an ISP the volumes and infections decrease, they i.e. have implemented a walled garden. If there is abnormal behavior the ISP could be asked if they have changed something.

5.3 Conclusions

With the knowledge from this chapter a reputation system can be designed in the next chapter. This chapter has shown the requirements and the design space. The design space shows all the options, where the requirements show the criteria a reputation system should oblige to. These two criteria are used in the next chapter to design alternatives, and evaluate these alternatives.

Section 5.1 shows the requirements originate from tree different fields: 1) Reputation and reputation systems, 2) ISP market and stakeholders and 3) the technical possibilities.

In section 5.2 the design space is shown. The design space consists of the dimensions identified in chapter 3, with the institutional and technical settings from chapter 2. Together they form a

design space to design a reputation system based on bot infected machines. Designs will be made in chapter 6. In this chapter the design space and requirements will be critical to develop a design. Some of the existing initiatives from chapter 4 are also used as basic framework for a design.

6. DESIGN

This report is not written from a specific point of view. There is no one single stakeholder who wants to create a reputation system for incentivizing botnet mitigation by ISPs. However there are multiple stakeholders with an interest in such a system.. This report is neutral in the way it handles perspectives, this means that not a single design can be made for a specific stakeholder. The issue of botnets is viewed neutral; this means that different options are linked to different stakeholders. These different stakeholders might prefer other designs, or would give more importance to specific design criteria. In Appendix A the interests for different stakeholders are identified.

Two alternatives have been identified. A stakeholder wanting to mitigate the problem of botnets, could start with such an alternative as a general design. Within the design there are still many factors on a more detailed level for them to determine.

Although two alternatives are provided, there is still a strong possibility that other alternatives are also feasible and might even provide with better results in some cases. There are so many different design variables and even more different methods to provide the content to these design variables, it is impossible to specify what the most ideal solution is. There are many other alternatives that will likely still give a good design.

Section 6.1 shows alternative 1. This alternative is loosely based on the framework by S&P. This design assumes cooperation between a set of stakeholders. These stakeholders work together to develop a reputation system. In section 6.2 alternative 2 is described. In this alternative there is no cooperation between stakeholders, the reputation system is developed by a stakeholder alone. In the two design there are therefore large differences. Alternative 1 is a much more comprehensive design, more factors are included. The way the reputation is calculated is much more precise, but on the other hand much more difficult to develop and maintain. For this reason it will require much more resources. This is more likely to be possible if different stakeholders work together to develop this alternative.

This is the core difference with alternative 2. In alternative 2 there is a single stakeholder, therefore less resources as funds or knowledge. A much simpler alternative is therefore for such a stakeholder much more feasible to develop.

Normally in providing designs there little differences between them, as it will be possible to determine effects of changing one design criteria for the entire design. In this chase it was chosen to provide two different designs for two possible scenarios, with the main difference in the scenarios being the available resources. There is more than one difference between these designs. They are likely possibilities given the complexity of measuring botnet activity together with the complexity resulting from the different stakeholders. Previous chapters did not put much research into costs and efforts. This is a limitation of the research, as it is required in development of a reputation system. The logic of the designs is based on this effort. Alternative 1 is a design which will require much cost and effort, while alternative 2 is much more resource friendly.

6.1 Alternative 1: S&P framework

In alternative 1 a design is set up to calculate a reputation for an ISP based on the S&P Framework. Obviously the framework currently focusses on financial aspects, instead of botnets. The current variables in the framework can therefore not be used, but the two tiered methodology as S&P provides is a good start for design.

The framework (by S&P) is a good framework to start with because:

- Works well in practice
- Adaptable to new situations (as other botnets or new datasets)
- The different views of stakeholders can be incorporated
- Can be turned into an automatic process
- The perspective of assigning ratings is similar to this situation
- The end result (the value) is easy to understand

Some of the arguments to be made in favor of using this initiative as a red line in designing a reputation system for ISPs are mentioned above. In section 4.3 it has become clear that such a system works well in practice. They are usually accurate, smooth, assign weights in the right way and relatively resistant to attacks. The only issue is the problem of wrong incentives at rating agencies, which contributed to the financial crisis of `07-`08. This is a specific incident, which is unrelated to adapting the framework for ISPs. ISPs operate in an already much regulated market (i.e. by the ACM), were rating agencies have (had) more freedom. The freedom rating agencies had gave the wrong incentives as: extraordinary profits by assigning as many positive ratings as possible. Since ISPs are strictly regulated, it can be assumed that these incentives do not occur when using this framework for ISPs. Altogether it can be concluded that the framework works well in practice.

The framework works well in practice, in its basis could be seen as a very general framework. Essentially there are initially some indicators for a base rating. This base rating is in a second tier changed; a lower or higher reputation is given according to some specific variables. This makes the framework very adaptable. In a new situation, for example a new botnet arises, or one is taken down, there can easily be a new indicator added or removed.

Because of its adaptability **(Req 25)**, it also incorporates different views of stakeholders. This is important since this report has a neutral stance, and incorporating different views is therefore important. Different stakeholders have different views, being adaptable is a bonus, because multiple views could be used in the scores. Another benefit of the framework is that a new metric can easily be incorporated.

The differences between how the framework is currently used and it could be used occur from the way S&P works. As identified in chapter 5, S&P is paid to assign a rating. In a situation where such a framework would be used to give a reputation to an ISP, it is unlikely that they want to receive a reputation, and want to pay for this. It is more likely that they are assigned one ‘involuntarily ’, therefore the perspective of assigning a reputation differs slightly.

Finally the end result of the S&P framework is a score as AAA. With a corresponding text about the meaning of this three letter score would give a brought group of stakeholders the knowledge to interpret it. For uninformed people, it is already something they know of. In the news often ratings as AAA are mentioned about banks or countries **(Req 14)**. People have unknowingly an idea what the score stands for. For the stakeholders interested in how such a score is assigned, the information about the framework could be made available. If they have a low score and see the indicators in the framework, they can connect the indicators to the score and form an idea on how it happened. For example ISPs could determine this way how to improve their score.

To conclude, the alternative using the S&P framework is a good start for a design since it is well working method, which can be adapted to the situation and to stakeholder needs.

Table 4 shows the design space. In this design space there are several values to choose from. If the framework S&P uses would be followed exactly, the design values would be:

- Context: Specific
- Information source: Objective & Subjective
- Governance model: Central
- Governance supervision: Indirect
- Computation engine: Simple Summation & Discrete
- Intended user: Self and stakeholders
- Communication: In reading and print, social media
- Cheating: Agents cheat, there are mechanisms to deal with this

In the next subsections these dimensions are given a content for this alternative. There are three subsections: a technical design, an institutional design and a process design. These sections respectively explain how the system works (technical), what the governance model and rules for the system are (institutional) and what the roadmap is to develop the technical and institutional design (process).

Scenario

The specific content to the design will eventually depend on the initiative taking stakeholder. Since this report looks at this problem from a neutral perspective, thus without a clear problem owner. The next section gives an advice. This advice is based on a setting of stakeholders.

In this scenario, there is a stakeholder with the goal to mitigate botnets by ISPs. The stakeholder also determined that a reputation system can contribute to botnet mitigation. A reputation system in practice is more than a value corresponding to a set of underlying values, it is also accepted into the society. The stakeholder knows this and for this reason cooperation from other stakeholders is required. The stakeholder wants an automatic system, which is first based on metrics about actual infected machines per ISP. The stakeholder wants to measure results over efforts. Efforts to prevent cheating is also very important for the stakeholder.

6.1.1 Technical design for alternative 1

The framework S&P uses is shown in Figure 10. The same type of computation engine in as with S&P is used since it can achieve a good performance and is understandable. The methodology from S&P and the output is applied in a similar way. The downside is that it requires more computational power. Given the scenario of multiple stakeholders working together, enough computational power is assumed. Another benefit of the engine is that the use of categories helps to overcome volatility issues in metrics (this is explained below). The methodology can generally be divided into two tiers. First, a basis rating, they refer to this as the anchor. For the anchor a set of company and industry values are used. These values refer to the most important metrics about a company. With these initial, most important values the anchor is created. The anchor can in a second tier be modified. S&P uses some modifiers to rate a company higher or lower. ISPs exist over longer time periods, therefore they can be seen as long living entities **(Req 5)**

Figure 12 shows how systematically the rating for an ISP could be determined **(Req 8)**. Changing the framework, as shown above, would set it in the right context **(Req 4)**. Essentially, for a design this system could also be used. In the first tier, the most important metrics should be used: which these are would depend on the actual stakeholder (coalition) who is implementing this solution. The first tier uses core metrics to setup a base rating or anchor. In the second tier, with modifiers the value of the base rating could be in/decreased, depending on scores. Using these modifiers is also a good method to incorporate other views of involved stakeholders **(Req 27)**. To satisfy their requests other modifiers could be used. This way they have influence on the ratings, but they are not incorporated as core metrics. The end result of the technical design would be a rating, as S&P uses from AAA (representing very good ISPs) to CCC- (representing very bad performing ISPs).

General technical architecture for alternative 1

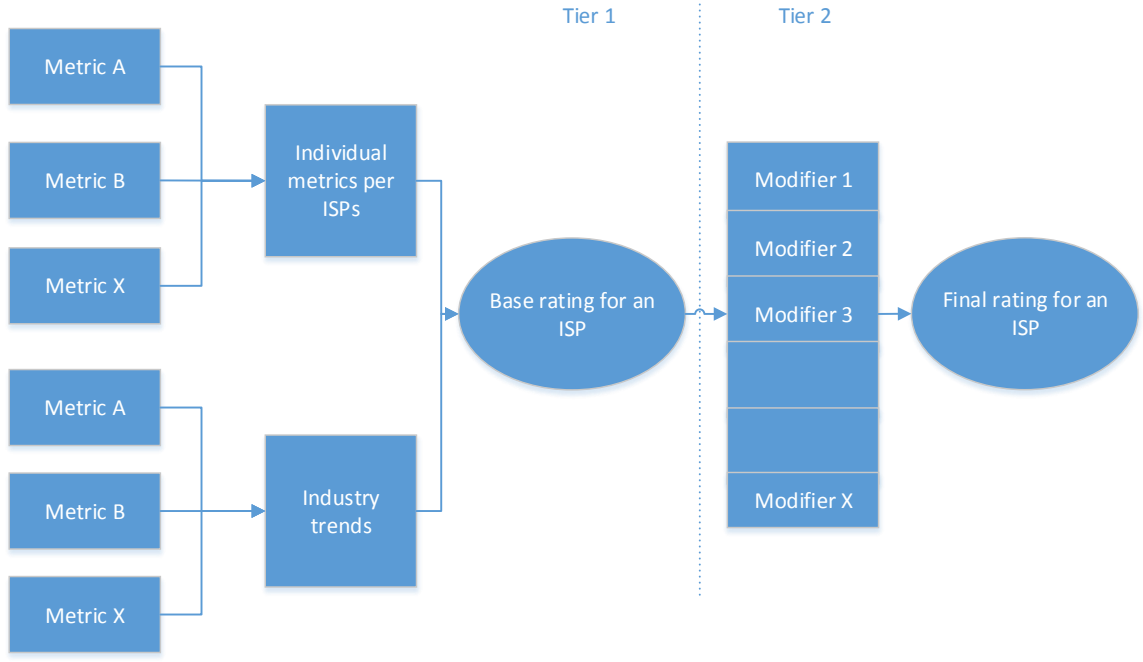

Figure 12 Technical architecture for alternative 1

Tier 1; Metrics

In this tier the base rating is compiled. The metrics used for this stage should be seen as the core metrics. These metrics have to be compared and merged to form such a base rating. Although these metrics should be seen as core metrics, they do not have to be of equal importance. This means that these metrics should be assigned to an importance factor. For example a weight which determines how important a metric is compared to the other metrics. The input metrics will not have the same scale of measurement, therefore there should also be a mechanism to make the different metrics integratable. For example metric A could represent a probability, where metric B might represent an absolute value. They have to form a base rating together, so they should be made comparable in some way to integrate them.

To deal with the problem of false positives, data sources could be compared to determine overlap. Daily values solves the problem of lease times, since most of the IPs do not change within a day. Sinkhole data is in chapter 2 identified to be free from false positives, so this is not an issue. Using multiple data sources reduces the risk of false negatives, it would also make a system therefore more reliable and valid **(Req 21 a,b,c,)**. One issue which is more difficult to account for is mitigation of symptoms instead of removing malware. ISPs can block for example different ports used by malware. This influences measurements drastically. If specific metrics decrease in numbers drastically for one ISP, it is indicated that the ISP has used some sort of filtering/port blocking method. In such a situation other ISPs or for other metrics (not using this port) the situation would be the same. Such behavior, as drastically seeing one metric fall in numbers would indicate such behavior. If such behavior is indicated the ISP could be contacted to verify.

In this scenario as metrics values about botnets should be used as core metrics: Infected IPs/botnet, spam volume, spam count all should be used, corrected for ISP size **(Req 19, 20, 24)**. The industry trends should also use these metrics, so instead of spam volume/ ISP, the industry trend would be the total spam volume **(Req 23)**. This also helps with the issue of takedowns **(Req 22)**

The way these metrics are measured also affects the score itself. Since botnets are constantly changing in size and activity, it makes no sense to calculate metrics over a long time period. On the other hand, if the values continuously change, there would also be a lot of volatility in the values. The behavior over the last weeks or months does give some knowledge over past performance. Van Eeten et al (2010) measured volumes and infected machines as daily averages and over a sum of three months. A reputation system should look at previous behavior, but also focus on current behavior.

