Countering Passive Reconnaissance Activities of Adversaries

~

Research into guidelines that aim to identify and reduce information in tenders that can be misused by adversaries.

MSc-thesis in Cyber Security

by Arjan Sibma

Cyber Security Academy

19-12-19

Version 1.0

CYBERSECURITYACADEMY

Colophon

Title	Countering passive reconnaissance activities of adversaries
Subtitle	Research into guidelines that aim to identify and reduce information in tenders that can be misused by adversaries.
Author	Ing. Arjan Sibma
Student number	S2328348
Supervisor	Dr. ir. Pieter Burghouwt
Second reader	Prof. dr. ir. Jan van den Berg
Institution	Cyber Security Academy
Faculty	Governance and Global Affairs
Module	Cyber Security: Thesis 2019 (8824CSTHE-CSA-1819FGGA)
Date	19th of December 2019
Version	1.0 – Finished

Abstract

This report describes the results of the research in the context of the Master's degree in Cyber Security. This study researched the problem that public tenders contain information that can be collected by hackers during their preparation for a cyber-attack. Adversaries can easily access this information and abuse it against the tendering organizations. Therefore, the main goal of this research is to establish guidelines aimed at identifying and reducing sensitive information in tenders, in order to prevent that malicious parties gather and use this information in the preparation of cyber-attacks against tendering organizations. To this end, the various concepts of open data, procurement, reconnaissance, cyber kill chain, and open source intelligence were examined. In addition, interviews were held to both identify the risks for tendering organizations due to the above identified problem and to evaluate on the established guidelines.

The research results have shown that the information in tenders is public due to the principles on which the rules regarding tenders are based. This is to offer fair opportunities to companies to win contracts through tenders. Due to the public nature of information in tenders, the comparison can be made with the concept of open data. As a result, the risks inherent to open data, such as abuse by malicious parties, also apply to information in tenders.

Further research into the reconnaissance activities of hackers has made it clear that hackers are looking for specific types of information in preparation for cyber-attacks. It has been determined through document analysis on real tenders and interviews with security professionals that these types of information occur in tenders. This means that malicious parties can use tenders to collect information about organizations that is relevant for the preparation of cyber-attacks, against the tendering organizations. As a result, the tendering organizations face risks with regard to the confidentiality, integrity and availability of company assets. In particular, the likelihood that such risks arise is increased because the information is easily accessible to malicious parties.

In order to prevent this, guidelines have been established. These guidelines should be used in follow-up research where a final solution is created that implements the described functionalities of the guidelines. The established guidelines focus in particular on identifying and reducing information that is relevant to hackers in tenders, before the tenders are made public. In this way the risks for tendering organizations can be prevented because this information can no longer be collected by malicious parties. The likelihood of the identified risks occurring is reduced. In addition, techniques have been suggested for these guidelines on which they can be implemented. The techniques regular expressions, text mining, comparison with known information, optical character recognition, and image recognition are discussed.

Furthermore, guidelines have also been established that focus on the practical side of a final solution and the fact that this solution must be used in an existing context: people, processes and organizations. The guidelines and the results of the study were evaluated in interviews with senior purchasers. From these interviews it can be concluded that a solution based on the guidelines is of added value in practice in order to reduce sensitive information in tenders and prevent risks for tendering organizations.

The results of this research thus result in recommendations for follow-up research, where the aim is to create an automated solution based on the guidelines that have been established.

Keywords: design science, tenders, open data, open source intelligence (OSINT), reconnaissance, cyber-attack, risks of open data, automation.

Table of Contents

Figures & Tables

1 Introduction

1.1 Opening

This document contains the master thesis for the Master of Science program in Cyber Security. The thesis is the final result of the module Cyber Security: Thesis 2019 (8824CSTHE-CSA-1819FGGA). The master thesis is also the final assignment to successfully end the Master of Science program in Cyber Security – part of the Faculty of Governance and Global Affairs of Leiden University – at the Cyber Security Academy (CSA). The CSA is an act of working jointly by Leiden University, Technical University Delft, and The Hague University of Applied Sciences.

Writing the master thesis results in *"a written report of research which has been carried out under supervision of a lecturer, but with a high degree of self-dependence. The thesis is the final project before graduation in the Master"* (Leiden University, n.d.). The master thesis has a course load of 15 EC. The goal of the thesis within the full context of the master program Cyber Security is to write the scientific report, with a high degree of independence, under supervision of a lecturer, within the specified time, using insights gained during the master program in Cyber Security and while writing the thesis.

1.2 Research context

When a government organization has to purchase a good, it must comply with legislation. This legislation in The Netherlands is called the Procurement Act (*Dutch: Aanbestedingswet*). Take the purchase of workplace hardware for a large organization as an example. The purchase of this must be done according to the Procurement Act. This way, organizations that want to take on the assignment have equal chances of getting the assignment. The details of the purchase are included in a tender. During this process, the contracting organization makes information public about what it wants to purchase. Organizations that want to win the contract can then subscribe to the tender on the basis of this information. All organizations must be given equal opportunities. Therefore, the Procurement Act is based on four principles: non-discrimination, equal treatment of organizations, transparency, and proportionality (Rijksoverheid, 2018). The details of the principles are explained later. In short, these principles ensure that the information on a tender is public to all organizations and that all organizations are treated equal. Due to this, the tender information is made public. Everyone can access this information.

Public information is generally perceived as positive. Governments, companies, and individuals benefit from it. However, public information can also be misused. Earlier research has shown that there are possible risks to the security of ICT infrastructure due to Open Government Data (Mauri, Mulas & Ariu, 2000). Another study concluded that open government data is useful for adversaries performing social engineering attacks (Labohm, 2018).

Adversaries collect relevant human, organizational, and technical information on its targets with the purpose of finding relevant information that can be used as input for their cyber-attacks (MITRE, 2000). The information that adversaries collect is called sensitive information during this research. According to the Cyber Kill Chain (Lockheed Martin, 2000) (CKC) framework, these activities are carried out during the reconnaissance phase.

The guidelines for buyers and others on the tendering process, made by the Ministry of Interior Affairs and the Ministry of Economic Affairs and Climate, does not include guidelines on the prevention of the inclusion of technical information. Neither are there guidelines for buyers considering the technical information in public tenders that might be abused.

In this thesis research is done on the risks for tendering organizations when sensitive information is collected from tenders by malicious parties who misuse this information for preparing cyber-attacks. Based on this, guidelines are defined that can help to develop a tool that reduces the sensitive information in tenders. The development of this tool is outside the scope of this research. This research stops with the guidelines and proposes recommendations for further research based on the guidelines.

1.3 Research relevancy

The procurement act is based on the four earlier mentioned principles, with the goal of providing fair chances to organizations and contracting authorities in the battle of winning tenders. Thus, the transparency of the data has a clear positive effect on the tender process. However, open data also has disadvantages to organizations. In the past there have been cases where personal or other sensitive information was unintentionally included in public tenders. It is hard to determine which information was gathered during reconnaissance after an attack happened. Therefore it is important to identify the risks that tendering organizations face and propose guidelines to minimize relevant public information that can be misused by malicious parties.

While the rules around tenders already exist since 1996 (PIANOo, n.d.), it is still valuable to do research on the risks arising from the transparency of the data in tenders and how these risks can be mitigated. A side-effect is that the research results can also support a discussion and raise awareness on the topic, for both security folks and people involved in the tendering process.

Outside of the scope of the research, readers gain insights on tactics and techniques used by adversaries to collect public information, on the tender process, and on the risks of sensitive information in tenders. These insights also enhances the cybersecurity awareness of the reader.

1.4 Problem statement

Certain organizations must purchase services, works and deliveries according to the tender rules (further: tendering organizations). The principles on which the law is based aim to benefit entrepreneurs and contracting authorities in providing fair chances for all in winning tenders. However, most complicated constructs, such as laws, might be dual edge swords: *"something that has or can have both favourable and unfavourable consequences"* (Merriam-Webster, n.d). It is possible that tenders contain information that is relevant for adversaries in preparing cyberattacks *(in the remaining parts of the thesis 'data that is relevant for adversaries in preparing cyberattacks' is referred to as 'sensitive information').*

In the past years the number of cyberattacks has increased dramatically. The quality, the sophistication, and the ingeniousness of the attacks have also grown tremendously. Although this is undoubtedly due to technological developments, the increasing wealth of open data may also play a role. A study by RSA (2015) shows how relatively easy it is to collect public information on persons and organizations. It demonstrates how a tremendous amount of intelligence can be gathered on almost every company and its employees. This information is used to identify weaknesses that can be exploited. It also shows that it doesn't require advanced knowledge to gather this intel. Adversaries have a growing pool of public resources with information to prepare for cyber-attacks. Information in tenders may also serve this purpose. In addition, the tender information tells adversaries something about the IST-situation and the SOLL-situation of an organization.

This thesis researches whether tenders contain information that can be collected and abused by malicious parties, and provides recommendations for further research to identify and reduce this sensitive information in order to prevent risks for tendering organizations.

1.5 Goal of the research

The main goal of this research is to establish guidelines aimed at identifying and reducing sensitive information in tenders, in order to prevent that malicious parties gather and use this information in the preparation of cyber attacks against tendering organizations.

These guidelines must be used when designing a final solution that can be used in practice to identify and reduce sensitive information from tenders. The final solution is *not* made in this research.

Research questions are determined to fulfil the goal of the research. These research questions are answered throughout the thesis.

1. What is the relationship between the information in tenders and cyber-attacks?
2. Which information that is relevant for adversaries can be collected from public tenders during the preparation of cyber-attacks?
3. What are the risks that tendering organizations face if adversaries collect this information from tenders during the preparation of cyber-attacks?
4. Which guidelines can be determined that aim to prevent the risks that occur if adversaries collect sensitive information from tenders during the preparation of cyber-attacks?
5. How can the established guidelines be implemented to create a solution that can be used to identify and reduce sensitive information in tenders in order to prevent that adversaries gather sensitive information during the preparation of cyber-attacks?

The following results are achieved by answering the research questions:

- Research question 1 examines the relationship between information in tenders and cyber-attacks by linking to open data and the risks of open data.
- Research question 2 examines the types of information that adversaries collect when preparing cyber-attacks. Subsequently, it is researched which of these types of information are also found in public tenders.
- Research question 3 examines which risks can be identified that tendering organizations face when adversaries have gathered this information from tenders. The causes that lead to the risks serve as the basis for the to be established guidelines.
- Research question 4 examines which guidelines can be determined that support the goal of identifying and reducing sensitive information from public tenders. With this, the risks identified in research question 3 are being prevented. On the basis of these guidelines a follow-up research can be initiated where a final solution is actually designed that reduces the sensitive information from public tenders.
- Research question 5 examines and describes how the guidelines can be implemented in a follow-up research in order to work towards an effective solution that reduces the sensitive information from public tenders.

1.6 Structure of the thesis

The outline of this is report is described below.

1. The first chapter ‘Introduction’ contains the introduction of the research. The goal of this chapter is to get the reader accustomed and introduced to the research topic. The chapter

contains the thesis overview, context and relevancy of the research, the problem statement, goal of the research, and the (main) research questions. This chapter highlights the problem and the relevancy of the research. A more in-depth study on the problem is done in chapters 3, 4 and 5.