To incorporate both previous behavior, but a higher value to new behavior, as core metrics these daily and 3-monthy averages could be merged into one indicator **(Req 6, 16)**. For example an exponential moving average over three months. An exponential moving average (EMA) can be trained to weigh new information heavier and value older information less **(Req 16)**. The EMA is recursive, it uses yesterday's moving average and today's value. Both get an importance, to calculate the EMA. The formula below shows the calculation of the EMA (Wikipedia, 2014):

$$EMA\ (t) = a * Y(t) + (1 - a) * EMA\ (t - 1)$$

- EMA = Exponential moving average
- t = Time
- a = weight (between 0 and 1)
- Y(t) = New information for time t

Such an EMA is calculated by multiplying a weight (a) to the new data (Y) and adding it to the old EMA multiplied by 1 minus the weight. Such an average (EMA) could be created for all the core metrics, so that they are all measured in the same way. Using an EMA would also help to mitigate the negative requirement of a reputation being a snapshot of the situation **(Req 26)**

However, in the base rating, metrics have to be made comparable. To be comparable metrics need to be of similar scales. It is unlikely that all metrics already have the same scale. To use them in the base rating, these scales should be made the same, so they are comparable and compatible. A way to do this is to give all the metrics a sort of normalization or standardization. With such a method all the metrics get the same score ranges. Although there are many different normalization techniques, an often used technique is the min-max normalization (Saranya & Manikandan, 2013). With the min-max normalization data is sanitized and rescaled (often to a scale of [0;1]). A downside of this approach is outliers. Outliers increase the minimum and maximum. The normal (non-outliers) values are therefore much closer on a scale. The formula for the min max normalization is:

$$X(i, rescaled) = \frac{X(i) - Min(X)}{Max(X) - Min\ (X)}$$

With:

- X (i) = the new observation
- Min (X) = the minimum value for X
- Max (X) = the maximum value for X

If there are 3 values for X; i.e. ISP A has on average 1 infected machine per subscriber, ISP B has 2 and ISP C has 3. The scores for ISP A, B, C on this metric would be 0, 0.5 and 1 (resp.). Such a rescaling should be done for all metrics. Although there are many other normalization methods appendix B does provide the min max function a good method for normalization. It rescales the values were others do not, but still gives a score range between 0 and 1. I.e. a Z-score could also be used, but this does not have the scale from 0 to 1.

The actual score for all the metrics can be calculated by:

$$\sum weight\ of\ a\ metric(x) * Normalised\ value\ of\ metric\ (x)$$

The actual valuation of the weights would be dependent on the preferences of the problem owner. The impact of the metrics could be used as an indication of the importance of the metric. For example, a botnet attacking the financial system might be considered worse than a botnet sending spam.

The different weights for all the metrics could a total sum of 1. For example if three metrics are used, A, B and C, A is most important, and the other two are of equal importance, the weights could respectively be 0.5, 0.25 and 0.25. This way the total score of all the metrics and weights would be a value between 0 and 1. A lower score would represent a lower base rating. Lower scores indicates less infected machines and lower botnet activity for an ISP.

The importance of a metric has to be determined for the weights. One way to derive such an importance is to "ask" or negotiate with stakeholders (see next sections). Another way is to determine the confidence in the dataset from which the metric is derived, the type of metric and the size of the metric.

The confidence in the dataset is related to a dataset being free from false positives, who measures it, is it known how it is exactly measured. If data is free from false positives, this raises the confidence in an occurrence actually being an infection. Using confidence in determining the weight helps to overcome the issue of false positives, as datasets with false positives are scored lower **(Req 21a)**.

Another factor that could influence the weight of a metric is how nasty the infection is. For example is a machine infected and only sending spam, or is it stealing credit card accounts? The purpose of the botnet is therefore a factor to be taken into account for the weight.

Finally a possible factor to influence the weight of a metric is the size of the metric. Two different scenarios: first a botnet with only a few thousand infections, second a botnet with few million infections. It would not be fair to give those botnets an equal weight in a reputation score, as one represents many more infected machines. On the other hand, the botnet with (relatively) few infections could represent a very nasty botnet, while the one with many machines only sends spam. For this reason both the type of infection and the size of the metric can be important for determining the weight of the metric in the reputation score.

Incorporating all three would help to increase a reputation scores validity, reliability and how it deals with false positives **(Req 21 a, b, c)**.

Tier 1: Base rating

As mentioned above, the metrics have to be merged into a base rating. S&P uses a system of letters, ranging from AAA to D. As mentioned above, a lower score represents a higher rating. These ratings do not have to be followed exactly, but a categorized score as in the finance world would be a good solution. Botnets and Spam are not consistent in their activity. There are times that the botherders are not using the botted systems, or the botted systems could be turned off. The differences in change of activity would give continuously different ratings. Using categorized scores as with AAA – D would mitigate this issue for a part, because as long as an ISP falls into a specific range, they would get the same rating. For this reason a categorized score should be used.

Using the proposed calculation method score between 0 and 1, a conversion table should be made here: the table below gives such an example (please note that the values in the table only serve as example and actual values will have to be based on testing, metric selection and preferences of stakeholders). Since lower amounts of botnet activity are preferred, a lower

score is preferred over a higher score; correspondingly lower scores in the table result in higher ratings.

Table 6 Example comparison table

BASE RATING	FROM >	TO
AAA	0.1	0
AAA-	0.2	0.1
….	…	…
CCC-	0.9	0.8
D	1	0.9

In the next tier, tier 2, this score can easily be adapted, because the modifiers simply increase the categorized scores up or down a notches.

Tier 2: Modifiers

After the base rating, the rating can incrementally increase or decrease depending on the modifiers. These modifiers have less importance than the core metrics in tier 1. The specific metrics to be used as modifiers have to depend on the different stakeholders who are implementing the system. These modifiers have two functions. First, the modifiers are used to tweak the performance of the rating system. These modifiers incorporate indicators which are not used for the base rating, but do have importance. Using modifiers flexibility to adhere to the requirements is created. A requirement which cannot be met using the core metrics can be satisfied by incorporating a specific modifier for it. Second, these modifiers can be used to incorporate demands other stakeholders might have. For example: the actor who is implementing this system wants to satisfy these stakeholders by using a specific metric or value, but determines a metric to be of less importance. This way the outside stakeholder's criterion is incorporated, but not as a core metric.

The base metrics are based on volumes and counts of botnets, in the metrics other variables could also be used. Some stakeholders, i.e. an ISP, might also want to incorporate mitigation efforts. There are also differences between ISPs, i.e. in size, but also in policies. With these modifiers provisions can be made to incorporate these differences. The following factors give some indications on how what kind of topic could be used as modifiers:

- **Capacity of the infected machines:** can be measured as average bandwidth. The idea is that higher bandwidth indicates the possibility for more botnet traffic.
- **Mitigation efforts:** Difficult to automatically measure. A survey with ISPs could be done where they indicate what kind of efforts they undertake to mitigate botnets. Obviously this cannot be done daily. A semiannual survey could be done to indicate the ISP efforts. Although it is labor intensive, this would give ISPs which are working on the problem,

but might have had bad luck the last months an increased grade for their efforts. Another methodology to indicate efforts could be to look at the behavior of an ISP over all the metrics. If all decrease, while it remains the same at other ISPs, this indicates that the ISP has increased their mitigation efforts (or the efforts have become more effective). In such a case the ISP could be contacted to verify if they have changed their mitigation methodology. Thus instead of contacting them to see what they are doing, they would be contacted to verify if they are doing something else.

- **Mitigation success of ISPs:** measured in average infection/botnet time and frequency of infection/botnet. If efforts of ISPs to mitigate the problem are working, the infection time and frequency of infections should decrease.
- **Botnet takedowns:** In case of a botnet takedown, one metric would drop drastically. To account for such behavior a provision could be made here to keep the rating smooth.
- **Strategic behavior:** What types of ports are ISPs blocking, i.e. port 25 or 53.
- **Provisions for cheating:** ISPs which are identified to be filtering traffic (see section 5.2) instead of removing the bots could be rated a notches lower. I.e. Google uses a similar method to decrease the ranking of a page which buys services from others to increase its rankings (see section 4.1). Such a provision would reduce the incentives for ISPs to represent the infected machines they host in another way. This should mitigate cheating from inside of the system **(Req 9,10)**

The above mentioned modifiers are indications of concepts to be used as modifiers. Other modifiers, based on data availability of requests from other stakeholders are also possible.

Tier 2; Final Rating

The final rating is basically a summation of all the notches the base rating is increased or decreased with. With the S&P Framework, the credit analyst has the option to increase/decrease the rating with a maximum of two notches. The analyst has to convince a board to do this, thus have valid arguments for its actions. In this design this is not incorporated, because the rating is much more often changed than a corporate credit rating. This would be practically impossible, since botnets are a day to day changing in their volume and numbers. The system has to be automatic. Manually adapting a rating for ISPs would be unfeasible given the rates of changes, so this step is therefore omitted. As a final step the rating should be published and disseminated to the public. For the public it is important that they can access the system, understand what the score means and that it is working **(Req 12, 13, 14)**. The score should show something about the expected near future **(Req 7)**. Over time it would be possible to determine this, by keeping a record about the previous scores and the values in the metrics. So if a score is very positive on day t and on day t+1 it becomes very negative, requirement 7 does not hold. So this can be evaluated regressively.

6.1.2 Institutional design for alternative 1

Following the framework, as by S&P, an ISPs should be rated by an external actor to ISPs. S&P is not affiliated with the companies they would rate. This means that an actor which is independent to ISPs in this system should assign ratings. This actor in turn should be regulated by government, as rating agencies are. In the scenario description it is already described that the initiative taking stakeholder can probably not develop a reputation system on its own. In practice complete independence as S&P have, might be less feasible.

S&P gets paid for many of their services, with these payments they fund their rating system. In determining the actor whom should assign reputations to ISPs this is more difficult. ISPs operate in an already competitive market, with low margins (Eyckelhof, 2014). Therefore they will not pay for a reputation, but they do have knowledge and interests. The business model S&P has is therefore not completely adaptable. The actor, who wants to develop the reputation system, will require funding, knowledge, or needs to have something to gain from the reputation.

An actor which is completely external and independent to this problem is therefore difficult to find, because they probably have less interest in this problem. This indicates that a reputation initiative should be done by parties that have funds and knowledge but also have the best interests possible in creating a system which is accurate and satisfies the requirements.

Given the need to have knowledge, funds and the correct interests, a consortium (in the form of a legal entity) of stakeholders might be more feasible. Such a consortium could consist of ISPs/ ISP association (knowledge), governmental agencies i.e. ACM (similar interests, appendix A), and academia (independence, knowledge), possibly security software companies (knowledge and funding). In this situation the governance model would be a central authority, consisting of multiple stakeholders around ISPs. The authority in turn is bound by laws as privacy laws and the telecom law. For ISPs this might be beneficial since such a rating would help them in their mitigation efforts (Appendix A).

The output of the reputation can be published on a special website, dedicated to ranking ISPs. In this alternative there is cooperation between stakeholders, this cooperation can be used in the dissemination of the reputation. Involved stakeholders, as market regulators could also publish the results on their websites. The initiative could get easier in such a situation.

How to publish the reputation score is very important. The large group of parties working together would have a larger audience and could reach more customers. The cooperation is therefore important in the distribution of the rating **(Req 11).** If the reputation score is published in such a way that it also provides the end users with more knowledge about the fact that ISPs conduct botnet mitigation, the score can be more effective. The uniformed users become more informed about the topic, increasing their information **(Req 1).** Take the example of an end user which does not know a lot about botnets and cybersecurity. In advertisement or media he has seen that ISPs are scored according to how much infections there are in the network. This end user becomes infected, gets a call from his ISP that he is infected. Where previously the ISP had to convince such a user that he is really infected, the user already knows that ISPs are contacting more of their clients in case of infection, so he is more easily convinced. The ISP in turn can spend less on explaining what they are doing. The communication of the reputation is therefore very important, as it not only gives information over how ISPs are doing,

it also could help with creating awareness. Dissemination thus not only contains a rating, but also the message that ISPs are undertaking efforts in mitigation, i.e. by contacting their infected customers. Advertising and media coverage could be used here.

The governance of such a system is similar to an IT governance platform. The reputation is IT platform, which is governed. In this governance several policies and decisions have to be made. Effective IT governance should address three questions (Weil & Ross, 2004). These are 1) *what* decisions need to made; 2) *who* should make these decisions; and 3) *how* these decisions are being made and how they will be monitored.

In such a consortium two types of authority exist. First of all the stakeholders have to determine the day to day management of the consortium, the operating authority. Second the strategic decisions also have to be made. Given the large number of involved stakeholders, a council consisting of representatives of these stakeholders should be created. Not all stakeholders have access to the same data sources, see 5.2.2. First it should be determined what access there is for these stakeholders: what can be measured. The core metrics and modifiers should be selected from this the list of possible metrics. The core metrics and modifiers have to be scaled from a numerical range to an ordinal scale. Therefore weights for metrics should be determined.

Such a council should decide over following concepts. The what, who, and how (Weil & Ross, 2004) and the identified criteria from the design space (section 5.2) result in the following criteria to be decided on:

- Selecting metrics (section 6.1.1)
    - What can be measured?
    - Core metrics
    - Modifiers
- Weights of metrics (section 6.1.1)
    - Weights for the base rating
    - Values for comparison tables
- Data (section 6.1.1; tier 1)
    - Who contributes what data?
    - Ownership
    - Storage
    - Access
    - Retention time
- Operational power (Weil & Ross, 2004)
    - Who is responsible for day to day decisions?
- Who should be in the governance council? (Weil & Ross, 2004)
    - How many members
    - Represented by whom?
    - What kind of legal entity should assign the reputation?
- Rules in the council? (Weil & Ross, 2004)
    - When is consensus reached? With a majority or does everybody have to agree?