2. The second chapter 'Research Design Methodology' contains the methodology that is used to structure the research. The goal of this chapter is to explain the reader how the research is structured in order that the reader understands the research steps and outcomes. The research type 'Design Science Research' (further: DSR) is used. This research methodology is explained in this chapter. In this chapter an overview of the thesis structure is visualized.
3. The third chapter 'Tenders as Reconnaissance Source' describes the relationship between information in tenders and cyber-attacks. Information on the tender process, open data, hacker activities, and previously misused data in cyber-attacks is elaborated on.
4. The fourth chapter 'Gathering Sensitive Information from Tenders' contains the conduction of a document analysis where information that is deemed sensitive is collected from tenders.
5. The fifth chapter 'Risks for Tendering Organizations' contains interviews with security professionals. Because of the way the interviews are structured, the professionals also collect information from tenders, after which they identify risks based on the information collected.
6. The sixth chapter 'Guidelines to Reduce Sensitive Information in Tenders' contains the guidelines that are established that help to realize a solution that can actually identify and reduce sensitive information in tenders. The guidelines are determined on the basis of information from the previous chapters, identified risks, a-like tools, and interviews with professionals involved in the tender process.
7. The seventh chapter 'Implementation of Guidelines through Automation' contains a description for the determined guidelines how they can be applied in practice. This chapter thus forms a concrete basis for the development of a final tool.
8. The eighth chapter 'Conclusion & Recommendations' reflects on the main goal of this thesis by drawing conclusions based on the conducted research. In addition, recommendations are given for follow-up research based on the guidelines.
9. The chapters that follow contain the references and the appendix.

2 Research Design Methodology

2.1 Opening

This chapter describes the underlying structure of this research: Design Science Research (DSR). Two theories related to DSR are discussed: the Design Science Research Methodology (DSRM) by Peffers, Tuunanen, Rothenberger & Chatterjee (2008) and the Three Cycle View of Design Science Research by Hevner (2007). These two theories contribute to the execution and the structuring of the research.

Although Design Science Research is not fully implemented because no solution is created, Design Science Research is still used as the underlying research methodology. This is chosen because the purpose of this research is to give recommendations in the form of guidelines, on the basis of which a tool can be created. For this reason, the steps prior to 'designing' are being followed. A follow-up research can then continue the research in a natural way following the DSR methodologies. Another reason is that the DSR methodologies also describe points of attention that are important for identifying the problem and determining the guidelines when the ultimate goal is to create a solution to an identified problem.

The DSR steps that are executed in this research are described in this chapter. The subsequent DSR steps that have not been carried out are briefly explained to get the complete DSR process clear for the reader.

The next paragraph describes the Three Cycle View of Design Science Research by Hevner (2007). The three cycles that are discussed *"must be present and clearly identifiable in a design science research project"* (Hevner, 2007). The third paragraph discusses the Design Science Research Methodology (DSRM) by Peffers, Tuunanen, Rothenberger, & Chatterjee (2008). Peffers presents a methodology for conducting design science research. The designed methodology consists of six phases that a researcher goes through when applying design science. The six phases are elaborated on. It is clearly stated to which stage of DSRM this research is being conducted. In the final paragraph an overview is given of the alignment between the Three Cycle View, DSRM, the research questions and the chapters in this report.

2.2 Three cycle view of design science research

Hevner (2007) argues that three specific cycles must be present in a design science research. In the figure below the research framework overlaid with the three cycles is displayed.

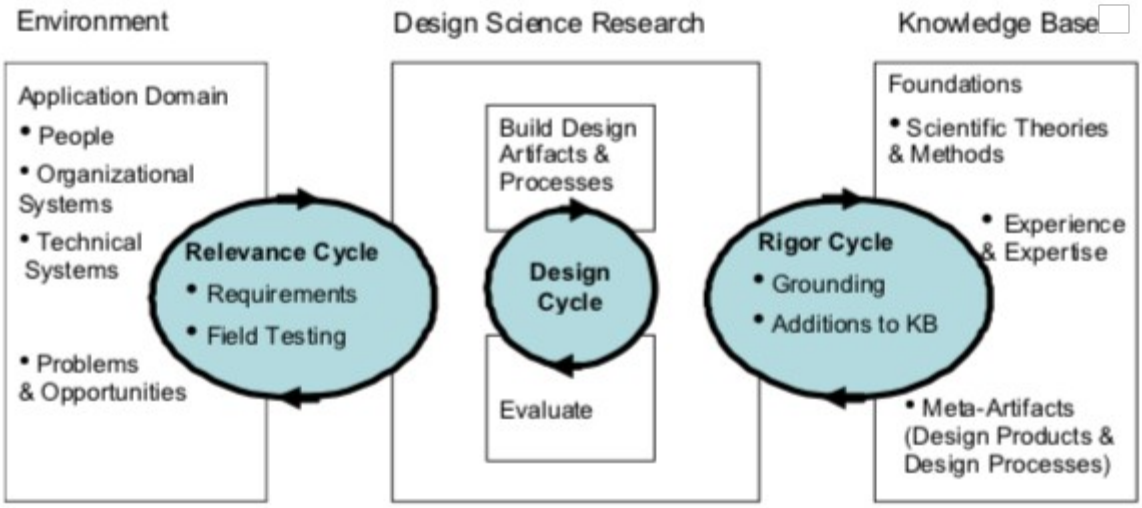

Figure 1: Three Cycle View by Hevner (2007)

Hevner explains the cycles as follows: *"The Relevance Cycle bridges the contextual environment of the research project with the design science activities. The Rigor Cycle connects the design science activities with the knowledge base of scientific foundations, experience, and expertise that informs the research project. The central Design Cycle iterates between the core activities of building and evaluating the design artefacts and processes of the research."* (Hevner, 2007)

2.2.1 The relevance cycle

Hevner (2007) notes that there is a certain desire to improve an environment that drives the motivation of design science research. The motivation is to introduce new and innovative artefacts and processes for building the artefacts. He states that design science research begins with identifying and representing opportunities and problems in an actual environment. In other words, *"the relevance cycles initiates design science research with an application context that not only provides the requirements for the research as inputs but also defines acceptance criteria for the ultimate evaluation of the research results. Does the designed artefact improve the environment and how can this improvement be measured?"* (Hevner, 2007).

This research starts with identifying and representing the problems and opportunities related to the information in tenders that can be misused. The application environment consists of the people, organizations, and systems involved in the tendering process. The risks that are identified and the guidelines that are established stem from these aspects of 'the application environment'. The 'acceptance criteria for the ultimate evaluation' are also the guidelines in some sort. The guidelines can be seen as requirements for a final solution. The guidelines can be used as acceptance criteria to evaluate if a final solution – an artefact – successfully solves the problem that tenders contain sensitive information.

2.2.2 The rigor cycle

The artefacts created by Design Science Research draw from a knowledge base of scientific theories and engineering methods. This knowledge base also contains the experiences and expertise that define the state-of-the-art in the application domain of the research and the existing artefacts and processes found in the application domain (Hevner, 2007). The knowledge base is consulted by the Rigor Cycle. Even though no final artefact is created, the knowledge base of scientific theories and engineering and design methods is consulted during the complete research process. Furthermore, the knowledge base is expanded by interviews with security experts, interviews with senior purchasers, and with new insights gained during this research.

2.2.3 The design cycle

The design cycle is about the design and development of artefacts. This is done by continuously designing, developing, and evaluating the designed artefact. The requirements that the artefact has to comply with stem from the relevance cycle. The methods that structure the development and design process, and the design and evaluation theories are drawn from the rigor cycle. Therefor it is important to state that the design cycle is highly dependent on the results of the relevance cycle and rigor cycle (Hevner, 2007).

Although DSR is the underlying methodology for this research, the design cycle is not present in this research. This has a valid reason, which has already been explained in the introduction to this study. Namely, this research aims to examine whether tenders contain sensitive information and then provide guidelines as recommendations that help create a tool that addresses the identified problem. In other words: this research results is the input for the design cycle. The guidelines that are determined can also be seen as the requirements that support the design cycle through the relevance cycle. The guidelines determine what the to be created artefact has to be capable of.

2.3 Design Science Research Methodology

The Design Science Research Methodology created by Peffers, Tuunanen, Rothenberger, & Chatterjee (2008) is visualized in the figure below. This model was created to help design science researchers to present research with reference to a commonly understood framework. The model presents a process sequence that a design science researcher goes through when conducting design science. The first two phases of DSRM are carried out in this study. The research result is the input for the third phase: Design & Development. This is also clearly indicated in the helicopter view in the following paragraph and in the figure below by the red square.

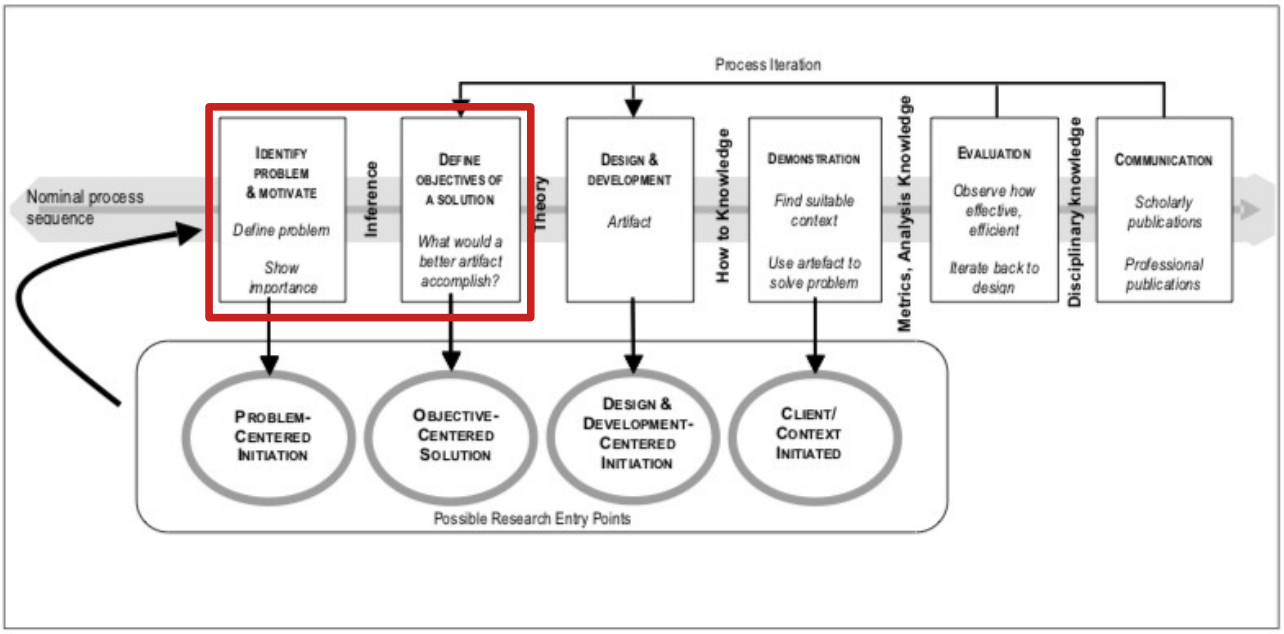

Figure 2: Design Science Research Methodology by Peffers et al. (2008). The first two phases are followed in this research indicated by the red rectangle.