    - Decisions for the board vs day to day?
    - How long can a representative be in a council
- Cheating (6.1.1; tier 2)
    - How should be dealt with cheating ISPs

Many of these concepts will have to be determined using negotiations. The next section gives an advice how to derive to this information.

6.1.3 Integrating the technical and institutional designs: the process design

The two previous sections described the institutional and design settings. The specific content in these designs has not been given yet. The two designs are described without a specific stakeholder to execute the design. This means that still a lot of issues are unknown. These unknowns are subject to negotiations and further exploration of the problem. This section gives a roadmap. This roadmap gives different issues to be discussed and decided on.

Much of the content and decisions will have to be based on consensus. For this reason it is more feasible to first select which stakeholders should be included, the governance types and rules and in a later stage starting to develop a reputation system. The other way around would give another result, a system would be designed which focusses on the ideas of the one whom designed it, and other stakeholders might not accept it.

There are four phases in developing this alternative:

1. Determining stakeholders to include
2. Negotiation with stakeholders
3. Developing the alternative
4. Implementation and maintenance of the alternative

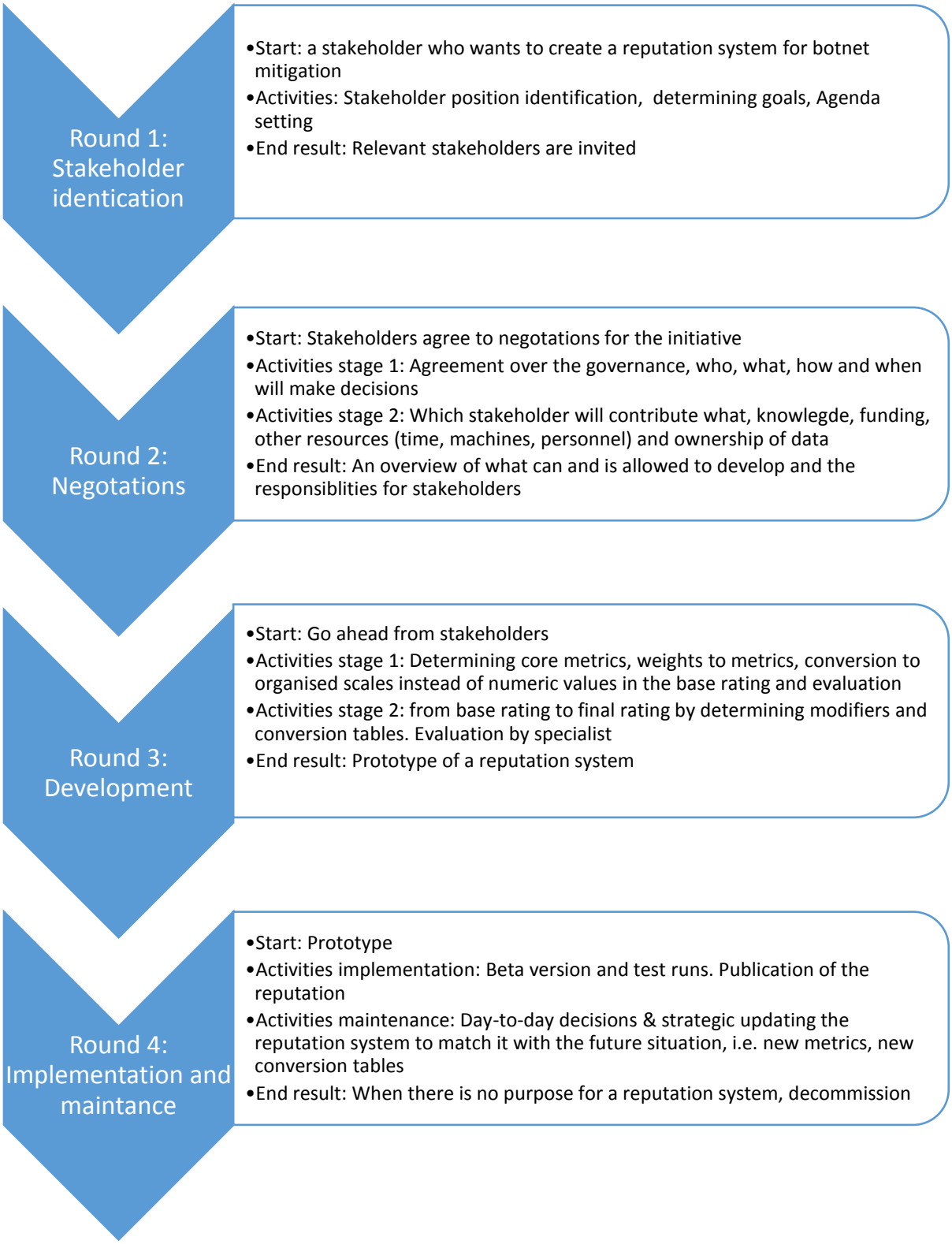

Figure 13 Roadmap phases for development

Round 1

In the first phase the initiative taking stakeholder should determine which other stakeholders there are (see appendix A) as well as their corresponding situations (as interests, financial situation, relationship to the problem of botnets, power). On this basis the initiative taking stakeholder should determine which stakeholders to invite in the next phase, and what topics should be discussed. In the negotiation phases these topics will be discussed. It is thus important for the initiative taker, it is clear for them what the interests of other stakeholders are. This way it can be determined who to invite.

Other stakeholders should be contacted and invited for negotiations. When contacting other stakeholders some initiative can already be taken to determine if their expected situation corresponds with their actual situation. For the other stakeholders it should be clear what the goal of the negotiations are: creating a reputation system for botnet mitigation by ISPs. An initial agenda should be set up by the initiative taker.

The end result of this phase is that the relevant stakeholders are invited for negotiations.

Round 2

The negotiations should be done in multiple stages. First the parties should be in general agreement about the reputation system. This means that it has to be determined how the governance body is organized, see section 6.1.2. Basically, this covers the questions:

- What decisions have to be made?
- Who will make these decisions?
- How will these decisions be made?
- How will these decisions be monitored?

The end result here should be an agreement with the different stakeholders. In this agreement it is decided who will be represented in the governance body. Given the many different stakeholders related to this problem and the need to have funding, independence but also knowledge to develop the system, it is likely that the end result here is that as a governance body: a council consisting of different stakeholders is set up. From the negotiations it should then be clear who would be represented in such a council and by whom. How many members it has, what kind of decisions this council should make etc.

Next to the way the reputation system will be governed, the reputation system should still be developed. The next goal is to determine which stakeholder contributes what? This can be in different form of knowledge, computer power, personnel, funding and also data.

With the data it should be determined how it is stored and how should be dealt with. With data there are questions regarding ownership, retention time, access, storage and security of the data. The data should not show information about specific IPs, but aggregations of IPs amounts per ISP. This way no data about ISP customers can be known to others, and is their privacy maintained **(Req 18).**

These above mentioned criteria can in turn be used to assign metrics and weights to metrics to the framework, see Figure 12 for the framework.

The criteria above should all be negotiated on. In the end of the negotiations it is important to have (De Bruijn, ten Heuvelhof, & in 't Veld, Process Management, 2012):

1. Consensus: stakeholders agree with each other
2. Commitment: The stakeholders have to commit to the initiative as they should also contribute. Therefore consensus and commitment are two different things. In case of no commitment;
3. Tolerance: a stakeholder will not obstruct a decision, but they will also not contribute to it **(Req 27)**

After this phase the initial reputation system can be developed.

Round 3

In the beginning of phase 3 it should be clear which stakeholder has which responsibilities, governance policies should be set and available and usable metrics should be determined. The starting point of development is an overview of which concepts can be measured and can be used: i.e. a metric can technically be measured, but a stakeholder could forbid using it.

The base rating should first be developed. To develop the base rating, the core metrics and weights to these metrics should be selected. With these metrics the base rating can be determined. The base rating converts numerical values from the core metrics to organized scales (the base rating). With expert knowledge, i.e. from stakeholders, it should be determined when behavior is considered good and bad to determine the conversion table from numerical values to a base rating with an organized scale.

In the framework, see Figure 12 Technical architecture for alternative 1, this corresponds to development of a base rating/anchor. This base rating gives a general idea of the underlying performance of the ISP. If a base rating is developed, the modifiers can be selected.

Modifiers use comparison tables. These comparison tables have ranges or values to determine the entities score on the specific modifier. These values should be set. These values basically distinguish the score of an entity in a range of being very bad, to very good. With specialists, i.e. from the stakeholders, it should be determined what kind of behavior is considered bad, and what kind of behavior is considered good.

The end result of this phase, is a prototype of a reputation system.

Round 4

The previous phase ended with a prototype. In this phase the reputation system should be commissioned for use. In this phase the prototype should be up and running for test runs, for a few months. By doing test runs the system can be validated and tweaked, i.e. by changing the conversion tables.

During these test runs, a beta version could be available for ISPs. This way they can be prepared for the eventual implementation of this system. They can already increase their efforts towards mitigation, so their reputation maintains once the system is opened to the public.

After these test runs, the system should be opened to the public and they should be made aware of it **(Req 11)**. After publication the shift goes from implementation to maintenance. The decisions to be made are on an operational level. Over time the system can change; new botnets arise, others demise. During maintenance there will have to be made strategic decisions about new botnets, the weights of metrics, and values in the comparison tables, to keep the system up to date **(Req 17)**. For example, if all ISPs increase their ratings over time, should the scoring criteria be changed to decrease the scores? Such decisions determine the long term success of the reputation system, because if it is not up to date anymore, it loses its purpose.

A reputation system which can be interpreted as valid, reliable and therefore a low amount of false positives, can be seen as an increase in the level of information. More information reduces the information asymmetry. An ISP which performs very badly would not be punished, but increases the likelihood of a visit from the market regulator to discuss this. This is only under the condition that it can be seen as low on false positives. Such a threat should help to enforce the incentives for mitigation **(Req 1, 2, 3)**.

This phase ends with the decommissioning of the reputation system.

6.2 Alternative 2: Borda Count as with Spamrankings

In alternative 1 there was cooperation between stakeholders to gain knowledge and resources. This alternative shows a design which is in contrary to alternative 1 in the way that the resources are much more limited. The previous alternative assumed a situation with cooperation between different stakeholders. In this alternative a single (set of) stakeholder(s) would be developing and implementing such a system. As with the previous initiative, this stakeholder is also outside of the ISP networks. They can only indicate botnet activity based on Spamtraps, sinkhole data and datasets as DShield.

There is a difference between this alternative and the previous alternative. In the previous alternative efforts are incorporated. Another difference between this alternative and alternative 1 is the available resources. The previous example assumes a coalition of stakeholders, with more stakeholders' resources as knowledge, data availability, funding and time might be higher. This alternative would be executed by some other stakeholder than an ISP. An ISP would have access to data from within their networks. Since in this the alternative is developed independently from ISPs, their requirement of also introducing mitigation efforts is less strict. Measurements can only be about outputs instead of efforts. The algorithm would therefore be based on automatically extracted data **(Req 19)**.

There is little to no cooperation and a reputation system is an initiative from a single stakeholder. This means that this stakeholder also has less resources. In the previous alternative different stakeholders had different resources as personnel, funds and knowledge. A single stakeholder will have less than a coalition. The design based on the S&P framework from section

6.1.1 could be used. In terms of resources however, this would be less feasible. Determining and tuning of the conversion tables requires time and knowledge and might for a single stakeholder be less feasible. The S&P framework is also computationally expensive, it would use more processing power, which also has to be available.

Less cooperation means less resources, for this reason a more resource friendly technical design would be more applicable. It is based on the technical design from Spamrankings (see section 2.3)

This stakeholder is outside of the ISP networks and can only use data from outside of these networks. The initiative taking stakeholder is thus not an ISP.

Table 4 shows the design space. In this design space there are several values to choose from. In this alternative the following the design values would be used:

- Context: Specific
- Information source: Objective
- Governance model: Central
- Governance supervision: Indirect
- Computation engine: Simple Summation
- Intended user: ISPs, Governmental organizations and involved end users
- Communication: In reading and print, social media
- Cheating: Agents cheat, there are mechanisms to deal with this

6.2.1 Technical design for alternative 2

In section 2.3 metrics have been defined. In one of the papers, an initiative called Spamrankings has been proposed. In this paper companies are ranked based on their spam volume and spam count ranking from two sources. For each source the volumes and spam rankings are calculated and the final score is a sum of the scores over different data sources.

In this form the system is too limited to be a good reputation system, but the initiative can be expanded to new metrics and ISPs. The metrics of spam volume and spam count could be used per ISP. However, ISPs have different sizes, so these factors should be compensated for the number of connections of the ISP (corrected for ISP size) **(Req 24)**. Another metric should be to compare the volumes and counts for the total volumes and counts **(Req 23)**.

By comparing metrics to the total volume a compensation is made for differences in botnet activity and possible botnet takedowns **(Req 22)**. If a botnet is taken down, the total values should decrease.

Similar metrics could be made for botnets, i.e. from sandbox or sinkhole data. Although these data sources could give a partial view of a botnet, they would still provide a good overview of infected machines.

Let's say for example there are 3 ISPs.