The following section describes the entry point of this research. The subsequent sections describe the six phases of DSRM. The first two phases are present in this research and are therefor described in more detail.

2.3.1 Problem centred initiation

According to Peffers et al. (2008) there are multiple entry points a design science research can start. The four different entry points are visualized in the above figure.

This research starts from a problem-centred initiation. The problem has been lightly identified that public tenders may contain information that is relevant for adversaries. This is related to the problem that has already been proven, namely that open data has negative consequences such as misuse of public information by malicious parties. Therefor, the study starts form this entry point. Following the nominal process sequence the next phase is 'Identify Problem & Motivate'. In this phase the problem that is identified being the initiation is further elaborated on.

2.3.2 Identify problem & motivate

Chapter 1 'Introduction' highlighted the problem and the relevancy of the research. The goal of this phase however, is to study the specific research problem more in depth and justify the value of a solution. *"Justifying the value of a solution accomplishes two things: it motivates the researcher and the audience of the research to pursue the solution and to accept the results and it helps to understand the reasoning associated with the researcher's understanding of the problem. Resources required for this activity include knowledge of the state of the problem and the importance of its solution"* (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2008). This phase of Design Science Research Methodology is extensively addressed in this research. The first three research questions are intended to thoroughly investigate and substantiate the problem. The result is that the identified problem – that public tenders contain information that is relevant for adversaries in preparing cyber-attacks – is confirmed. This is also the motivation why the problem must be solved. As figure 2 points out, the objectives of a solution are inferred from the identified problem. In the perspective of this research this means that the guidelines that are determined are derived from the identified problem in this phase of the research. This is further elaborated on in the next section.

This phase of DSRM has a clear relationship with the Three Cycle View by Hevner (2007). The information needed for identifying and motivating the problem are in the 'Environment'. The environment is therefore frequently consulted. Methods and scientific theories are drawn from the 'Knowledge base' (Hevner, 2007) to conduct the research.

The first three research questions are answered in this phase of the research. Reasoning from the 'Environment' by Hevner (2007), the people, organizational systems, and technical systems aspects are consulted to answer the research questions. In other words: the tendering process, tenders themselves, literature on open data, and literature on hacker activities are thoroughly researched to determine whether there is a relationship between cyber-attacks and information in tenders, and to determine what information types adversaries use to prepare cyber-attacks. A document analysis is done in this phase to research whether these information types are present in public tender documents. In addition, interviews are held with security professionals with the aim of determining whether tenders contain sensitive information, and to identify risks that tendering organizations face as a result of the sensitive information that adversaries can gather and misuse. Professionals who deal with tenders on a daily basis are also interviewed. The purpose of these interviews is to determine whether they recognize that public tenders contain sensitive information and whether there are possibilities for a solution to identify and reduce this information from public tenders. The foundations to, for example, conduct this research, conduct the document analysis, and prepare the interviews are drawn from the 'Knowledge base' (Hevner, 2007). The new insights gained from this phase, such as the information types that adversaries search for and the link between these types and public tenders, are an addition to the 'Knowledge base'.

The chapters ‘Tenders as Reconnaissance Source’, ‘Gathering Sensitive Information from Tenders’, and ‘Risks for Tendering Organizations’, together form the ‘Identify Problem & Motivate’ phase.

2.3.3 Define objectives of a solution

The goal of this phase is to “*infer the objectives of a solution from the problem definition and knowledge of what is possible and feasible. The objectives can be quantitative, e.g., terms in which a desirable solution would be better than current ones, or qualitative, e.g., a description of how a new artefact is expected to support solutions to problems not hitherto addressed. The objectives should be inferred rationally from the problem specification. Resources required for this include knowledge of the state of problems and current solutions, if any, and their efficacy”* (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2008). This phase of Design Science Research Methodology is extensively addressed in this research. The objectives are qualitative: a description of how a new artefact is expected to support solutions to problems not hitherto addressed. The qualitative description – the determined guidelines of how a new artefact is expected to identify and reduce sensitive information in tenders – is the result of this study. The research ends with proposing the guidelines for further research. The resources required for determining the guidelines – the knowledge of the state of problems – are derived from the previous phase, where it is made clear what the relationship is between data in tenders and cyber-attacks, what information types can be gathered from tenders, and what the risks are that occur for tendering organizations when adversaries collect sensitive information from tenders. This phase results in the theory (figure 2) that is needed to start designing and developing the final solution. The fourth and fifth research question are intended to determine the ‘objectives of a solution’.

This phase of DSRM has a clear relationship with the Three Cycle View by Hevner (2007). For the establishment of the guidelines and making recommendations to implement the guidelines in a final solution, such as using machine learning for pattern recognition, the knowledge gained from the 'Environment' is used as input. The identified problem that public tenders contain sensitive information that can be misused by adversaries – otherwise expressed – can be seen as opportunities to solve the problem. From the 'Knowledge base' methods and scientific theories are used to conduct the research. Existing artefacts (tools) are also used to support the guidelines by taking examples of how these existing a-like artefacts are used in practice. In terms of Hevner (2007) the established guidelines are the requirements that a final solution should meet. The guidelines can therefor also be used as acceptance criteria to evaluate a final solution later on in follow-up research.

The last two research questions are answered in this phase of the research. The previous phase has resulted in a clear and identified problem, based on the 'Environment' aspects pointed out in the previous section (Hevner, 2007). In order to determine the 'objectives of a solution', we look at where the problems lie to turn them into possibilities: the determined guidelines. The risks that are identified and the causes that lead to these risks in the previous phase of the research are analysed in order to extract guidelines that aim to identify and reduce sensitive information in public tenders – to prevent the identified risks. In addition, the 'Knowledge base' (Hevner, 2007) is consulted. This in particular to further substantiate the guidelines by giving possible solutions how it can be applied in practice for each guideline. This is done by consulting meta-artefacts; existing artefacts that are used as examples. The new insights gained from this phase are an addition to the ‘Knowledge base’. The guidelines, together with the substantiation of how the guidelines can be implemented in practice by proposing different techniques, serve as the recommendations for further research to start designing a final solution, based on the guidelines.

The chapters 'Reducing Sensitive Information in Tenders' and 'Exploration of the Guidelines' together form the 'Define Objectives of a Solution' phase.

2.3.4 Important note for the reader

The research ends after the 'Define objectives of a solution' phase. The report will move on to the conclusions and the recommendations for follow-up research. As explained, in terms of Design Science Research, the guidelines are the requirements that are fed into the design cycle by the relevance cycle. The guidelines are the basis on which a final solution can be made.

The other DSRM-phases are briefly described in the sections below to clarify Design Science Research. The recommendations for further research are input for the next DSRM-phase: 'Design & Development'.

2.3.5 Design & development

The goal of this phase is to create the artefact. An artefact can be a construct, model, method, or new properties of systems. *"Conceptually, a design research artefact can be any designed object in which a research contribution is embedded in the design"*, according to Peffers, Tuunanen, Rothenberger, & Chatterjee (2008).

The guidelines established in this study should be used to support in building an artefact. In terms of the Three Cycle View (Hevner, 2007), the guidelines provide the Design Cycle with the requirements for the development of the artefact to the identified problem.

This phase is not conducted in this research.

2.3.6 Demonstration

The goal of this phase is to demonstrate the created artefact – the final tool – after it is designed in the previous phase. This can be done by an experimentation, simulation, case study, proof, or other appropriate activity (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2008). The guidelines that are created in this research are the acceptance criteria for the demonstration which have to be met.

This phase is not conducted in this research.

2.3.7 Evaluation

The goal of this phase is to observe and measure how effective and efficient the designed artefact supports a solution to the earlier defined problem. *"Conceptually, such evaluation could include any appropriate empirical evidence or logical proof. At the end of this activity the researcher can decide whether to iterate back to step three to try to improve the effectiveness of the artefact or to continue"*, according to Peffers, Tuunanen, Rothenberger, & Chatterjee (2008).

The guidelines proposed in this study can be used in the evaluation of the demonstration of the final tool. The guidelines are the acceptance criteria. On the basis of the acceptance criteria, for example, it can be assessed to what extent the guidelines (i.e. the solution directions) have been implemented.

This phase is not conducted in this research.

2.3.8 Communication

The goal of this phase is to "*communicate the problem and its importance, the artefact, its utility and novelty, the rigor of its design, and its effectiveness to researchers and other relevant audiences, such as practicing professionals, when appropriate. In scholarly research publications, researchers might use the structure of this process to structure the paper, just as the nominal structure of an empirical research process (problem definition, literature review,*

*hypothesis development, data collection, analysis, results, discussion, and conclusion) is a common structure for empirical research papers. Communication requires knowledge of the disciplinary culture" (*Peffers, Tuunanen, Rothenberger, & Chatterjee, 2008).

This phase is conducted in this research. No artefact has been created and not all the phases of DSR have been executed. The results of this research however are communicated through this thesis.

2.4 Overview

In the figure below the overview of this research is visualized. It is a representation of the Three Cycle View by Hevner (2007), the Design Science Research Methodology by Peffers et al. (2008), the determined research questions and the chapters in this thesis. The relationship that is visible shows that, although design science is not fully executed, a clear design science research is being conducted.

In every chapter that follows a short retrospective is given to the theoretical approach: how does the execution of this chapter relate to design science research.

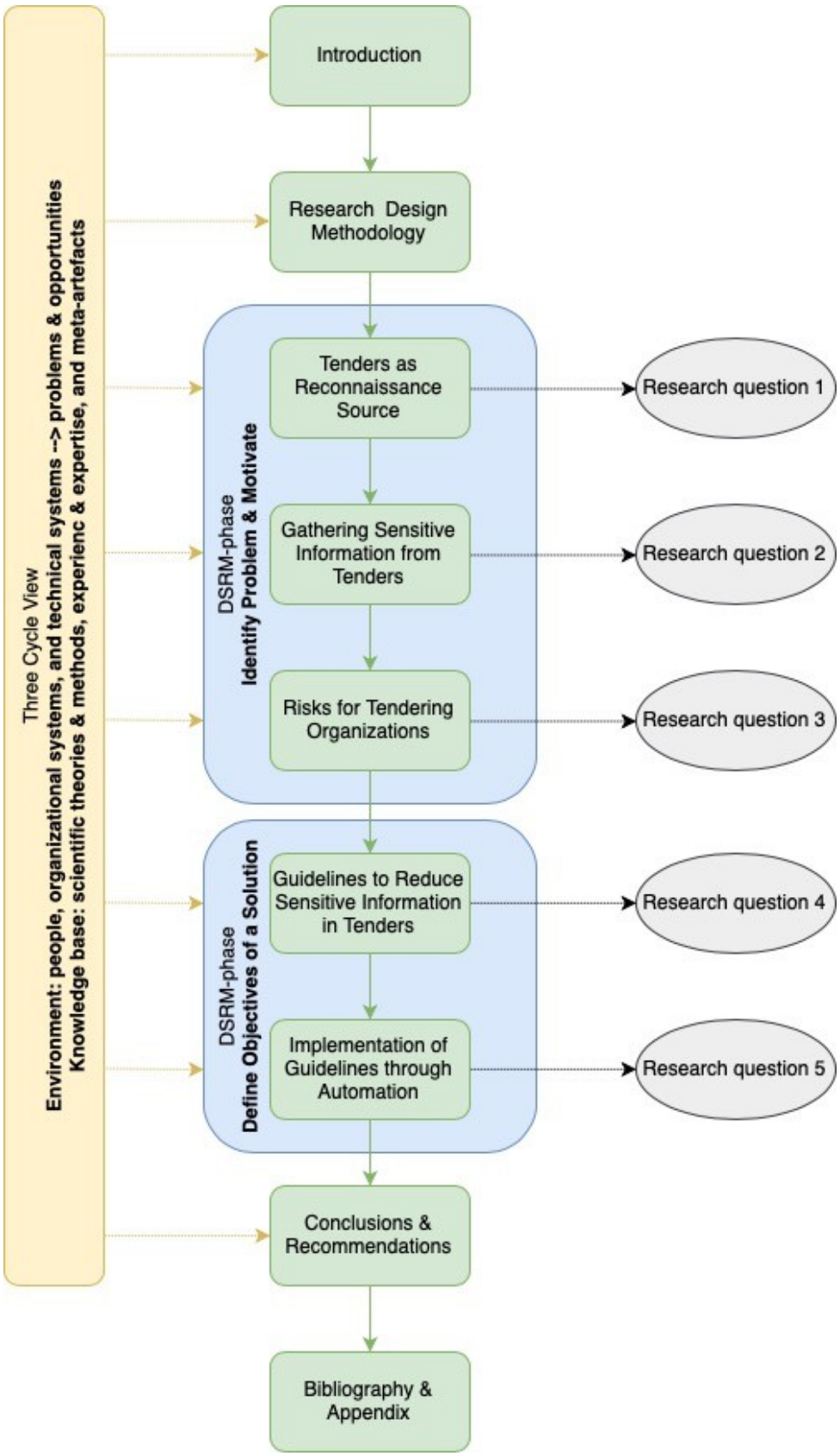

Figure 3: Alignment between Three Cycle View by Hevner (2007), the Design Science Research Methodology by Peffers et al. (2008), the determined research questions and the chapters in this thesis.

3 Tenders as Reconnaissance Source

3.1 Opening

The first chapter ‘Introduction’ stated that tenders can contain sensitive information. This problem is studied in-depth in this chapter. The goal of this chapter is to determine the relationship between the information that tenders contain and the information that is used in preparing cyber-attacks.

The following two paragraphs mention the research question that is answered in this chapter and how this is approached. The subsequent paragraphs contain the implementation thereof.

3.1.1 Research question

In this chapter the first research question is answered:

1. What is the relationship between the information in tenders and cyber-attacks?

This research question aims to determine the relationship between the information in tenders and cyber-attacks in order to identify the highlighted problem in chapter 1 of this study.

3.1.2 Approach

Literature study is mainly used to answer the first research question. The 'snowballing method' has been used for collecting literature, articles, and relevant information on websites (further: literature). Among others, the following search terms have been used on google scholar and university libraries: "tenders" "risks", "risks of open data", "open data", "reconnaissance", "open data" "reconnaissance", "cyber" "open data", and "cyber kill chain". In addition to whether literature turned out to be relevant, the publication year of the source found was also taken into account. Other literature have been derived from the sources found from these keywords. In addition, Dutch central government websites have been used for information about tenders and the MITRE website, which was known to the author, was used for information on hacker activities.

A start has been made with analysing information about tenders. In particular to clarify why the information in tenders is made public. The PIANOo website has been frequently consulted for this. Subsequently, to describe the relationship between data in tenders and cyber attacks, a start was made with analysing literature about the risks and disadvantages of open data. Literature has been selected that addresses the risks related to abuse of open data. This is because these risks then also inherently apply to 'open data' in tenders and therefore possibly affect contracting organizations. After making this connection, literature was studied on the activities of hackers before they actually carry out cyber-attacks. Information about the reconnaissance phase of the Cyber Kill Chain and information from the MITRE website about the techniques and tactics that hackers apply have been analysed. In particular, information regarding what information adversaries are looking for has been taken into account and included in the research. This so that it can be analysed later in the study whether this information is included in tenders to actually identify risks for contracting organizations. Of the information types that adversaries look for that have been determined from the literature, examples are analysed of how the data has been abused. In this way, literature study results in the relationship between data in tenders and cyber attacks.

In terms of design science research, the relevant information that comes from the literature is consulted from the 'Environment'. The identified problem also arises from this. The methods are provided from the 'Knowledge base' to answer this research question. Answering this research question contributes to the DSRM phase 'Identify Problem & Motivate'.

3.2 Information on tenders

Before delving in to the link between tenders and open data, the existence of tenders, the process of tenders, and the principles where tenders are based on, are explained in order to enhance the understanding of the subject.

3.2.1 A tender

Simple said, a tender refers to the process whereby governments (and other large institutions linked to a government) invite other organizations to bid on large projects whereof the government is the client. This happens for example when a government owned datacenter wants to obtain all hardware components for its datacenter. The needs of the datacenter are documented and the documents are made public to contractors. Contractors can then subscribe to a tender and have a chance of winning the tender and the order of all hardware.

As of the year 2014 the Dutch Government gave a meaning to the definition tender, which is: *"Tender is the purchase by public contract of works, supplies or services by one or more contracting authorities from economic operators chosen by these contracting authorities, irrespective of whether the works, supplies or services have a public destination or not.*"(translated from Dutch) (Europa decentraal, n.d.).

Investopedia defines 'Tender' as followed: "*For Public Sector companies, a tender refers to the process whereby institutions invite bids for large projects that must be submitted within a finite deadline. This is done according to both European and Local laws*" (Investopedia, 2019). Both definitions clearly are in line with the above given example.

3.2.2 Existence

In the year 1996 the Agreement on Government Procurement (further GPA) has come into force within the European Union. This agreement was made in relation to the World Trade Organization (further WTO). Countries that are part of the agreement must let each other take part into their home markets for public contracts under certain conditions. (PIANOo, n.d.)

The reason why the GPA was made is due to a rise in consciousness that an economic crisis is related with a collapse in international trade. This consciousness arose after the second world war. This philosophy was the start of the General Agreement on Tariffs and Trade in 1947, the precursor of the WTO. Organizations like WTO have the goal of enhancing the international trade by means of agreements such as the GPA. In 2011 the WTO-countries made a change to the GPA. The new GPA brought more clarity and transparency and ensured that the same terms are valid for all countries that take part in the GPA. (PIANOo, n.d.)

The European Union gave substance to the GPA in the European Tendering Guidelines of 2004 and to the new version of the GPA in the European Tendering Guidelines of 2014. The European Tendering Guidelines are meant for the members of the European Union, where the GPA is meant for countries that are outside the European Union but are part of the WTO, China for example. (PIANOo, n.d.)

As the European Tendering Guidelines are Directives, the members of the European Union have to translate the Directives into National laws. The Directives that are most important are:

- Directive 2014/24 for the awarding of public contracts,
- Directive 2014/25 for the awarding of contracts for water, energy, transport and postal services, and

- Directive 2014/23 for the awarding of concession contracts.

The European Directives are translated into the Dutch Law 'Aanbestedingswet 2012'. The revises in the European guidelines, as a result of the change in the GPA, have been adjusted in the 'Aangepaste Aanbestedingswet 2012' as of the year 2016.

3.2.3 Tender process

When an organization needs a certain product, a form of service or an architectural or civil engineering work, this is translated into an assignment. The client determines what must be purchased, the market is oriented, the purchasing strategy and purchasing conditions are determined. It is determined which procurement rules are at issue, whether there is a need to tender, which procurement procedure is applied and the choice for the award criterion to be applied is made. A decision is also made about any suitability requirements that the client wishes to set to the contractor (supplier). The client determines how he wants to specify the assignment and how he may involve the market in this. The client goes through a number of steps in the preparation of the assignment. (PIANOo, n.d.a)

Public organizations must place assignments on the market according to certain procedures. Parties can then register for the assignment. The contract is awarded to the tenderer whose tender best fits the requirements of the client. The agreements that are made with the winner and the client are set out in an agreement. This entire process is referred to as 'tendering'. (PIANOo, n.d.a)

If a contract exceeds a certain threshold value, it must follow the European procurement rules. The threshold values differ per type of assignment and are not relevant to this study.

The prescribed procedures are followed by the client whereby the principles of objectivity, transparency, proportionality and non-discrimination are applied. The principles relevant to this study are explained in detail later. (PIANOo, n.d.a)

The following steps are followed during the process according to PIANOo: 1. Announcing the tender, 2. Applying on the tender, 3. Selecting contractors, 4. Registering to the tender, 5. Awarding the tender to an organization, and 6. Completing the process. The first step is relevant for this study because here information on tenders is made public. (PIANOo, n.d.b)

(European) procurement procedures start with an announcement. This tells the market that a client is going to place an assignment for which entrepreneurs are asked to register. This announcement takes place on TenderNed.nl. In principle, this announcement only applies to contracts above the European procurement thresholds. Nevertheless, orders below the threshold are also announced to stimulate the market, for example. Contracts via the European procedure are published via TenderNed in the Official Journal of the European Union on TED (Tenders Electronic Daily). TED offers providers in 23 languages free access to potential assignments.

During the procedure, information is made public[1] for the market so that the market can best orient itself on the assignment. It is noted by PIANOo that clients do not disclose sensitive information from or about third parties. Sensitive information from the client itself is not spoken about. A statement on information with respect to information security is also not noted. PIANOo also states that detail requirements for the contract must be clearly stated so that all tenderers have equal access. (PIANOo, n.d.c)

1 In this context, public means that anyone with an account on TenderNed and/or TED can view the announcement. An account can be created by anyone with a valid e-mail address. This is accounted for as public and open in the context of this study.

3.2.4 Tender principles

The rules for tendering are based on four principles: non-discrimination, equal treatment, transparency, and proportionality. In short, they mean the following (PIANOo, n.d.):

- Non-discrimination: a distinction based on nationality may not be made.
- Equal treatment: discriminatory factors must not be present in a tender, so that one party has more chances than another party. All parties must be treated objectively and in the same way and everyone will receive the same information.
- Transparency: With a description that can be read by everyone, it must be clear in advance what is expected.
- Proportionality: Specifications, requirements and criteria must be relevant and proportionate to the nature and scope of the assignment.

It can therefore be deduced from the above that the information is made public on the basis of the four principles. In order to inform tenderers in advance about the assignment as well as possible clients, the contractor discloses many details publicly. The transparency and the proportionality principles may be the underlying reason *if* sensitive information is published in tenders. It might be a balancing act between just giving enough information for contractors to orientate, and releasing too much information which might be sensitive.