Table 7 Simplified example of design alternative 2

ISP	SPAM COUNT/ SUBSCRIBER	SCORE SPAM COUNT	BOTNET (A) COUNT/ SUBSCRIBER	SCORE BOTNET (A)	TOTAL SCORE
A	0.1	0	0.06	2	2
B	0.2	1	0.01	0	1
C	0.3	2	0.05	1	3

In Table 7 a very simple example of the ranking is given. For all the metrics to be used the score of an ISP on this ranking should be determined. With this score, ISPs can be ranked. In this example there are 3 ISPs. Low values on the metrics is better than high values (since sending out less spam is better than more spam). The ISP with the highest scores on the metrics get for that metric the score of 2 (there are 3 ISPs, the entity has the highest score, and thus gets the highest rank of 1, 3-1=2). The score is systematically calculated by the number of ISPs minus the rank of the ISP on the metric **(Req 8)**. The total score is the sum of all scores and should show information about the expected near term future behavior **(Req 7)**. New metrics can be incorporated in the total score, making the initiative adaptable and easier to keep up to date **(Req 17, 25).**

Possible metrics to use for this system are:

- Number of infected machines per subscriber for spam, Dshield or sinkhole data or botnet data
- Volumes per subscriber for spam or botnet data
- Total spam/botnet volume
- Total number of infected machines
- Duration of infection
- Frequency of infection
- ISP Characterization
    - Nr of subscribers
    - Market share

As similar data sources as with section 6.1.1 are used, the same issues and solutions for false positives can be applied **(Req 21)**. By looking at both ISP size and total values, corrections for differences between ISPs and global trends are made. These metrics are the measurable concepts available **(Req 20)**.

Botnets change constantly and their activity is not constant during each day. For this reason these metrics should not only look at a daily ranking. Measurements over longer time should also be included in the ranking, i.e. by using metrics two times. Also using longer timespans (i.e. by summations from daily averages), should decrease the cheating possibilities (Quarterman, Linden, Tang, Lee, & Whinston, 2013) **(Req 9, 10)**. Once with a scale of a day and once with a longer timescale to account for previous behavior **(Req 6, 26)**. It could also be calculated with an

exponential moving average (see par. 6.2.1), as it incorporates past behavior but prioritizes current behavior **(Req 16)**. Looking at longer timescales helps to account for cheating, over time it is more maintaining the score by cheating. Longer time spans smoothens the metrics to decrease volatility of underlying metrics **(Req 15)**

6.2.2 Institutional design for alternative 2

In this scenario there is little cooperation between stakeholders to develop a reputation system. The assumption is that the initiative taker has the knowledge and data access to measure from outside of a network the above metrics. For the governance the stakeholder can just decide everything for themselves, as long as they oblige by governmental policies and regulations (i.e. about privacy). The reputation system should correspond to the actual underlying scores of the ISP, meaning that the reputation score cannot show a bad score, while in practice the ISP has almost no infected machines in its network. This is to ensure it is not opposed by involved stakeholders **(Req 27).**

Where in the previous alternative, the focus was on what kind of decisions had to be made between stakeholders, the focus is here less on who can make decisions etc. In alternative 1 the institutional design has some characteristics similar to politics, where everything needs to be negotiated, with an end result as an agreed compromised content.

In this situation the focus is more on spreading the results of the reputation system. In the previous alternative the reputation could be communicated by a group of stakeholders with a larger audience. In this alternative the communication has to be done by the stakeholder itself. Publication of the reputation is thus very important **(Req 11)**. News articles, social media and advertisements are important here. The initiative taking party should also try to get contacts with i.e. consumer organizations to communicate the reputation. As an end result some kind of dashboard or platform should be developed to show the reputation. People should easily understand the dashboard, therefore it should be easy to use, accessible and always work (**Req 12, 13, 14).**

6.2.3 Integrating the technical and institutional design: the process design

In the two previous sections the institutional and design settings are described. The specific content in these designs has not been given yet, only an indication of what could be used as content. A part of the design includes also the stakeholder wishes. Since the designs have been made with an open stakeholder, there is some room to maneuver. The stakeholder can choose to develop a reputation system on his or hers data availability or preferences. Figure 14 shows the roadmap for design.

Roadmap

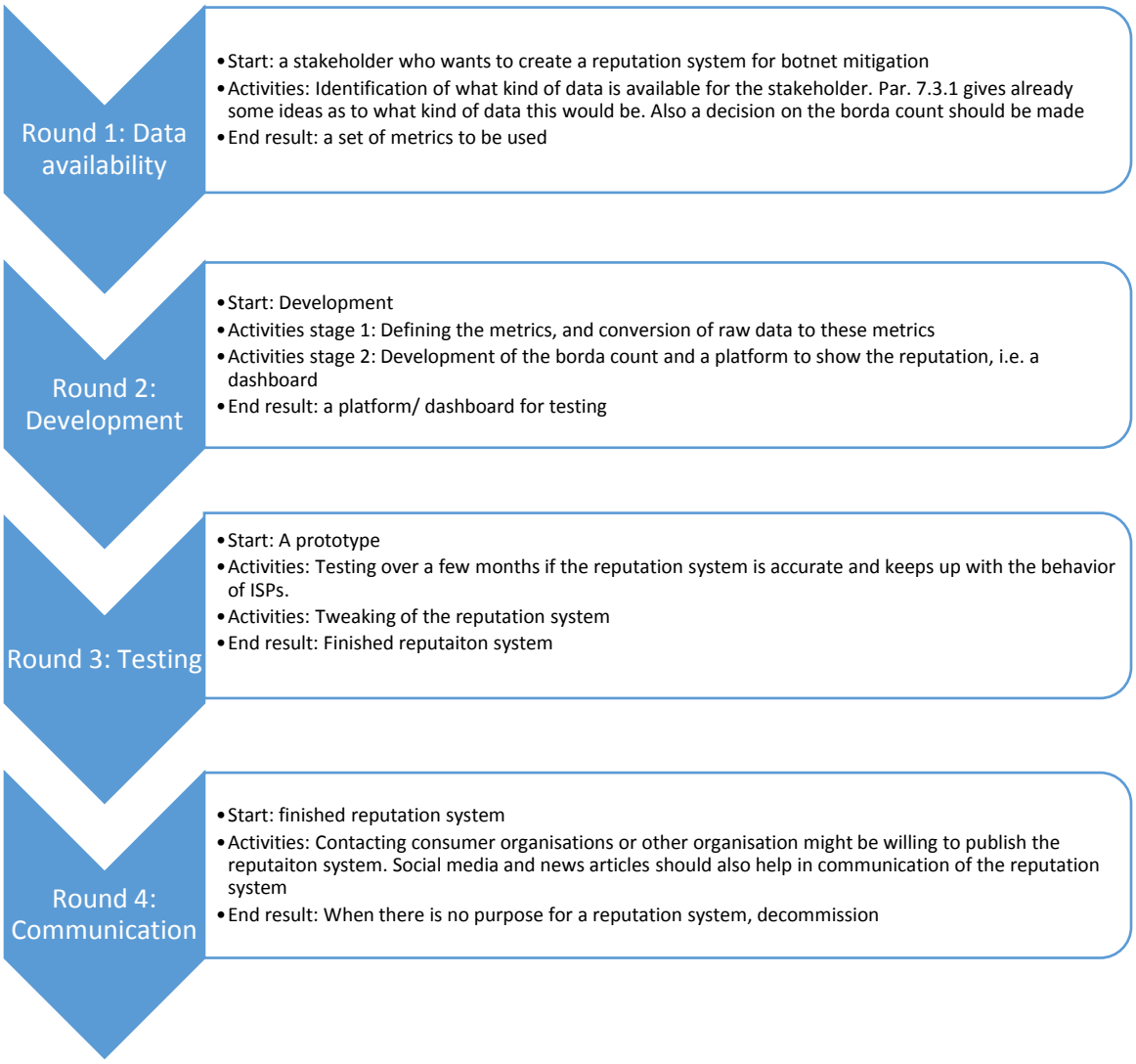

Figure 14 Roadmap for alternative 2

In this integration the focus is more on development, testing and publication. There are four rounds in developing this alternative:

1. Data availability: in this round the stakeholder should determine what kind of data/metrics are available. Also a decision should be made over the type of reputation system, i.e. a borda count.
2. Development: in this phase the data is turned into metrics and the reputation system is developed. The end result of this round is a platform which shows the reputation, to be tested.
3. Testing: during testing the actual accuracy of the reputation systems should be determined. A test version should run next to the metrics to see how well it is performing. Changes could be made in the reputation system to adapt for these issues.

The performance here needs to be tuned so it is valid and reliable **(Req 21 a,b,c).** A reputation system which can be interpreted as valid, reliable and therefore a low amount of false positives, can be seen as an increase in the level of information. More information reduces the information asymmetry. Therefore testing and tuning is very critical. An ISP which performs very badly would not be punished, but increases the likelihood of a visit from the market regulator to discuss this. This is only under the condition that it can be seen as low on false positives. The threat of a visit should help to enforce the incentives for mitigation **(Req 1,2,3)**.

4. Communication: this round is most important. As identified it is good to have a reputation system, but if nobody knows about it, it is useless. For this reason the reputation system should be communicated well **(Req 11)**. Ideally some organizations with a larger consumer base (i.e. a consumer organization) do help here. News articles, social media, and advertisement could also be used for communication.

The next section evaluates the different designs.

6.3 Conclusions

This chapter showed two possible designs for a reputation system. The design space and requirements have been used to describe the design. Each design consists of three parts: a technical design, an institutional design and a design integrating the two. The integrating design, called the process design, shows a roadmap for development.

In section 6.1 a design based on the S&P Framework is proposed. In this case the fictional case of executing such a design with a coalition of stakeholders is envisioned. The S&P framework uses two tiers to assign the rating, first a base score with the core metrics, second a final score. The final score is derived from the base score and modifying indicators (See section 6.1.1).

In section 6.1.2 the institutional design is given. Since the design is envisioned with a coalition of stakeholders, the institutional design focusses on determining and dividing ownership rights and responsibilities. The outcome of such an institutional design would largely be based on negotiations.

The process design of section 6.1.3 has four stages: 1) Determining stakeholders to include, 2) Negotiation with stakeholders, 3) Developing the alternative and 4) Implementation and maintenance of the alternative.

The development (3) and implementation (4) are very important. In this stage the accuracy of the reputation system is determined. To achieve accuracy it should be a reliable, valid system with measures to prevent cheating and minimalize the issue of false positives (See section 6.1.1). If the system is resistant to FP, reliable and valid: the core requirements of realigning incentives and mitigating botnets are more likely to be met, i.e. as it also potentially increases pressure from market regulators (see section 6.2.2 or appendix A). On the assumption that during stages three and four the system is tuned to be accurate the requirements for design are met in this alternative.

In table 8 the alternatives are compared to requirements from ch. 5. For ever alternative an indication is given with regard to how they score on each requirement. There are five possible values:

1. "++" = the requirement is met
2. "+" = the requirement is likely to be met
3. "+/-" = it is unknown if the requirement can be met
4. "-" = the requirement is likely not to be met
5. "- -" = the requirement is not met

The alternatives are not developed yet, therefore scores for requirements are indications. As seen above, there are five possible values. The first value, "++", is given if it is relatively certain that a requirement will be met after development. The second value, "+", is given is a requirement is likely to be met, given enough resources. A difference between one or two plusses is that the first one is definitely possible to be met, while the second one ("+") can be met, but at what cost or level of effort. For this reason a distinction between the two is made. The third value, "+/-", is awarded if it is unknown if a requirement is to be met, or it could go either way (positive or negative). The structure is the same for the negative values, "-"and "- -", a double minus indicates a requirement definitely not to be met, where a single minus shows that a requirement is likely not to be met.

Requirements 4, 5, 6, 8, 12, 13, 14, 16, 17, 18, 19, 20, 22, 23 and 24, 26 for alternative 1 are awarded a "++". Those requirements can be met by taking them into account when developing the system. For example, the requirement of being a systematic algorithm, meaning that given a certain input, the same output can be expected, is a matter of programming. So under the assumption that development of an alternative is done correctly those requirements can be met and therefore are given a double plus. Other requirements are also based on the effects and quality of the reputation and how it well it is implemented into society. The two tiered methodology also makes the initiative adaptable, so requirement 15 is awarded a "++".

Requirement 21 is important. The alternative uses mechanisms to deal with false positives (i.e. by implementing confidences in data sources). However it is to some extent impossible to know for some data if it is botnet data and if there actually is an infection. Although mechanisms are used to deal with false positives, it is unknown if they will work in practice. There is a good probability that it would, but it is something that should appear in practice. For this reason it is unknown and it is awarded a "+/-". If it is possible in practice it will be a difficult procedure, so it would require much knowledge and resources to develop it.

Reliable and valid (21 b & c) are likely to be met, but it is a costly procedure. For every metric its importance should be determined, requiring many resources. Also the conversion from numerical values to categorized scores are likely to provide a reliable and valid algorithm, but again resource costly. So it is possible, but at what cost.

Past behaviour can be taken into account by using data from the past. However a reputation system should also show expected future behaviour. The odds are that a reputation system would do this to some extent, as the number of infections do not suddenly all would disappear. However there can be new infections or new mitigations. For the near future it is possible that

the reputation would show the expected behaviour, but on a short timescale. For this reason requirement 7 is given a “+”.

In the design provisions for cheating have been included. With such provisions it could be possible to determine cheating. For this reason the requirements 9 is given a “+”. Attacks from outside of the system would be someone trying to force the system offline. Technically this could be possible, it is difficult to determine if the alternative is resistant to this. It has to appear in practice therefore it is unknown (“+/-“).

Requirement 11 is awarded a double plus as the cooperation between stakeholders would be able to distribute a rating to large audience. The coalition also provides a larger bases for consensus, therefore requirement 27 can also be met.