3.3 Risks of open data relate to information in tenders

There are certain similarities between the concept of open data and the principles where tenders are based on. To define open data two definitions are considered:

- Open Knowledge Foundation (OKF): "*is a global non-profit organisation focused on realising open data's value to society by helping civil society groups access and use data to take action on social problems*" (Open Knowledge Foundation, n.d.). OKF defines open data as: "*Open data is data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and share alike*" (Open Knowledge Foundation, n.d.a).
- The European data portal (EDP) "*gathers and collects the metadata of government information that is available on the public data portals of European countries. It also provides information regarding the provision of data and the benefits of its reuse*" (European Data Portal, 2016). EDP defines open data as: "*Open data is data that anyone can access, use and share. Governments, businesses and individuals can use open data to bring about social, economic and environmental benefits*" (European Data Portal, n.d.).

Both definitions state that the data can be accessed and used by everyone, which is the case with tender data. According to the definition of EDP open data benefits governments, business and individuals. Taking the data in tenders into account this is the case. Making the data transparent, treating all equal and be it proportional, governments, businesses and individuals benefit from the tender data.

While the openness of data has many advantages for the tenderers and the client it may also have disadvantages. Multiple studies have raised awareness for the possible disadvantages. Open data mainly gets defined from the perspective that people have no malicious intentions. Examples are the definitions of the Open Knowledge Foundation and the European Data Portal given above. From these definitions it can be deducted that open data can be accessed, used (for whatever purpose?) and shared by anyone. Plus, that is has certain benefits. If we take these definitions into

account from the perspective of hackers, they suddenly have a completely different meaning. Hackers are also able to access, use, and share open data; therefor, open data may also benefit their activities. OKF also states that an important aspect of open data is that "*the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets*" (Open Knowledge Foundation, n.d.). It is unlikely that hackers value and honour this aspect. Adversaries will not care if the data is provided under these terms.

A paper by Zuiderwijk & Janssen (2014) notes that open data on the one hand has benefits (transparency, strengthening the economic growth) but also stresses that the risks and disadvantages, like privacy violation and possible misuse and misinterpretation, should be taken into account when it comes to open data. Another study by Labohm (2018) states that governments are increasingly providing open data. An analysis made in this study shows that this increase has beneficial effects for social engineering activities.

In view of the above information, it can be argued that the risks identified in previous studies that arise from the disclosure of data also apply to tenders. Later in the study, an analysis is made of the specific risks that arise from public data in tenders. The disadvantages and risks in this chapter may occur.

Two questions arise from the aforementioned facts, which are answered below: how do hackers collect public data, and which public data has previously been abused by hackers?

3.4 Collecting information to prepare a cyber-attack

To understand how hackers prepare and execute their attacks, we briefly discuss the phases that hackers go through. We then go deeper into the phase where hackers collect data in order to obtain as much information as possible about their targets. Then we will discuss the tactics, techniques and procedures (TTP's) that hackers use.

3.4.1 The reconnaissance phase

The Cyber Kill Chain by Lockheed Martin (2015) (further CKC) is a frequently used model to perform a structured analysis on hackers' methods. This model with its phases is visualized below.

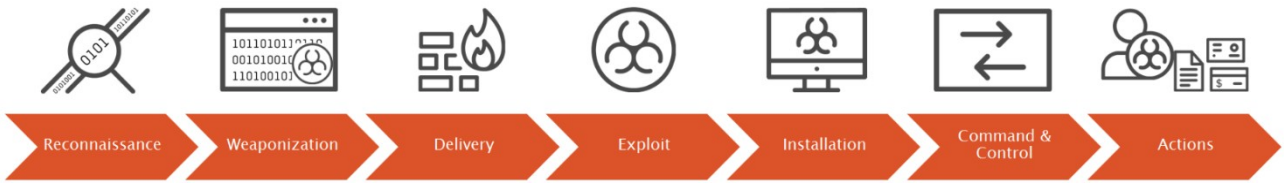

Figure 4: Cyber Kill Chain by Lockheed Martin (2015)

Lately however, the CKC has been widely criticized. Because of this other Kill Chains have been developed with other phases. In a research by Pols (2017) different Kill Chains are compared and a new 'Unified Kill Chain' has been designed. The Unified Kill Chain and the original kill chains all start with the reconnaissance phase. Thus, it is safe to say that hackers start their activities with reconnaissance. In the CKC, the reconnaissance phase is defined as: *"Researching, identifying and selecting targets. This may consist of passive reconnaissance of open source intelligence or active reconnaissance where internet facing systems are probed for potential weaknesses"* (Pols, 2017). As can be derived from the definition, this study focusses on passive reconnaissance; gathering open data that can be used in preparing attacks. In other words, hackers collecting sensitive information from tender documents.

3.4.2 PRE-ATT&CK framework

MITRE is a not-for-profit organization that aims to tackle problems that challenge safety, stability and well-being. They provide innovative and practical solutions in the cyber security sector, among others. MITRE states that "*a comprehensive security plan does not begin or end at the perimeter, but instead leverages an understanding of the full lifecycle of a cyber adversary*" (MITRE, n.d.). They note that the pre-attack activities of hackers are executed mostly outside the enterprises field of view. The wealth of information on the internet is being used to take advantage over the defenders. Organizations have to realize that the defence of the organization stretches outside of the organization's logical perimeters. An aspect of the cyber defence therefor is to minimize public sensitive information.

MITRE developed the ATT&CK framework for the private sector, government, and in the cybersecurity product and service community to aid the development of threat models in order to assist companies' cyber defence. The ATT&CK framework is a knowledge base of adversary tactics and techniques based on real-world observations (MITRE, n.d.). In relation to the CKC, the ATT&CK framework focusses on the phases Delivery to Action on objectives. After the development of the ATT&CK framework MITRE produces the PRE-ATT&CK framework. MITRE defines this framework as: "*PRE-ATT&CK provides the ability to prevent an attack before the adversary has a chance to get in. The 15 tactic categories for PRE-ATT&CK were derived from the first two stages (recon and weaponize) of a seven-stage Cyber Attack Lifecycle [2] (first articulated by Lockheed Martin as the Cyber Kill Chain® [3]). This framework captures the tactics, techniques, and procedures adversaries use to select a target, obtain information, and launch a campaign. The framework lists the ways that adversaries perform each tactic and provides the ability to track and organize adversary statistics and patterns*" (MITRE, n.d.). The figure below visualizes how the tactics are related to the reconnaissance phase.

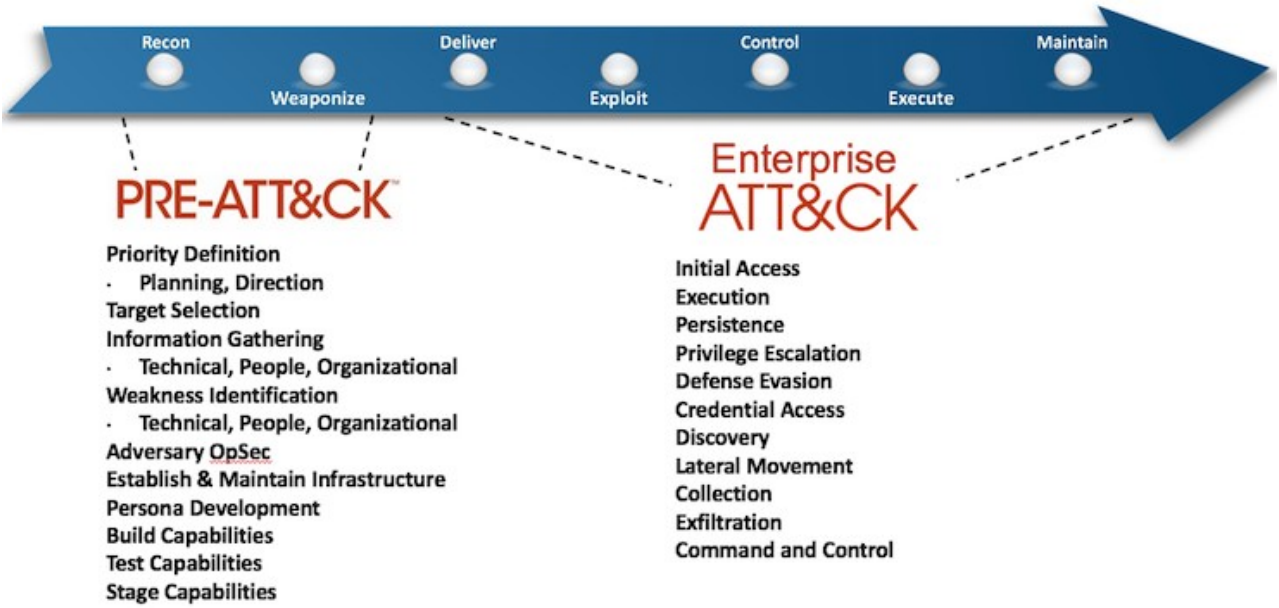

Figure 5: Relation between the tactics and the reconnaissance phase (MITRE, n.d.).

3.4.3 PRE-ATT&CK tactics

For this study the tactics under 'Recon' are of importance, because these are the tactics where open data is obtained. The 'Weaponize' phase is the next phase, where this data is actually being used to prepare a cyberattack. The table below explains the PRE-ATT&CK tactics that are defined by

MITRE and are relevant for this study. Annex A contains the remaining tactics with the motivation why they are not relevant for this study.

Table 1: Relevancy of the PRE-ATT&CK Tactics

PRE-ATT&CK Tactic	Description (MITRE, n.d.)	Motivation relevancy
Technical Information Gathering	*"Technical information gathering consists of the process of identifying critical technical elements of intelligence an adversary will need about a target in order to best attack. Technical intelligence gathering includes, but is not limited to, understanding the target's network architecture, IP space, network services, email format, and security procedures."*	The study focusses on the gathering of open data. This includes technical information.
People Information Gathering	*"People Information Gathering consists of the process of identifying critical personnel elements of intelligence an adversary will need about a target in order to best attack. People intelligence gathering focuses on identifying key personnel or individuals with critical accesses in order to best approach a target for attack. It may involve aspects of social engineering, elicitation, mining social media sources, or be thought of as understanding the personnel element of competitive intelligence."*	The study focusses on the gathering of open data. This includes people information.
Organizational Information Gathering	*"Organizational information gathering consists of the process of identifying critical organizational elements of intelligence an adversary will need about a target in order to best attack. Similar to competitive intelligence, organizational intelligence gathering focuses on understanding the operational tempo of an organization and gathering a deep understanding of the organization and how it operates, in order to best develop a strategy to target it."*	The study focusses on the gathering of open data. This includes organizational information.

The tactics technical information gathering, people information gathering, and organizational information gathering are relevant for this study. All three tactics are about gathering information needed for cyber-attacks. From these tactics it can also be concluded that different types of information are relevant for malicious parties. Most people think that only technical information is relevant to prepare a cyber-attack, but also information about the organization and its employees seems of importance. It is abundantly clear that public sources are used for this. As can also be determined from the techniques that belong to these tactics. Techniques provide a more detailed picture of the attacker's activities than the tactics. The tactics indicate at a high level what the attacker wants to achieve with his work, where a technique describes what the adversary actually does.