Finally, the first 3 requirements. Such requirements would to some extent be based on the results for other requirements. Theoretically, a reputation system could give incentives to underperforming ISPs to improve botnet mitigation. Extra factors as some pressure from market regulators for underperforming ISPs would help to reinforce requirement 2 and 3. These requirements are related to many other requirements, of which it is unknown if they will be met. Therefore it is likely that in practice requirements 1,2 and 3 will be met, but again at what cost or efforts. This is a limitation of the research, as costs and efforts to develop a reputation system are not yet researched.

For requirement 1 there is another issue. A reputation system would certainly help to decrease the information asymmetry. It is a source of information, and per definition does information help to decrease the asymmetry. How much does the information asymmetry decrease; this is unknown, therefore the requirement is scored “+/-“.

Table 8 Alternatives compared to requirements

A REPUTATION SYSTEM SHOULD	ALTERNATIVE 1	ALTERNATIVE 2
1. DECREASE THE INFORMATION ASYMMETRY IN THE MARKET	+/-	+/-
2. HAVE A POSITIVE EFFECT ON BOTNET MITIGATION	+	+/-
3. REALLIGN INCENTIVES FOR ISPS	+	+/-
4. CORRESPOND TO THE RIGHT CONTEXT	++	++
5. ONLY GIVE A REPUTATION TO LONG LIVING ENTITIES	++	++
6. TAKE INTO ACCOUNT PREVIOUS BEHAVIOR	++	++
7. SHOW EXPECTED FUTURE BEHAVIOR	+	+
8. BE A SYSTEMATIC ALGORITHM	++	++
9. BE RESISTANT TO CHEATING FROM INSIDE THE SYSTEM	+	+
10. BE RESISTANT TO ATTACKS FROM OUTSIDE OF THE SYSTEM	+/-	+/-
11. BE WELL-DISTRIBUTED	++	+/-
12. ALWAYS WORK	++	++
13. ACCESSIBLE FOR THE INTENDED USER	++	++
14. USABLE FOR THE INTENDED USER	++	++
15. DEAL WITH VOLATILITY OF UNDERLYING METRICS	+	+
16. ASSIGN MORE VALUE TO NEW OBSERVATIONS	++	++
17. UP-TO-DATE	++	++
18. MAINTAIN PRIVACY OF CUSTOMERS OF ISPS	+	+
19. BASED ON AUTOMATICALLY EXTRACTED DATA	++	++
20. BASED ON MEASURABLE CONCEPTS	++	++
21. A. BE ABLE TO DEAL WITH FALSE POSITIVES	+/-	-
B. RELIABLE	+	+/-
C. VALID	+	+/-
22. BE ABLE TO DEAL WITH BOTNET TAKEDOWNS	++	++
23. ACCOUNT FOR INDUSTRY TRENDS	++	++
24. DIFFERENTIATE BETWEEN THE SIZE OF THE ENTITIES	++	++
25. ADAPTABLE	++	++
26. NOT REPRESENT A SNAPSHOT THE SITUATION	++	++
27. NOT BE OPPOSED BY THE INVOLVED STAKEHOLDERS	++	+

Section 6.2 shows alternative 2. Alternative 2 is based on a single stakeholder developing such a system. In such a situation resources might be more limited. A less complex system might therefore be more feasible. In alternative 1 the focus was to a large extend on negotiations. In the case of a single stakeholder this would be much less.

The computation engine from alternative 1 requires more resources (see Table 2). The categorized scale requires much more computational steps. Developing the comparison tables is also costly in terms of resources as time and knowledge.

For this reason a decision was made to continue on the Spamrankings initiative. This initiative was introduced in section 2.4. In 6.2.1 the technical design was given. This design shows an expansion on the Spamrankings initiative. Where the initiative originally only used spam sources per company the focus is shifted to multiple sources (also sinkhole data) to ISPs. For the metrics a ranking can be made by determining the score on the ranking for every metric. Using a Borda count the different scores turned into a ranking.

Table 8 shows how initiative 2 scores on the requirements. For many of the requirements it is the same as with alternative 1. Below only is explained on which requirements they differ. For example, many requirements are met in alternative 1, as they are a result of development (i.e. requirement 8). This is the same for alternative 2.

The main elements where alternative 1 and 2 differ is the cooperation and the score calculation. Both use the same metrics, resulting in factors as cheating being the same. The system is much simpler in its calculations and i.e. does not give weights to metrics as with initiative 1. It is therefore more difficult to determine if this initiative is going to be accurate. The table above shows for requirement 21 a,b,c the value “+/- or -”. The previous alternative uses confidence levels to reduce effects of false positives, also it makes a distinction between the purpose of a bot, and the size of the metric. Alternative 2 aggregates evenly over all metrics. Although it is not impossible to create a system that is reliable, valid and resistant to false positives, it would require much resources. Compared to alternative 1 is would perform less there requirements. For this reason they are scored lower than alternative 1.

Another issue with this alternative is to distribute it to all the intended users. How will it be disseminated? Alternative 1 had a cooperation between many stakeholders. For the dissemination, these stakeholders could be used. In this alternative this is not the case, increasing the difficulty to distribute it. For this reason it is scored as a “+/-“. It is not impossible to do it, but unknown if people will be aware of it.

The lack of cooperation between stakeholders might also result in some stakeholders opposing the alternative, resulting in a lower value in the table compared to alternative 1 for alternative 2 on requirement 27. Finally, given the many unknowns in this alternative it impossible to determine if requirements 1, 2 and 3 are met. This is partly the same issue as with alternative 1. For this reason they are scored as a neutral value.

This chapter used the information from chapters 1 to 5 to develop and show two designs. These are not the only possibilities as in practice there will be many more. Both designs are likely to oblige by many of the requirements. Each design also has a roadmap. The information from the

designs together with the roadmap and the design space of chapter 5 can be used to either develop such a design, or to determine other designs for future work.

In the next chapter the conclusions, limitations and recommendations for future work are described. It evaluates all the chapters so far to answer the research questions.

7. CONCLUSIONS, RECOMMENDATIONS AND FUTURE RESEARCH

Previous chapters have discussed the topic of (incentivizing) botnet mitigation by ISPs. A method to incentivize ISPs to increase their mitigation efforts is a reputation system. Chapter 2 identified the current situation by researching the topics of botnets, botnet activity measurement and the institutional settings. In this chapter a research gap was identified.

The situation is so that ISPs can mitigate the problem of botnets, but have the wrong incentives to undertake enough mitigation efforts to mitigate botnets. Increasing ISPs botnet mitigation efforts requires to publish a reputation based on the infected machines they host. There are already initiatives to rank systems to the level of spam they host, however these initiatives are too limited still. They only provide a ranking based on spam, while spam is an indicator of botnets, but certainly not the only one. The initiatives do not specifically focus on ISPs, they either rank autonomous systems or they rank companies.

As there was no literature about a reputation system for ISPs. The topics of reputation and reputation systems have been researched in chapter 3, while chapter 4 shows existing initiatives from other scientific fields. The knowledge from the current situation and reputations is turned into requirements and a design space in chapter 5. Chapter 5 served as the bridge from theory in chapters 1 – 4 to design in chapter 6.

This chapter concludes on these chapters by answering the research questions in section 7.1. Section 7.2 limitations of the research are given. They show what is still unknown after doing the research and give possibilities for further research.

7.1 Conclusions

In chapter 1 the main research question and sub questions were identified. This section answers them in the same order. The main research question for this chapter is:

"How could a reputation system to incentivize botnet mitigation for ISPs be constructed?"

With the corresponding research questions (Rq):

1. What is the current situation regarding:
    a. ISP market? **(Rq 1a)**
    b. Botnet detection? **(Rq 1b)**
    c. Measuring botnet activity? **(Rq 1c)**
2. How should reputation be defined and what are its dimensions? **(Rq 2)**
3. What are reputation systems? **(Rq 3)**
    a. What are the dimensions of reputation systems? **(Rq 3a)**
    b. How do reputation systems work? **(Rq 3b)**
    c. What are the objectives for a reputation system? **(Rq 3c)**
4. How are reputation systems used in practice **(Rq 4)**?
    a. Which are effective? **(Rq 4a)**
    b. Which can be adapted to a reputation system for botnet mitigation? **(Rq 4b)**
5. What are the requirements for a reputation system based on botnet activity? **(Rq 5)**

6. What are the possibilities for designing a reputation system based on botnet activity measurements? **(Rq 6)**
7. What are the challenges in designing a reputation system based on botnet activity measurements? **(Rq 7)**

7.1.1 Conclusions from the ISP market, botnets and botnet mitigation

The ISP market is described in section 2.1. ISPs are part of a wider range of internet intermediaries. Table 1 shows the intermediaries and their purposes. The ISP provides access to the internet. The Dutch ISP market is very competitive and ISPs operate on a low profit margin. There are two categories of ISPs. One type of ISPs offers internet services, but also owns the underlying infrastructure (i.e. KPN or UPC). The other buys bandwidth from KPN. Between internet connections there is also a difference in the way the user accesses internet. The network of KPN is based on the telephone line or fiberglass, were UPC and Ziggo offer internet via the coax cable (TV cable). Many ISPs operate regional, this means that the end user cannot simply choose between all ISPs, they can only choose between the ISPs that do offer their services in the region.

An ISP consists of one to many autonomous systems. An AS is: “a set of routers under a single technical administration, using an interior gateway protocol and common metrics to determine how to route packets within the AS, and using inter-AS Routing protocol to determine how to route packets to other ASes (Rekheter, Li, & Hares, 2006, p. 3)”. Every autonomous system has a number (ASN), a unique integer for identification. These ASN are used for the identification which IP address corresponds with which ISP. An end user is thus connected to one of the ISPs ASes. These ASes together make up the ISP.

ISPs are strictly regulated. They have to oblige by Dutch telecom law. The Authority Consumer and Market is the regulator for ISPs. They ensure that i.e. KPN opens its network for other ISPs to use. So they keep the market competitive. The ACM helped with initiatives as the Abuse hub, to decrease botnet activity.

The current situation regarding the ISP market **(Rq 1a)** can best be described as a very competitive regulated market. The ISP market is forced to be competitive by the ACM, resulting in lower profit margins for ISPs. On the other hand ISPs are expected to do something about the issue of botnets.

Botnets can be used for multiple purposes as: DDoS attacks (Distributed Denial of Service attack), gathering information, phishing scams, identity fraud, political protests, terrorism and finally sending spam. Botnet activity per IP can be indicated in various ways (Section 2.1). The first method to indicate botnet is by measuring spam data. Botnets and spam are related (section 2.1.1-2.1.2). In about 90% of the cases of spam, it is send by a bot. Spam usually originates from botnets. If it is measured that an IP is sending spam, there is about a 90% probability that this spam originates from a bot. Spam can be measured by honeypots. Such a honeypot is set up to explicitly attract spam, i.e. to map spam data.

There are also other data sources which indicate botnet activity. An example is the DShield dataset. In such a dataset log offending IPs from firewalls and Intrusion detection systems

throughout the world are shown. Both spam data and DShield data have false positives. With a false positive the data suggests infection, but in practice there is no infection.

Bot infected machines can also be identified from sinkholes. With a sinkhole an entire view for a specific botnet is given. The data is free from false positives as it only contains infected machines with the specific bot. Sinkhole data is from botnets that are taken over, the command and control structure is taken over. The botnet is therefore deactivated, but the infected machines remain infected (only they do not harm).

Botnet activity can thus be measured from three kinds of sources **(Rq 1b)**: honeypots, existing datasets as DShield and sinkholes. These measurements are from outside of an ISP network. There is a difference between measurements outside of a network of ISPs and inside of ISPs networks. An ISP would be able to identify infected machines even better, since they are within a network of an ISP and can therefore exactly see what kind of requests their customers make.

Data from sinkholes, honeypots, DShield and spamtraps have been used to develop metrics about botnets **(Rq 1c)**. Potential metrics can be (section 2.3): Unique infected sources per subscriber based on various data sources (Spamtraps, sinkholes etc), volumes of infection per data sources, frequency of infection and duration of infection and identifying features about the ISP (Size, market share and bandwidth).

Such metrics are already used to i.e. rank autonomous systems or companies based on spam data. There are no real efforts yet to provide a single ranking for ISPs based on multiple metrics. Van Eeten et al (2010, 2011) have shown performances of different ISPs for different metrics.

7.1.2 Conclusions from reputation and reputation systems

Since no reputation system is yet available for showing infected machines per ISP, the concepts of reputation and reputation systems are researched in chapter 3. From literature it is defined what reputation is and what its dimensions are.

"A reputation is the degree to which one party has confidence in another within the context of a given purpose (Hoffman, Zage, & Nita-Rotaru, 2009, p. 3)*"* **(Rq 2)**

A reputation can generally be classified onto two dimensions **(Rq 2)** (Josang, Ismail, & Boyd, 2007):

1. From general to specific
2. Based on human or machine feedback

Reputation is set in a context. This context drives the level of specificity or generality of a reputation. The degree of confidence is based on information. Such information can be obtained from two types of data sources: Human feedback or machine feedback. The first one is often subjective, while the second is objective (sections 3.1.2.1-2).

Reputation is part of a reputation system. A reputation system is an automated method that collects, distributes and aggregates feedback about a participants' past behavior **(Rq 3)** (section 3.2). A reputation system can be separated into five dimensions **(Rq 3a):** the computation engine, the reputation assigner, the intended users, communication of the reputation system

and cheating. The concepts of reputation and reputation systems merge together into a reputation metric.

The reputation is formed in three stages: 1) data conversion; 2) calculation and 3) communication **(Rq 3b)**. The figure below shows this process. To assess the quality of a reputation system there are four criteria (see section 3.2.1): accuracy, weighting towards current behavior, robustness against attacks and smoothness **(Rq 3c)**.