3.4.4 PRE-ATT&CK techniques

Forty-two techniques are identified by MITRE in the PRE-ATT&CK framework that are part of the three relevant tactics. Not all of these are relevant to delve into in the context of this study. *It is important to emphasize that this research focuses on 'passive reconnaissance'; the collection of public information. 'Active reconnaissance'; whereby the attacker actively uses tooling aimed at the targets' network is not taken into consideration.*

The table below describes the techniques which are determined relevant for this study, under which tactic they belong and why they are relevant.

Table 2: Relevancy of the PRE-ATT&CK Techniques

PRE-ATT&CK Technique	Belongs to tactics	Description (MITRE, n.d.)	Relevancy
Acquire OSINT data sets and information	Technical Information Gathering People Information Gathering Organizational Information Gathering	*"Open source intelligence (OSINT) is intelligence gathered from publicly available sources. This can include both information gathered on-line, such as from search engines, as well as in the physical world."*	With OSINT hackers can search tender documents for relevant information.
Determine 3rd party infrastructure services	Technical Information Gathering Organizational Information Gathering	*"Infrastructure services includes the hardware, software, and network resources required to operate a communications environment. This infrastructure can be managed by a 3rd party rather than being managed by the owning organization."*	If information about 3rd party infrastructure services are included in tender documents they are searchable for adversaries.
Determine domain and IP address space	Technical Information Gathering	*"Domain Names are the human readable names used to represent one or more IP addresses. IP addresses are the unique identifier of computing*	If domain and IP address space are included in tender documents, they

		devices on a network. Both pieces of information are valuable to an adversary who is looking to understand the structure of a network."	are searchable for adversaries.
Determine external network trust dependencies	Technical Information Gathering	*"Network trusts enable communications between different networks with specific accesses and permissions. Network trusts could include the implementation of domain trusts or the use of virtual private networks (VPNs)."*	If external network trust dependencies are included in tender documents, they are searchable for adversaries.
Determine firmware version	Technical Information Gathering	*"Firmware is permanent software programmed into the read-only memory of a device. As with other types of software, firmware may be updated over time and have multiple versions."*	If firmware version information is included in tender documents, they are searchable for adversaries.
Discover target logon/ email address format	Technical Information Gathering	*"Email addresses, logon credentials, and other forms of online identification typically share a common format. This makes guessing other credentials within the same domain easier. For example, if a known email address is first.last@company.com it is likely that others in the company will have an email in the same format."*	If logon/email address formats are included in tender documents they are searchable for adversaries.
Enumerate externally facing software applications technologies, languages, and dependencies	Technical Information Gathering	*"Software applications will be built using different technologies, languages, and dependencies. This information may reveal vulnerabilities or opportunities to an adversary."*	If externally facing software applications technologies, languages, and dependencies are included in tender documents they are searchable for adversaries.

Identify security defensive capabilities	Technical Information Gathering	*"Security defensive capabilities are designed to stop or limit unauthorized network traffic or other types of accesses."*	If security defensive capabilities are included in tender documents, they are searchable for adversaries.
Identify web defensive services	Technical Information Gathering	*"An adversary can attempt to identify web defensive services as CloudFlare, IPBan, and Snort. This may be done by passively detecting services, like CloudFlare routing, or actively, such as by purposefully tripping security defences."*	If web defensive services are included in tender documents, they are searchable for adversaries.
Map network topology	Technical Information Gathering	*"A network topology is the arrangement of the various elements of a network (e.g., servers, workstations, printers, routers, firewalls, etc.). Mapping a network allows an adversary to understand how the elements are connected or related."*	If network topology information is included in tender documents, they are searchable for adversaries.
Identify groups/roles	People Information Gathering	*"Personnel internally to a company may belong to a group or maintain a role with electronic specialized access, authorities, or privilege that make them an attractive target for an adversary. One example of this is a system administrator."*	If information on groups/roles is included in tender documents they are searchable for adversaries.
Determine centralization of IT management	Organizational Information Gathering	*"Determining if a "corporate" help desk exists, the degree of access and control it has, and whether there are "edge" units that may have different support processes and standards."*	If information on the centralization of IT management is included in tender documents, they are searchable for adversaries.

Determine physical locations	Organizational Information Gathering	*"Physical locality information may be used by an adversary to shape social engineering attempts (language, culture, events, weather, etc.) or to plan for physical actions such as dumpster diving or attempting to access a facility."*	If information on physical locations are included in tender documents, they are searchable for adversaries.
Identify business processes/tempo	Organizational Information Gathering	*"Understanding an organizations business processes and tempo may allow an adversary to more effectively craft social engineering attempts or to better hide technical actions, such as those that generate network traffic."*	If information on business processes/tempo is included in tender documents they are searchable for adversaries.

The above table shows that adversaries gather a wealth of information that might be publicly accessible for everyone, without actively aiming at the target its network with the risk of setting off alarm bells. The techniques thereby clearly indicate what kind of information adversaries are looking for during the reconnaissance. The techniques are therefore seen in this study as the information types that adversaries are looking for. The techniques also clearly indicate that adversaries have many options at their disposal that help them with this. A paper by Sanghvi & Dahiya (2013) on Cyber Reconnaissance also lists information types that are gathered by adversaries. This lists clearly shows a comparison with the MITRE techniques. Sanghvi & Dahiya (2013) notes that passive reconnaissance is also known as 'foot printing': minimizing the interaction with the target which generates alert in logs. With the goal of gathering information on the victim without alerting the victim. An example given in the paper has a resemblance with the subject of this thesis. *"If any company puts an advertisement for opening for a system administrator and asking proficiency in Linux environment then it can be inferred easily that this company use Linux environment widely"* (Sanghvi & Dahiya, 2013). It shows that data which is meant to be open can be relevant for an adversary.

3.4.5 Previously misused data in recent cyber-attacks

Since the tactics and techniques in the PRE-ATT&CK framework of MITRE are based on real world examples it can be concluded that the above-mentioned information-needs give a realistic view. This is also confirmed by reports of cyber-attacks that have happened.

In his study Pols (2017) shows how a Red Team compromised a critical supporting ICT asset from an unauthorized and external perspective. The visualization shows how public sensitive information was misused for phishing mails and social engineering. The report states that the target organization was passively researched through open source intelligence and that the information gathered was used in preparation of further attacks. The report mentions domain and contact information. In the next phase of the Cyber Kill Chain, Weaponization, the Red Team registered domain names that were notably similar to the primary domains collected during the reconnaissance. Afterwards, phishing mails were sent to the employees whose e-mails were found

during reconnaissance. In the end, the Red Team gained domain administrator privileges to any workstation, server and account that is part of the primary domain.

Another case is the well-known cyber attack Stuxnet. There are various theories about how the attack was set up and what was done in terms of preparation. New theories were recently published, stating that the Netherlands was also involved. In any case, it is clear that a lot of public information has been collected about the components used in the nuclear power plant that has been attacked. Examples of public information that was misused given in reports are:

- Behpajooh website mentioning it had installed Siemens S7-400 Programmable Logic Controllers, as well as the software and communications systems. These components were specifically targeted by Stuxnet. Behpajooh had been linked to Iran's illegal procurement activities by the US federal court earlier. (Wired, 2017)
- The Iranian Government published photographs of the Iranian President observing monitors at the Natanz uranium enrichment plant in Iran. It is suspected that these photographs were used in preparing and building the worm Stuxnet, as the photographs clearly show a part of the plants' setup.

Again, these examples show that different information types are being misused by adversaries. In addition, it is very likely that hackers abuse more types of information than is known. Many attacks are not detected, which means that no forensic investigation is taking place. In addition, the types of information that hackers collect and abuse will most likely increase because the information on the internet is also increasing.

3.5 Summary of 'Tenders as reconnaissance source'

The goal of this chapter was to determine the relationship between the information that tenders contain and the information that is used in preparing cyber-attacks. The links established in this chapter show that risks arising from open data are inherently risks for information in tenders because this information is public due to the principles on which tenders are based. It has also been demonstrated, using the MITRE techniques and the examples of previous cyber attacks, that adversaries use public information when preparing and executing cyber attacks. Theoretically it can be stated that tenders are a possible source for reconnaissance activities.

We can conclude that:

- Tender information is public due to the principles on which tenders are based.
- Risks of open data such as abuse of information are inherently risks for data in tenders.
- Adversaries gather public information during the reconnaissance phase.
- Tenders are a source of information relevant for adversaries preparing cyber-attacks.

After a cyber-attack is it difficult to determine where the adversaries gained the information from they used during the attack. Adversaries use multiple sources for reconnaissance, so the same information can also be found in different places.

4 Gathering Sensitive Information from Tenders

4.1 Opening

The previous chapter has resulted in the theoretical proposition that tenders can be used as a source for the reconnaissance activities of adversaries. The goal of this chapter is to determine if sensitive information can be gathered from public tenders and if so, to show it in practice.

The following two paragraphs mention the research question that is answered in this chapter and how this is approached. The subsequent paragraphs contain information on open source intelligence, describe the approach of the analysis to gather information from tenders, and the results of this analysis. The final paragraph summarizes this chapter.

4.1.1 Research question

In this chapter the second research question is answered:

2. Which information that is relevant for adversaries can be collected from public tenders during the preparation of cyber-attacks?

This research question aims to determine which information can be gathered from tenders by doing a document analysis where it is tried to gather sensitive information from tenders.

4.1.2 Approach

The literature study primarily focused on the topic of Open Source Intelligence. This from the perspective how OSINT can be gathered and used for specific purposes; therefor a book was primarily used that discusses OSINT-techniques. Other articles on OSINT have been studied in terms of definitions and facts. Furthermore, literature on how a document analysis should be structured is used for executing the document analysis. The method of O'Leary (Triad, 2016) was chosen because it was easy to use and it covered document analysis from a broad perspective. The document analysis itself is executed using existing tenders. The MITRE techniques are considered the information types that adversaries search for during reconnaissance. A manual analysis is done by comparing information in tenders with the information types. The process was administered in Excel. The steps of the document analysis are explained in detail in chapter 4.4.

In terms of design science research, the literature on document analysis comes from the 'Knowledge base'. The tenders that are analysed and the aspects on the basis of which they are analysed come from both the 'Environment' and the 'Knowledge base'. Answering this research question contributes to the DSRM phase 'Identify Problem & Motivate'.

4.2 Open source intelligence

OSINT is not a new phenomenon from the internet era. Public information has always been used by individuals, companies and governments for their own gain. During the Second World War, resistance fighters listened to the radios to gather information to plan and improve their own activities. Companies analyse news reports about competitors to adjust their own strategies and business operations. Investors analyse news articles and economic articles to determine their own course and strategies. However, due to the Internet the amount of information has grown tremendously and the accessibility has become easier. In an article, Marr (2018) notes that 2.5 quintillion bytes of data is created each day at the current pace and according to Internet World Stats (2019), 58.8% of the world population uses the Internet. This indicates that the amount of data publicly available still has room for large growth.

A book on open source intelligence techniques, by Bazzel (2018) describes 'What is OSINT' very nicely. The book is called 'Open Source Intelligence Techniques: resources for searching and analysing online information'. The author spent 18 years as a government computer crime investigator, focussing on open source intelligence. His book is used by government agencies as training manuals for intelligence gathering and securing personal information. He states that OSINT can mean many things to many people, depending on who is asked and in what context. And with this context in mind he describes OSINT. "*Officially, it is defined as any intelligence produced from publicly available information that is collected, exploited and disseminated in a timely manner to an appropriate audience for the purpose of addressing a specific intelligence requirement. For the CIA, it may mean information obtained from foreign news broadcasts. For an attorney, it may mean data obtained from official government documents that are available to the public. For most people, it is publicly available content obtained from the internet*" (Bazzel, 2018).

A study on OSINT and OSINT applications defines OSINT as: "*all information and knowledge that can be gathered from publicly available sources*" (Tuominen, 2019).

In the context of this study, OSINT means 'publicly available content obtained from the internet'. Where publicly available in this study means: information accessible by registering without a creditability or identification check. The internet in this study is narrowed down to websites that contain tender information. This is also in line with the definition proposed by Tuominen (2019): where 'publicly available sources' refers to 'publicly available content'.

The question with OSINT is however, how do adversaries gather this information? While this can be done manually: scrolling through websites, reading articles and cherry-picking what may be useful, this can be done more effective and efficiently. Tools exist for open source intelligence gathering that are used by government bodies (military, security, law enforcement), international organizations, business corporations, penetration testers, hackers, criminal organizations, terrorist groups, and privacy-conscious people (Tuominen, 2019). On top of the existing tools, tasks can also be automated. The usage of tools and scripting manual tasks is outside of the scope of this study.

4.3 Document analysis method

This section describes how the document analysis is approached. After that, the results are discussed.

The gathering of information from tenders is done manually using qualitative document analysis in the context of this research. *Keep in mind that hackers might do this differently, using certain tools and scripts. The goal however is the same: gathering as much sensitive information from public tenders as possible.*

O'Leary outlines 8 steps that should take place in document analysis (Triad, 2016). Below it is explained how these 8 steps fit in the document analysis that is done in this research:

1. *Create a list of texts to explore*: the list of texts during the experiments are texts that originate from tenders on TenderNed. The list is a result of a search query build with ICT-related keywords.
2. *Consider how texts will be accessed with attention to linguistic or cultural barriers*: texts are accessed via the internet. Credentials used are that of the author of this study. Linguistic or cultural barriers do not occur.

3. *Acknowledge and address biases:* there are no biases. During the document analysis the text will be looked at from a hackers perspective; meaning that everything will be analysed with the question 'what can a hacker do with this?'.
4. *Develop appropriate skills for research*: no advanced skills are required. During the document analysis the texts are analysed using the information types from table 2 first column (chapter 3.3.3). These are the precise information types that are searched for in the documents.
5. *Consider strategies for ensuring credibility*: credibility is ensured by saving all tender documents used in the experiment. Also, interviewees with a security background will analyse the same tenders without knowing the information types beforehand. If the outcome of the interviews is approximately the same, credibility is also achieved.
6. *Know the data one is searching for*: During the document analysis the texts are analysed using the information types from table 2 first column (chapter 3.3.3). These are the precise information types that are searched for in the documents.
7. *Consider ethical issues*: no confidential information can be accessed. Only public tender information. The sensitive information that is gathered however, is not pasted 1 to 1 in this thesis due to security risks for the tendering companies, which is the whole purpose of this study.
8. *Have a backup plan*: security experts will be interviewed. If this document analysis has no results, the backup plan is that the security experts might gather something.

O'Leary describes the 'interview technique' as a technique to explore the documents (Triad, 2016). The texts are treated as a respondent or informant that provides the researcher with relevant information. During the experiment "questions" are "asked" to the document (f.e. searching in documents on the keyword "VLAN" or "Switch"). The frequency and amount of occurrences within the documents of what is searched for is organized. (Triad, 2016)

4.4 Document analysis execution

The following steps are executed:

1. As has been described earlier the tenders on https://www.tenderned.nl are protected by a login mechanism using only username & password. *Everyone can make an account, so it can hardly be called a preventive mechanism.* Step 1 is logging into https://www.tenderned.nl

Figure 6: Screenshot of the login screen on https://www.tenderned.nl (in Dutch).

2. After logging in, a search query is used to search for tenders. The keywords used in the query are all IT-related because it is assumed that adversaries are mainly interested in IT-

related information (deducted from table 2 first column (chapter 3.3.3)). The keywords are also chosen because they are objects that organizations have to purchase mainly by tendering. The following query is used: *"applicatie" "infrastructuur" "systeem" "hardware" "software" "ict" "it" "security" "tools"*. The query gives 60 results.

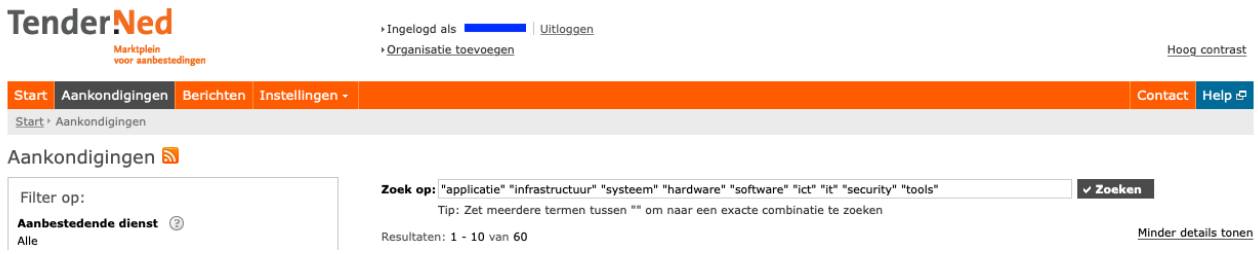

Figure 7: Screenshot of the search query used on https://www.tenderned.nl (in Dutch).

Other keywords have been searched but did not yield more results. This can be deduced from the number of search results in the figure below. This is also 60. The figure shows that more keywords are searched. Note that new tenders can be added daily and existing tenders have an expiration date. Replaying the document analysis is therefore difficult.

Figure 8: Screenshot of a search query with more keywords on https://www.tenderned.nl. No new results are gained. Amount of results is still 60. (In Dutch)

3. The resulting tenders and related documents are analysed in order. This is done manually by reading the tender documents and correlating the information with specific information types. These information types are the MITRE-techniques as explained in the previous chapter. When a tender that is being analysed adds no new information it means that saturation has been achieved. This process is tracked in Excel and described in the next paragraph. Tenders that show up as a result from the query but have no clear relation with the keywords are not examined.

4.5 Document analysis results

A summary of the results is visualized in the table below. The results of the search query (documents associated with the tenders) have been downloaded. Subsequently, the analysis of the tender documents was started in ascending order of the search results. Tenders that are not related to the keywords but that appear in the search result have not been analysed. The same applies to tenders that do not have any documents attached to them or only the standard mandatory documents associated with a tender. Of the 60 search results, between 15 and 20 tenders were assessed whether they were usable for analysis. Of these, it became apparent after eight analysable tenders that saturation had been achieved. *Exactly which tenders are analysed is not included in this report because sensitive information about the contracting organizations is then indirectly disclosed. Details can be viewed upon request.* After this amount no new information was gained because it showed that every analysed tender possessed sensitive information. Every new tender that was analysed contained information that could be linked to the information types that

adversaries search for. The table shows the MITRE techniques, translated into types of information that malicious parties seek during the reconnaissance activities. This is done because the MITRE techniques clearly indicate what types of information adversaries seek for when executing the technique. This is also mentioned in table 2 column 4. The analysed tenders are listed in the columns that follow indicated by 1 to 8. The analysis is done manually. Information related to the information types was searched in the tender documents for the specific tender that was analysed. This was done by studying the table of contents of the documents, reading relevant chapters and searching for specific keywords such as IP address, version, network, service desk, map, process diagram, domain, VLAN, switch, router, firewall, process, et cetera. As soon as information was found in a document, a *V* was listed in the table below. In addition, a separate table has been kept for each tender analysed, whereby the columns do not list the various tenders, but the documents pertaining to 1 tender. It may therefore be that different information has been found several times related to a specific information type under 1 tender, but in different documents.

Table 3: Overview of the document analysis results

		Tender #							
	Information type searched for in the tender	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	3rd party infrastructure services	*V*			*V*		*V*	*V*	
2	domain and IP address space				*V*				*V*
3	external network trust dependencies				*V*	*V*	*V*	*V*	
4	firmware version				*V*		*V*		
5	target logon/email address format	*V*	*V*	*V*		*V*	*V*	*V*	
6	externally facing software applications technologies, languages, and dependencies				*V*	*V*	*V*		*V*
7	security defensive capabilities		*V*		*V*		*V*		*V*
8	web defensive services						*V*		
9	network topology	*V*	*V*		*V*	*V*	*V*		*V*
10	groups/roles						*V*	*V*	
11	centralization of IT management				*V*				
12	physical locations	*V*	*V*		*V*				*V*
13	business processes/tempo	*V*	*V*			*V*	*V*	*V*	

The results clearly show that sensitive information can be found in tenders. All analysed tenders apparently contain sensitive information that can be related to the information types described in the second column. As such, it can be stated that adversaries can collect information from tenders that is relevant to them in preparing cyber-attacks.

To emphasize the sensitivity of information found, the following sections provide a number of examples of information found and how this information is relevant to malicious parties. These examples also indicate the need for this research to be conducted into the sensitive information in tenders because relevant information can be found from a hackers' perspective.

4.5.1 Domain and IP address space

Information on domain and IP address space give an adversary a clear understanding of the layout of the network. Information gathered is on the different firewall levels, the security zones defined on the firewall, which VLANs are part of which security zone, and what the use is per VLAN. The information also shows the hierarchy in zones: which zone has access to what other zones. Another document contained an overview of all the access switches, over different physical locations of one company. It includes the room number the switch is in, the management IP address and also the

type of switch. Another tender contained a document with an overview of all core- and access switches, its identification numbers, physical location, and type of switch.

4.5.2 Target logon/email address format

Information on logon/email address format can be very useful for adversaries in multiple ways. For starters it is very useful to know the email address format to perform a targeted phishing attack. Furthermore, tenders also contained information on the format of login credentials. Giving away the fact that the email addresses of employees are used as username.

4.5.3 Network topology

For adversaries it is very useful to understand the topology of a network, application, or other information system. It tells the adversary what the best path is to reach its goal. Information gathered contained multiple topologies of networks, but also that of new applications to be developed. The topologies also showed the data flows between components and what protocols are allowed. Other (third party) applications that are used to communicate with are also included in the overviews. Another tender included an overview of an application, all text fields ID's, allowed values, and comments on maximum and minimum length for example.

4.5.4 Physical locations

Information on physical locations was also found. Maps of the physical locations were included in a number of tenders. It was also often indicated per room for which it was intended, so also computer rooms where possible servers are located. Access points were also visible on floor plans that might be of interest to malicious parties for installing rogue devices.

4.6 Summary of 'Gathering sensitive information from tenders'

The goal of this chapter was to determine if sensitive information can be gathered from tenders. This is done through the document analysis. The approach of the document analysis was explained. Existing tenders from https://www.tenderned.nl were downloaded. The tender documents were then analysed based on specific information types. These information types come from the MITRE techniques. The results of the document analysis have shown that tenders contain sensitive information. Malicious people can therefore use tenders to gather information during the reconnaissance phase, in preparation for a cyber attack.