The overview on how reputation systems work, can be seen as a scientific contribution of this research. The combination and integration of the views of the different authors give an overview about reputation and reputation system. It can be interpreted as design criteria for reputation systems.

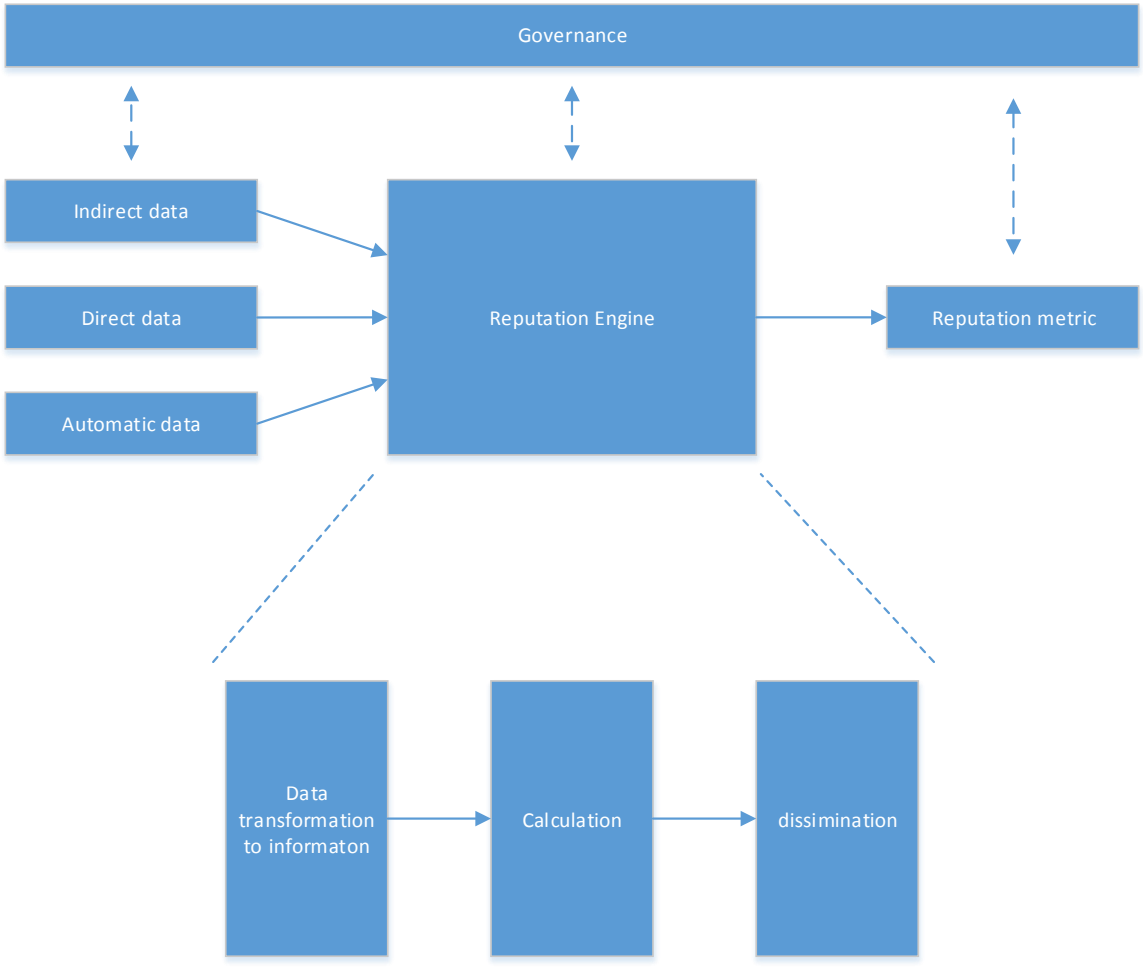

Figure 15 Data to reputation; the reputation system

Theory on reputation and reputation systems gave insight in how reputation system work and what they consist of. However, they do not show what makes them effective in practice. For this reason existing initiatives were described.

In chapter 4 existing reputation systems are introduced, characterized and evaluated. This was done for five different systems: PageRank, eBays feedback forum, corporate reputation (Reputation quotient and MAC index) and reputations in the financial world, the credit ratings.

Many of these initiatives oblige by the objectives set in section 3.2.1 (accurate, preference over current behavior, smooth and resistant to attacks). The well working initiatives have in common that they oblige by the criteria. Another factor in their success is the methodology for feedback gathering and reputation communication are very well established. Generally it can be determined that the effective initiatives are PageRank, Reputation quotient and Credit ratings **(Rq 4a)**

Some facts have been identified or confirmed. One of these is the question: should a reputation system measure efforts or results? A reputation, where also the efforts are measured, would introduce a more complete overview of the mitigation efforts. It is based on what is valued higher: efforts or results. If efforts are to be used as a reputation system, an initiative as the reputation quotient could be used for determining a reputation for ISPs.

The methodology by S&P (Sec. 4.3) is a good framework for design of a reputation system for ISPs. Both output and efforts can be incorporated in this methodology. It uses categorized scales and would therefore be a good method to deal with volatility of botnet measurements. The framework would have to be adapted for use in another context **(Rq 4b)**.

7.1.3 Conclusions on the requirements and design space

The knowledge from chapters 1 to 4 is combined into a set of design requirements and a design space. The design space shows all the possibilities, were the requirements show what a design has to oblige by.

The requirements for the design are listed below. A reputation system for ISPs based on botnet activity should **(Rq 5)**:

1. Decrease the information asymmetry in the market
2. Have a positive effect on botnet mitigation
3. Realign incentives for ISPs
4. Correspond to the right context
5. Only give a reputation to long living entities
6. Take into account previous behavior
7. show expected future behavior
8. Be a systematic algorithm
9. Be resistant to cheating from inside the system
10. Be resistant to attacks from outside of the system
11. Be well-distributed
12. Always work
13. Accessible for the intended user
14. Usable for the intended user

15. Deal with volatility of underlying metrics
16. Assign more value to new observations
17. Up-to-date
18. Maintain privacy of customers of ISPs
19. Based on automatically extracted data
20. Based on measurable concepts
21.
    a. Be able to deal with false positives
    b. Reliable
    c. Valid
22. Be able to deal with botnet takedowns
23. Account for industry trends
24. Differentiate between the size of the entities
25. Adaptable
26. Not represent a snapshot the situation
27. Not be opposed by the involved stakeholders

In section 5.1. and Table 3 the requirements are explained and their place of origin is shown. The knowledge from all previous chapters also identified a design space. The design space merges the dimensions from reputation and reputation systems with the context of ISPs and shows all the possibilities for design. It is listed what the possibilities for each would be in the context of designing a reputation system for ISPs. The criteria and possibilities are **(Rq 6)**:

- **Context**: should the reputation only be for ISPs based on botnet activity. Or should it be more general to also include other intermediaries or other topics than botnets?
- **Information source:** should metrics as infected machines be used, or should also efforts of ISPs for mitigation be measured? The efforts are measured by interviews and therefore more subjective.
- **Governance model:** Who would govern the reputation? This determines how the reputation is derived. Is this a stakeholder alone, or a coalition of stakeholders? Or are ISPs themselves assigning reputation?
- **Governance supervision:** Depending on the stakeholder implementing the system there are different types of supervision on the system possible. Be it a legal context to which the governance has to apply to, or a direct supervision.
- **Intended users:** who are the intended users of the reputation? This can be ISPs, Government, consumers or businesses.
- **Computation engine:** Which computation engine is possible in which context? Possible engines are Simple summation, Bayesian, Discrete, Flow, Belief models.
- **Communication:** How can intended users be reached? Possibilities include Digital platforms, Print, Word-to-mouth or Social media.
- **Cheating:** A reputation system can either assume that there is no cheating, that cheating cancels itself out in the long term or assume cheating and take countermeasures. In the context of ISPs it is possible to misrepresent infections.

However, an ISP which is really working on mitigation would see over time lower infection rates for multiple metrics. Where an ISP who is cheating would suddenly have no infections for a specific metric. Abnormal behavior is an indication of cheating.

The challenges in designing a reputation system in the context of ISPs based on the number of infected machines they hosts lies in the combination of data availability, resources and the involved stakeholders.

The integration between chapter 2 (the institutional and technical settings) and the knowledge about reputation systems can be seen as a scientific contribution. Although there are still a lot of unknowns, it gives information over how to create reputation systems for incentivizing botnet mitigation.

The sole purpose of developing a reputation system is to realign incentives for ISPs to increase mitigation efforts so that the information differences in the market becomes smaller and the number of infected machines also becomes smaller (see requirements 1,2,3). It is therefore especially important that the reputation system shows the right reputation, meaning that it should be low on false positives, reliable and valid (see requirement 21). Knowledge, time, computer power, access to data, funding and cooperation are required **(Rq 7)**.

7.1.4 Design conclusions

For this reason two possible designs have been provided. These designs are from a neutral point of view, meaning not from the point of view from a specific stakeholder. The designs differ in the way these resources are available. The first design shows an ideal situation were different stakeholders work together in developing a reputation system. There are more resources, therefore a more comprehensive design can be developed. The situation identifies other issues as how to deal with all the stakeholders contributing to the design. The second design shows a less comprehensive design, employed by a single stakeholder with less resources.

The figure below shows a technical representation of the first design. Specifics can be found in section 6.1.

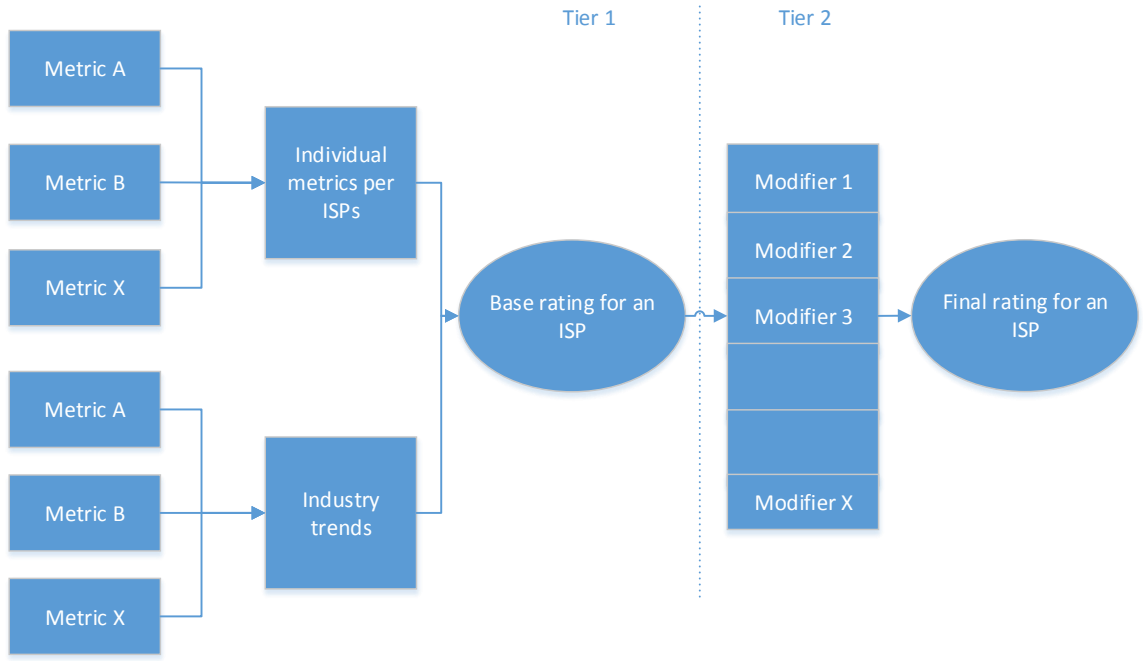

Figure 16 Technical architecture for alternative 1

Negotiations determine many of the design features. Governance is very important in this design. Essentially the What, Who and how questions for this design have to be determined by these negotiations: 1) *what* decisions need to made, 2) *who* should make these decisions, and 3) *how* these decisions are being made and monitored.

In the development of such a design four phases can be identified:

1. Determining stakeholders to include
2. Negotiation with stakeholders
3. Developing the alternative
4. Implementation and maintenance of the alternative

The second alternative continues on one of the existing initiatives identified in chapter 2, Spamrankings. For all metrics to be used, a ranking is made per metrics per ISP. The total score would be a summation over all rankings for all metrics. Such a system would be computationally much easier to develop compared to alternative 1. The previous alternative uses comparison lists which requires more processing power and knowledge to derive values for the tables. To achieve a good reputation system this way much of the attention has to go to the distribution of the reputation. How should it be published? Testing and validating the reputation system is also important here. Since rankings per metric are just summed it might be inaccurate.

Altogether; theoretically it seems possible to design a reputation system for ISPs based on the infected machines they host. False positives are a large issue, questioning accuracy. Given enough resources it would be plausible that a reputation can be developed to be accurate, in the context of reliable, valid and low on false positives.

The end result of such a design, or another design based on the design space from Ch. 5, can be the contribution of this research to society. By developing a reputation system, the threat of botnets can be reduced, solving a societal problem.

The last section describes limitations of the research and correspondingly gives possibilities for further research.

7.2 Research limitations and further research

This research has the assumption that ISPs do find some way to pay for the mitigation costs or someone will pay for the mitigation costs ISPs have. This is a clear limitation of the research. Increasing mitigation efforts for ISPs would result in high costs for them. Legally they are not responsible for their customers being infected, as this is the responsibility of the end users. Scholars have identified government or end users to be candidates to help with the mitigation costs. This assumption could be turned into a recommendation for further research. Determining what the costs for ISPs would be, who would be able and willing to pay for mitigation and how these costs than should be distributed. On the one hand this is an ethical discussion: is it fair to burden ISPs with the problems and costs of botnet mitigation. On the other hand it is also an economical problem, with the above questions of how much would it cost and who is going to pay.

Such a study could be part of a bigger question. The question is therefore: What is the business case for botnet mitigation (by ISPs). This research has discussed a method to incentivize botnet mitigation. Incentives can be seen as intangible costs or benefits and could therefore be a part of the research. As shown in ch.2, botnet mitigation for ISPs is costly. But are there also benefits for ISPs? A good reputation could become an asset. It could be that customers of ISPs become more loyal to an ISP if they are helped with removing malware. Such a study could also help to evaluate the effects of a reputation system, what is its value (with regard to requirements 1, 2 and 3).

Another research limitation is the level of depth and unknowns in the design. This results from the situation that no clear problem owner is identified yet as to implement such a solution. The designs have to be expanded and developed. A limitation of the research is that this has to be coupled with the stakeholder selection and negotiation. This can be overcome by a stakeholder implementing a design.

The design which have been provided both are possible to develop. The question is however, at what level of effort or costs. The requirements have failed to specify or take into account costs for development. Although it is likely that the designs will be possible, or even meet the requirements, it is unknown how much effort it would take. This adds to the limitation of the level of depth of designs. This level of depth is (partly) due to the fact that these are open designs, without a specific stakeholder. Further research could be to use the design space and requirements in the context of a specific stakeholder. By determining what kind of resources the stakeholder has available and what their goals in a reputation system are, a much more specific design can be made. The main recommendation for further research is therefore to specify a design in the context of a given stakeholder.

The research did not provide evaluation methods. Evaluation of a reputation system is difficult, because there is little to compare it to. The research did not provide a method to evaluate the actual accuracy of the reputation. It is impossible to determine all the infected cases and know how big the botnet problem exactly is. Therefore problems arise for evaluation, as the total values are indications and there is nothing to really compare it to. Partly this is a known problem with spam/botnet data. This research also did not provide a method to evaluate the effectiveness/accuracy. A possibility for evaluation is developing several alternatives and compare their scores in practice. Experts could also be used. Another possibility for evaluation could be simulation. Differences in the underlying metrics could then be compared to the differences in the reputation metric. If in such a case it appears that an ISP scores very high and bad on a specific metric, but the reputation is still very good, the accuracy of the reputation system could be questioned. This is a method of falsification, so it would only give reputation systems that do not work. It wouldn't determine a reputation system to be accurate.

BIBLIOGRAPHY

Abusehub. (2014). *Abuse information exchange home page*. Retrieved 10 1, 2014, from https://www.abuseinformationexchange.nl/

ACM. (2014). *Toezicht op internet*. Retrieved 10 21, 2014, from ACM: https://www.acm.nl/nl/onderwerpen/telecommunicatie/internet/toezicht-op-internet-en-zakelijke-netwerkdiensten/

Akerlof, G. (1970). The Market for Lemons: quality uncertainty and market mechanism. *Quarterly Journal of Economics*, 488-500.

Alperovich, D., Judge, P., & Krasser, S. (2007). Taxonomy of Email Reputation Systems. *ICDCSW'07 Proceedings of the 27th International Conference on Distributed Computing Systems Workshops* (pp. 27-34). Washington: IEEE Computer Society.

Anderson, R. (2001). Why Information Security Is Hard- An Economic Perspective. *Proc. 17th Annual Computer Security Applications Conference* (pp. 358-365). Accociation for Economic Service.

Ang, S., & Wight, A. (2009). Building Intangible Resources: The stikiness of Reputation. *Corporate Reputation Review*, 21-32.

BBC. (2013). *Internet's 'bad neighbourhoods' spread scams and spam*. Retrieved 12 13, 2013, from BBC technology: http://www.bbc.co.uk/news/technology-21798829

BBC.com. (2014). *Google warned by EU to make changes or face fine*. Retrieved 10 3, 2014, from BBC news: http://www.bbc.com/news/technology-29325580

Campbell, K., Gordon, L., Loeb, M., & Zhou, L. (2003). The Economic Cost of publicly announced information security breaches: Empirical evidence from the stock market. *Journal of Computer Security, 11*, 431-448.

Chess, B., & McGraw, G. (2004). Static Analysis for Security. *IEEE computer society*, 32-35.

Cooke, E., Jahanian, F., & McPherson, D. (2005). The Zombie Roundup: Understanding, Detecting and Disrupting Botnets. *Proceedings of the Usenix Sruty*. Retrieved from https://www.usenix.org/legacy/event/sruti05/tech/full_papers/cooke/cooke_html/

De Bruijn, H. (2006). *Prestatiemeting in the publieke sector.* Den Haag: Lemma.

De Bruijn, H., ten Heuvelhof, E., & in 't Veld, R. (2012). *Process Management.*

Dingledine, F., Freedman, M., & Molnar, D. (2000). Accountability measures for peer-to-peer systems. In F. Dingledine, M. Freedman, & D. Molnar, *Peer-to-peer: Harnessing the power of disruptive technologies.* O'Reilly Publishers.

DShield.org. (2014). *Information about DShield*. Retrieved 10 25, 2014, from DShield.org: https://www.dshield.org/about.html

Economides, N. (2007). *The economics of the internet. The new palgrave dictionary of economics.* London: Macimillan.

Economides, N., & Tag, J. (2012). Network neutrality on the Internet: A two-sided market analysis. *Information Economics and Policy*, 91-104.

Eeten, M. v., Asghari, H., Bauer, J., & Tabatabaie, S. (2011). *Internet Service Providers and Botnet Mitigation: A Fact-Finding Study on the Dutch Market.*

Elliot, C. (2010). Botnets: To what extent are they a threat to information security? *Information security technical report*, 79-103.

ENISA. (2011). *Botnets: Detection, Measurement, Disinfection & Defence.*

Eyckelhof, C. (2014).

FCC. (2013). *List of members for CSRIC*. Retrieved 9 10, 2014, from FCC: http://transition.fcc.gov/bureaus/pshs/advisory/csric3/List%20of%20CSRIC%20Members.pdf

Fombrun, C. (1996). *Reputation: Realizing Value from Corporate Image.* New York: Harvard Business School Press.

Fombrun, C., & Gardberg, N. (2000). Who's Tops in Corporate Reputation. *Corporate Reputation Review*, 13-17.

Fombrun, C., & Shanley, M. (1990). What's in a Name? Reputation building and corporate strategy. *Academy of Management Journal*, 233-258.

Fombrun, C., & Van Riel, C. (1997). The Reputational Landscape. *Corporate Reputation Review, 1*(2), 5-13.

Fombrun, C., Gardberg, N., & Sever, J. (2000). The reputational quotient: a multi-stakeholder measure of corporate reputation. *Journal of Brand Management, 7*(4), 241-255.

Gotsi, M., & Wilson, A. (2001). Corporate reputation: Seeking a definition. *Corporate Communications: An international Journal, 6*(1), 24-30.

Hall, R. (1993). A Framework Linking Intangible Resources and Capabilities to Sustainable Competitive Advantage. *Strategic Management Journal, 14*, 607-618.

Herder, P., & Stikkelman, R. (2004). Methanol-Based Industrial Cluster Design: A Study of Design Options and the Design Process. *Ind. Eng. Chem. Res.*, 3879-3885.

Hoffman, K., Zage, D., & Nita-Rotaru, C. (2009). A survey of attack and defense techniques for reputation systems. *ACM Computing Surveys*, 1-31.

Johnson, E., & Goetz, E. (2007). Embedding Information Security into the organization. *IEEE Security and Privacy*, 16-24.

Josang, A., & Golbeck, J. (2009). Challenges for Robust Trust and Reputation Systems. *5th international Workshop on Security and Trust Management* (pp. 1-12). Saint Malo, France: Elsevier.

Josang, A., & Ismail, R. (2002). The Beta Reputation System. *15th Bled Electroninc Commerce Conference e-Reality: Constructing the e-Economy*, (pp. 1-14). Bled.

Josang, A., Ismail, R., & Boyd, C. (2007). A survey of trust and reputation systems for the online service provision. *Decision support systems*, 618-644.

Kaner, C., & Bond, W. (2004). Software Engineering Metrics: What Do They Measure and How Do We Know. *10th international Software Metrics Symposium, Metrics* , (pp. 1-12).

MAAWG. (2007). *BIAC and MAAWG Best Practices for ISPs and Network Operators.* Messaging Anti-Abuse Working Group.

Moore, T. (2010). The economics of cybersecurity: Principles and policy options. *International journal of critical infrastructure protection*, 103-107.

Moreira Moura, G. (2013). *Internet Bad Neighborhoods.* Ipskamp Drukkers B.V.

Mui, L., Halberstadt, A., & Mohtashemi, M. (2002). Notions of Reputation in Multi-Agent Systems: A review. *AAMAS '02 Proceeding of the first international joint conference on autonomous agents and multiagent systems*, (pp. 280-287).

Mui, L., Mohtashemi, M., & Halberstadt, A. (2002). A computational model of trust and reputation. *Proceedings of the 35th Hawaii International Conference on System Sciences*, (p. 9).

NCSC. (2014). *Homepage NCSC*. Retrieved 10 1, 2014, from www.ncsc.com

Net-security. (2013). *nearly a third of all computers are infected with malware*. Retrieved `9 22, 2013, from Net-security.org: http://net-security.org/malware_news.php?id=2404

NMAP.org. (2014). *Port scan specification*. Retrieved 11 10, 2014, from NMAP.org: http://nmap.org/book/man-port-specification.html

OECD. (2011). *The role of internet intermediaries in advancing public policy objectives. Forging partnerships for advancing policy objectives for the Internet economy, Part II.*

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The Pagerank Citation Ranking: Bringing Order to the Web. *Stanford Infolab*.

Pfleeger, S., & Cunningham, K. (2010). Why Measuring security is hard. *IEEE Security metrics*, 46-54.

Pieterse-Bloem, M. M. (2014, 10 20).

Politie. (2013). *Botnets in Europa bestreden*. Retrieved 11 26, 2014, from Politie.nl: http://www.politie.nl/nieuws/2013/december/6/11-botnet-in-europa-bestreden.html

Politie.nl. (2013). *Cybercrime*. Retrieved 11 6, 2013, from politie.nl: http://www.politie.nl/onderwerpen/cybercrime.html

Puri, R. (2003). *Bots & Botnet: An overview.* SANS Insitute InfoSec Reading Room.

Quarterman, J., & Whinston, A. (2010). Fireeye's ozdok botnet takedown in spam blocklists and volume observed by iiar project. *Proceedings of the 48th North American Network Operators Group.* NANOG 48.

Quarterman, J., Linden, L., Tang, Q., Lee, G., & Whinston, A. (2013). Spam and Botnet Reputation Randomized Controlled Trials and Policy. *TPRC*.

Rao, J., & Reiley, D. (2012). The Economics of Spam. *The Journal of Economic Perspectives, 26*(3), 87-110.

Rekheter, Y., Li, T., & Hares, S. (2006). *RFC 4271 - A Border Gateway Protocol 4.* Network Working Group. Retrieved from http://tools.ietf.org/html/rfc4271

Resnick, P., Zeckhauser, R., Friedman, E., & Kuwabara, K. (2000). Reputation Systems. *Communications of the ACM, 43*(12).

S&P. (2014a). *S&P General Criteria: Principles of Credit ratings*. Retrieved 5 14, 2014, from Standardandpoors.com: http://www.standardandpoors.com/prot/ratings/articles/en/us/?articleType=HTML&assetID=1245366284668

S&P. (2014b). *Criteria Corporate General Methodology*. Retrieved 5 14, 2014, from Standardandpoors.com: http://www.standardandpoors.com/prot/ratings/articles/en/us/?articleType=HTML&assetID=1245367512383

Sabatier, J., & Sierra, C. (2005). Review on Computational Trust and Reputation Models. *Artificial intelligence review*, 33-60.

Saranya, C., & Manikandan, G. (2013). A study on normalisation techniques for privacy preserving data mining. *International Journal of Engineering and Technology*, 2701-2704.

Saxton, M. (1998). Where do Reputations Come From? *Corporate Reputation review, 1*(4), 393-399.

SenderBase.org. (2014). *Last days spam data per country*. Retrieved 5 12, 2014, from SenderBase: http://www.senderbase.org/static/spam/#tab=4

Shadowserver. (2014). *Information Honeypots*. Retrieved 10 29, 2014, from https://www.shadowserver.org/wiki/pmwiki.php/Information/Honeypots

Shadowserver.org. (2014a). *Stats Conficker*. Retrieved 11 1, 2014, from Shadowserver.org: http://www.shadowserver.org/wiki/pmwiki.php/Stats/Conficker

Silva, S., Da Silva, r., Pinto, R., & Salles, R. (2013). Botnets: A Survey. *Computer Networks*, 378-403.

Smith, E. (2008). *Bringing Down Wall Street as Rating Let Loose Subprime Scourge*. Retrieved 5 30, 2014, from Bloomberg: http://www.bloomberg.com/apps/news?pid=newsarchive&sid=ah839IWTLP9s

Spamrankings.net. (2014). *Fag Spamrankings.net*. Retrieved 5 8, 2014, from Spamrankings.net: http://cloud.spamrankings.net/faq.html?ud=d3649ab1-45f4-47e2-ac6f-8b8d0cfc5612

St. Sauver, J. (2012). *FCC CSRIC WG7 Botnet Metrics.*

Tang, Q., Linden, L., & Quarterman, J. (2013). Improving internet security trough social information and social comparison. *WEIS*.

Techzine. (2013). *Nederlandse ISPs gaan de strijd aan met botnets*. Retrieved 11 15, 2013, from Techzine.nl: http://www.techzine.nl/nieuws/36413/nederlandse-isps-gaan-strijd-aan-met-botnets.html

Telecomwet. (2014). *Dutch Telecomlaw*. Retrieved from http://wetten.overheid.nl/BWBR0033401/geldigheidsdatum_05-02-2014

van Eeten, M., & Bauer, J. (2008). *ECONOMICS OF MALWARE: SECURITY DECISIONS, INCENTIVES AND EXTERNALITIES.* STI WORKING PAPER 2008/1.

van Eeten, M., Asghari, H., Bauer, J., & Tabatabaie, S. (2011). *Internet Service Providers and Botnet Mitigation: A Fact-Finding Study on the Dutch Market.*

van Eeten, M., Bauer, J., Asghari, S., Tabataie, S., & Rand, D. (2010). The role of internet service providers in botnet mitigation: An emperical analysis based on spam data. *The Ninth Workshop on the Economics of Information Security.*

van Erp, J. (2014). Naming and Shaming of Corporate Offenders. New York: Springer Science.

Wagner, C., Francois, J., Dulaunoy, A., Engel, T., & Massen, G. (2013). Ranking ASs Providing Transit Service to Malware hosters. *IFIP/IEEE International symposium on Integrated Network Management*, (pp. 260-269).

Wartick, S. (2002). Measuring Corporate Reputation: Definition and Data. *Business & Society, 41*(4), 371-392.

Weil, P., & Ross, J. (2004). *IT Governance: How top performers manage IT decision rights for superior results.* Harvard Business School Press.

Wikipedia. (2014). *Moving average*. Retrieved 10 25, 2014, from http://en.wikipedia.org/wiki/Moving_average

Wood, D., & Rowe, B. (2011). Assessing home internet users' demand for security: will they pay isps. *Workshop on the economics of information security.*

APPENDIX A STAKEHOLDERS AND INTERESTS

This appendix identifies the point of view of some stakeholders. These stakeholders are ISPs, Security software vendors, end users, governmental organizations and malicious internet users. For all these stakeholders their description, role, relationship to ISPs, attitude towards botnets and attitudes towards a reputation system are described. Much of the information is based on interviews.

Internet Service Providers and abuse information exchange

Description

ISPS are the bridge between the end user and the internet. Therefore an ISP knows which IP address belongs to which user. Although formally they are not responsible for their end users being infected, it has become some sort of responsibility for them. ISPs can see the internet traffic of their users, it is for them possible to determine which user is infected. However, privacy laws prevent them for doing this. They also have to rely on external sources to provide them with information of their customers being infected. They are the bridge between end user and internet, they know which IP adress belongs to which end user. Measuring botnet activity is possible based on IPs. Therefore it is possible for them to determine which user is infected. They have the power to warn or quarantaine this user (van Eeten M. , Bauer, Asghari, Tabataie, & Rand, 2010).

Role

The ISP provides the access to the internet. They are part of a broader group of internet intermediaries.

Attitude towards botnet mitigation

Many of the ISPs have started a collective: the abuse information exchange. In this initiative these ISPs are working together to share knowledge about how to (quickly) deal with botnet mitigation. Also they do measurements about botnets at the Abusehub. The Abusehub in turn informs the corresponding ISP about such an infection. It is for the ISPs self to decide if and what kind of action to undertake.

Therefore it can be said that many ISPs are working on the problem. However there are differences in how well each ISP performs in terms of botnet mitigation. Some are (relatively) very clean, while others have much more infections in their network. The infected ISPs therefore can be a burden for them. They do not all have the same attitude towards botnet mitigation.

Attitude towards a reputation system

In mitigating this problem they would like to be as much in control as possible. In every solution, as with a reputation system, they would at least want to be regularly consulted. Some consulting is also feasible to check if metrics and assumptions correspond with how their networks work.

For ISPs that are performing well a reputation score could help with awareness amongst customers. So if such an ISP determines one of their customers infected, the customer has heard

of ISPs and botnet mitigation. The increase in information for end users can help ISPs. A reputation system could also be used by an ISP for comparing how well they are doing with other ISPs.

Sub group Rogue ISPs

Similar as regular ISPs, they have the same rules to follow. The difference is that these ISPs ignore the problem of botnets. They do not want to invest in mitigation measures. Therefore they have many infected end users. Other ISPs are affected by these bad ISPs, because their end users are attacked by the misbehaving infected end users of the bad ISP. Eventually it will become worse and worse for them.

Data and knowledge providers

Description

These organizations can provide data about infected machines per ISP/ or IP. Such organizations can be security companies (anti-virus software providers), Microsoft but also research institutes as TU Delft or data providers as Dshield/Shadowserver

Role

A part of their role is to deliver their services to ISPs and large companies (i.e. anti-virus companies). Another role for these such stakeholders can be to help in the problem of measuring botnet activity. Such companies can provide data or knowledge.

Relationship with ISPs

Many organizations buy security software from these vendors. Their successes are in a way based, because other people are infected with malware (for security solution vendors). If ISPs are putting in more efforts to mitigate botnets, there could be opportunities for these companies to sell security solutions to ISPs.

For research institutes and data providers as DShield this does not really exist. Their point of view is more related towards botnet mitigation.

Point of view towards botnet mitigation

For commercial companies, it might be that they gain from the process of botnet mitigation. The more neutral research institutes the point of view is improving the world by providing services for i.e. mapping botnet activity.

Attitude towards a reputation system.

If such a system helps to increase the mitigation efforts it should be positive.

Authority Consumer and Market (ACM)

Description

With regard to cybersecurity there are many activities the ACM does, as the ACM is the regulatory agent enforcing the telecomlaw. I.e. The ban over sending spam and distributing malware. Specified in the telecomlaw art. 11.

With regard to botnets and spam the ACM has specific projects as “Nederland Schoon”. Such a project is aimed at mapping and mitigating the malware within autonomous systems.

The ACM is also involved with research into spam and the market behind malvertising (the online advertisement chain and its weaknesses.

Role

The role of the ACM is reactive. Is someone is determined to be infected, the ACM could warn. With regard to cybersecurity their role is therefore not one of assigning penalties, but warning and by using the argument of societal responsibility mitigating problems.

Relationship with ISPs

The ACM is the regulatory body which enforces net neutrality i.e. in the market for ISPs. Therefore they can also be seen as a regulatory body over ISPs. With regard to botnet mitigation and ISPs the ACM’s role is more complex. As ISPs are formally not responsible for their customers being infected, the ACM can only ask ISPs to help their customers mitigate these infections.

Attitude towards botnet mitigation

With projects as “Nederland Schoon” a goal of the ACM is for the Netherlands to set an example for the rest of the EU. The attitude towards botnet mitigation is therefore positive.

Attitude towards a reputation system

A system that assigns scores to ISPs in term of how many infected users there are in their network would in general not lead to any kind of punishment for an ISP. If an ISP seems to be performing worse than others, the system might initiate a visit from the ACM to such an ISP to discuss this. This would only be on the condition that the score is reliable and valid. In such a setting they could be a user of a reputation initiative.

Uninformed end user

Description

An uninformed end user is someone with little knowledge about security/cybersecurity and the threat of botnets. Someone is usually unaware of the problem. If they become infected with a machine they often do not know about it.

Role

The end user subscribes to an ISP. Formally the end user is responsible for removing their own malware.

Relationship towards ISPs

ISPs contact infected customers. Such a customer is often unaware of ISPs contacting customers. So when they get a notification of being infected, this is a surprise for them. The ISPs service desk has to convince such a customer that they are in fact infected by malware, that they are speaking to a legitimate source (the ISP). This requires a lot of time, increasing costs for ISPs.

Attitude towards botnet mitigation

The unawareness about the topic can create aversion towards influence from an ISP when they try to help in removing bots. Some customers even switched to other ISPs because of this aversion.

Attitude towards a reputation system

Publishing reputation scores for ISPs, together with explanations over possible courses of actions ISPs can take would increase the awareness for these customers. A score and explanation might help to reduce the efforts ISPs have to take to convince a customer that they are in fact infected. This means that the uninformed users are a potential intended users, because they little knowledge a score should be well explained to them. It should be linked with possible courses of action with ISPs could be do towards mitigation.

Informed end user

Description

Such an informed end user has knowledge about the topic. This can be because they were infected once and already have had some contact with an abuse desk from an ISP.

Role

The end user subscribes to an ISP. Formally the end user is responsible for removing their own malware.

Relationship towards ISPs

ISPs contact infected customers. An informed customer is satisfied after he or she is convinced that they really have been helped by the ISP. This enforces their relationship to the ISP.

Attitude towards botnet mitigation

Usually not negative. However there are customers who are aware of the problem but do not want to contribute to it or remove the malware from their machines. This applies for the unsatisfied customers because of botnet mitigation

Attitude towards the reputation system

For the satisfied customer of an ISP it becomes more clear how well their ISP is doing, reinforcing the band with the ISP.

Criminals & malicious internet users

Description

Organized criminal organizations, people from inside of networks, ethical hackers are all possible malicious internet users (Cooke, Jahanian, & McPherson, 2005). They benefit from the discordance between the different stakeholders. This group of stakeholders will desire to maintain the status quo.

Role

On the other side of the problem. They cause the problem. This group of stakeholders has a gain

Attitude towards botnet mitigation

They actively oppose mitigation. They will try to undermine the mitigation efforts, or look at new initiatives which are unaffected by mitigation efforts

APPENDIX B COMPARISON OF DIFFERENT NORMALISATION TECHNIQUES

For normalization there are many techniques. Min-mam normalization and Z-score standardization are identified to be the most common used (Saranya & Manikandan, 2013). The other normalization techniques described below give ratios of a value compared to a total, mean or logarithmic scale.

The benefit of the first to is that they rescale the variable. For all different variables using this technique the scale would therefore be the same. The downside is that outliers in data will influence the scale. This means that normal data is converted to a smaller interval.

	DESCRIPTION	FORMULA
MIN-MAX	Min-max normalization performs a linear alteration on the original data. The values are normalized and scaled to a value between 0 and 1	X(i, rescaled) = (X(i) - Min(X))/ (Max(X) - Min (X))
STANDARDISATION/Z-SCORE	Data is normalized based on the mean and standard deviation	Z-score X(i) = X(i) - avg(X)/ (st. dev (X))
NORMALIZE BY TOTAL	With a ratio, values are corrected by dividing them by a value. For example a total value	X(i, norm) = X(i)/ total value (X)
NORMALIZE BY MEAN	With a ratio, values are corrected by dividing them by a value. For example by the mean	X(i, norm) = X(i)/ avg (X)
NORMALIZATION BY LOGARITMIC RATIO	A log scale has the benefit over the previous two that the difference between 1 and 2 and 100 and 101 is absolute the same. With a logarithmic value the difference between 1 and 2 gets more value than the	X(i, norm) = LN(X(i)/ Y(i))

	difference between 100 and 101.

In the table X(i, norm) stands for the normalized value of metric x. Max (x), Min (x), average (x) etc show a maximum, minimum or average value for all the values of the observed metric. LN shows the natural logarithm and st. dev the standard deviation of X.

APPENDIX C REQUIREMENTS COMPARED TO EXISTING REPUTATION SYSTEMS FROM OTHER FIELDS

The table below show how the different existing initiatives score on the requirements. An "X" shows that the requirement is met. If it is not met, the field is left blanc.

However not all requirements were met. This means that there are differences in theory and practice. For example: not all systems are resistant to cheating (PageRank, eBay), previous behavior is always not taken into account (PageRank) and some also give a reputation to a short living entity (eBay). Therefore requirements 5, 6, 9 and 10 are not met in practice. S&P gives ratings twice a year, therefore the rating is not up to date (req 17). The field of corporate finance uses questionnaire data, it is questionable if such a system is always reliable (req 21). In appendix C a table is provided which shows how the requirements correspond to the existing alternatives.

Table 9 Reputation requirements compared to existing initiatives

A REPUTATION SYSTEM SHOULD	THEORY ON RS (CH. 3)	PAGE-RANK	EBAY	CORPORATE REP	FINANCE
4 CORRESPOND TO THE RIGHT CONTEXT	X	X	X	X	X
5 ONLY GIVE A REPUTATION TO LONG LIVING ENTITIES	X			X	X
6 TAKE INTO ACCOUNT PREVIOUS BEHAVIOR	X		X	X	X
7 SHOW EXPECTED FUTURE BEHAVIOR	X	X	X	X	X
8 BE A SYSTEMATIC ALGORITHM	X	X	X	X	X
9 BE RESISTANT TO CHEATING FROM INSIDE THE SYSTEM	X		X	X	X
10 BE RESISTANT TO ATTACKS FROM OUTSIDE OF THE SYSTEM	X	X		X	X
11 BE WELL-DISTRIBUTED	X	X	X	X	X
12 ALWAYS WORK	X	X	X	X	X
13 ACCESSIBLE	X	X	X	X	X
14 USABLE	X	X	X	X	X
15 DEAL WITH VOLATILITY OF UNDERLYING METRICS	X	X	X	X	X
16 ASSIGN MORE VALUE TO NEW OBSERVATIONS	X	X	X	X	X
17 UP-TO-DATE	X	X	X	X	
21 BE ABLE TO DEAL WITH FALSE POSITIVES, RELIABLE AND VALID	X		X		X
23 ACCOUNT FOR INDUSTRY TRENDS				X	X
24 DIFFERENTIATE BETWEEN THE SIZE OF THE ENTITIES		X	X	X	X