We can conclude that:

- Sensitive information related to the information types can be gathered from public tenders.
- Adversaries can use public tenders as a source for reconnaissance activities.
- Information in tenders might be extra relevant for adversaries because it describes both the current and desired situation.
- Information in tenders also reveal information on other organizations due to dependencies in networks.

The document analysis has its limitations. The analysis is highly dependent on reading the tender documents. When more sensitive information is in a document, it is possible that this information has been read over. The documents also contain images. When images contain words, they will not appear in the search results when specific words are searched for in the documents. In practice a hacker will automate the manual tasks and/or use tools to analyse the results. While automating this would probably yield more results, the goal of gathering sensitive information was still achieved.

5 Risks for Tendering Organizations

5.1 Opening

The previous chapter showed that sensitive information can be found from tenders. The goal of this chapter is to describe the risks that occur for tendering organizations. This is done by interviews with security professionals and by describing and visualizing risk scenario's using the CORAS model.

The following two paragraphs mentions the research question that is answered in this chapter and how this is approached. The paragraphs that follow describe how the interviews are structured, the interview results, and visualizations of the risks. The final paragraph summarizes this chapter.

5.1.1 Research question

In this chapter the third research question is answered:

3. What are the risks that tendering organizations face if adversaries collect this information from tenders during the preparation of cyber-attacks?

This research question aims to determine the risks that tendering organizations face if sensitive information is collected from their tenders. Adversaries can misuse this information in their advance.

5.1.2 Approach

Interviews were used to answer this research question. The purpose of the interviews is to identify the risks that tendering organizations face due to the sensitive information that can be collected from tenders by adversaries. The interviews are structured on the basis of the document analysis in the previous chapter. The persons interviewed are security professionals and professionals who have a function related to security. In addition, the CORAS method is used to visualize the risks that are identified by the interviewees. Literature has been consulted about this method.

In terms of design science research, the interview results stem from the 'Environment'. The visualization of the risks is based on a methodology that emerges from the 'Knowledge base'. Answering this research question contributes to the DSRM phase 'Identify Problem & Motivate'. After this chapter, the first three research questions are answered. The three research questions together form the implementation of the DSRM phase 'Identify Problem & Motivate'.

5.2 Interview approach

To confirm the results of the document analysis and to identify the risks that arise as a result, security professionals and professionals with a job related to security are interviewed. Seven professionals were approached for the interview, five responded. Their job functions are: a Chief Information Security Officer within the Dutch Government, the Head of Intelligence Engineering Architecture Team within a global security firm, a Senior Audit Manager within the Dutch Government, the Head of the Cyber-forensics & Security team of a mid-sized security firm, and a Cyber Security Engineer within a National Bank. The invitation and closure e-mails are included in Annex B. The people are part of the researcher's network.

The interviews are structured and are therefore conducted by email. The questions are structured as such that no follow-up questions are needed. The interview contains 7 questions. The questionnaire used is included in Annex C. The questions are structured in such a way that the interviewee independently reads the tender documents, whereby he or she searches for

information that he deems sensitive, without giving directions such as the information types. In this way the interviewee performs the document analysis as in the previous chapter but without a reference to the information types. The interviewees purely use their own knowledge and experience to recognize sensitive information. The following points are taken into account when selecting interviewees:

- The interviewees are not direct colleagues of each other; direct colleagues tend to have the same vision.
- The interviewees are from both private and public organizations; people who work at private organizations have other interests than people who work at public organizations.
- The interviewees hold various positions relative to each other (strategic, tactical and operational); depending on the level at which someone works, you deal of other interests and risks.
- The interviewees have different years of experience with security and/ or tenders; younger people with less experience might have a fresh view.

These aspects have been included in the selection of interviewees to ensure that the knowledge distribution is as large as possible.

5.3 Interview results

Five of the seven interviewees responded to the interview questions. The interviewees received the same tender documents as used during the document analysis in the previous chapter. The interviewees have indicated in various ways what information they have identified as relevant to hackers. This by copying and pasting text, taking screenshots of text and images and referring to page numbers. It was also very interesting to see that interviewees themselves already described situations in which the information found was placed in a larger context; the information found was a puzzle piece of a planned attack. For example version numbers of a firewall or e-mail addresses of specific people. The information received from the interviews was then analysed and ordered by plotting the information found by the interviewees on table 3 from the previous chapter. See table 4 below. This is done by adding a *blue V* in the relevant column. Example: column 4a is interviewee 'a' who has analysed the 4th tender from the document analysis. The interviewees did not so much discover new types of information that adversaries search for during the interviews. However, information was discovered related to an already established information type that was not identified during the document analysis.

The results of the interviews show that all the respondents can identify sensitive information from the tender documents. Note that the lists of information types were not given to the interviewees. They responded solely based on their knowledge and experience.

Table 4: Sensitive information gathered by the interviewees indicated by a blue V. 5 interviewees analysed a tender. Every interviewee has its own column indicated with an 'a' or 'b' after the specific tender they analysed from the document analysis.

		Tender												
	Information type searched for in the tender	1	2	3	4	4a	4b	5	5a	6	6a	7	8	8a
1	3rd party infrastructure services	V			V	V				V		V		V
2	domain and IP address space				V		V						V	
3	external network trust dependencies				V	V		V		V		V		
4	firmware version				V		V		V	V				
5	target logon/email address format	V	V	V				V	V	V	V	V		
6	externally facing software applications technologies, languages, and dependencies				V	V		V		V			V	
7	security defensive capabilities		V		V		V			V	V		V	V
8	web defensive services									V				
9	network topology	V	V		V	V	V	V	V	V	V		V	V
10	groups/roles									V		V		
11	centralization of IT management				V									
12	physical locations	V	V		V	V	V						V	
13	business processes/tempo	V	V					V	V	V	V	V		V

The interviewees were also asked which risks they could identify due to the sensitive information they gathered from the tenders. The following answers were given (sentences have been adjusted for readability):

1. *"By listing the applications used for office automation, an adversary can search for known vulnerabilities in these applications. The organization faces the risk that if the vulnerabilities are not patched, they are vulnerable to an attack. The concerning organization notes that Skype for Business is used. Skype has had several security vulnerabilities"* (Interviewee 1).

2. *"With regard to the management of the IT environment, the starting point is that, in the long term, external parties perform full LAN / WLAN management, on the understanding that the concerning organization wants to handle a number of standard activities with a view to flexibility of IT services to the organisation. So, there is another party that could take over the entire network. If I am interested in the concerning organization that can also be done through that other party. I can find out for example by signing on the tender."* (Interviewee 1)

3. *"If I want to physically explore the location as an attacker, a map is very useful. I can now figure out where there are probably network cables that I could possibly tap and implement my rogue devices there too. I can also register for the viewing days and then have free access (possibly under supervision) to the building."* (Interviewee 1)

4. *"In this specific case, the Public information on the specific Tender, that relates to the purchasing of the WAN solution and network maintenance definitely constitute to an attractive target for the attacker. The attacker gets immediate insight into the draft agreement understanding details such as when the maintenance of the system would occur. This is a nice information for the attacker since he can plan his attack during the same maintenance window when the network is most vulnerable. The following information in my opinion constitutes for attractive information for the attacker:*

    - *The overall Network WAN diagram gives a great insight into how the entire Network is setup. This is essentially a great starting point for the attacker since he can easily spot the week points and focus his attack there rather than iteratively trying to find a weak spot.*

- *The information on Standard Service Window is also very useful for the attacker since he can plan his attack during the network maintenance, for example by somehow introducing a networking hardware into the system or even compromising the supply chain.*
- *Where there is information about Optical Distribution Frame and CPE, is quite beneficial to the attacker, since he can utilize this information for performing a more passive recon for example by Eaves dropping in optical Network. This is possible in combination with understanding where the fibre network exactly lies. Eaves dropping via Channel Access by directly accessing the optical channel via fibre tapping is possible.*
- *Exposing the network WAN diagram opens up for eaves dropping via channel access, eavesdropping via optical network, compromising the supply chain to gain access."* (Interviewee 2).

5. *Due to the information given on the overall network WAN a hacker has great insight into how the entire network is setup. The diagram visualizes this. This is a great starting point for the attacker since he can easily spot the weak points and focus his attack there rather than iteratively trying to find a weak spot."* (Interviewee 3).
6. *"The combination of insights into the systems and the ways how they tie together is for an adversary relevant. In particular, the information from the systems and versions can be used by hackers to exploit known vulnerabilities to enter the system. Because the architecture is also shared, the hacker can, once inside, determine the route to the specific goal. Furthermore, the documents contain interesting information as contact information, but also information on the applications and maintenance windows."* (Interviewee 3)
7. *"The IP-ranges are exposed in two Excel documents. This is relevant for an attacker when he/she entered the network. This can be seen as a map of the network, this ensures that the attacker knows where he is located within the network and what route is easiest to get somewhere. Also, the names of switches and the type numbers are published in a list. Some switches contain security vulnerabilities. The list helps an adversary search for a switch that is vulnerable. Although the switches are not placed externally, it will aid in lateral movement activities."* (Interviewee 4)
8. *"The documentation shows that the security of the organization is very dependent on 1 firewall. It is also stated that encryption is not being implemented everywhere. Further, it is noted that only OSI layers 1 are up to 4 are inspected. This means that attacks on the application layer will not be detected. The document includes the type of firewall (an older Fortigate). Metasploit contains a module that abuses Fortigate vulnerabilities. Social engineering is also made easier because it is noted that multi factor authentication is not being used."* (Interviewee 4)
9. *"Below the screenshots (not included) and descriptions of the paths I would take are listed. The information on how messages are built up that applications use are good to have in a later stadium.*
   - *Mail information and contact information of an employee can be found, together with a mail address to send digital invoices with an explanation how they should be submitted.*

- *Contact information of two employees listed in an Excel file.*
- *Information on the implementation of the share of the concerning organization. Via Social Engineering this might be exploited to gain more information.*
- *Information on the printers that the concerning organization uses. Possibility this can be exploited.*

I would start with executing social engineering attacks specifically on one of the employees who's contact information was found. According to the information it is clear that this person is involved in the tender." (Interviewee 5)

5.4 Visualization of the identified risks

In this paragraph risk numbers 3, 8, and 9 from the previous paragraph are visualized using the symbols of the CORAS modelling language. These risks have been selected because they have a different attack path. CORAS is a method for conducting security analyses and it provides a language to model risks (Sourceforge, 2015). Visualizing the possible risk scenario's is expected to enhance the understanding of the risks and their underlying cause. Namely that sensitive information is misused. The symbols in the figure below are used.

Figure 11: CORAS building blocks (not all of the symbols are used).

The visualisations in this chapter show:

- Which information that is gathered from tenders is misused in the specific risk scenario. These information is gathered from the tenders by the interviewee.
- How an adversary can misuse the specific type of information. It is possible that the adversary also consults other information that can be gathered from the Internet.
- How a specific asset is damaged by the attack of the adversary and how the sensitive information found in the specific tender relates to this damage.

In addition to the visualization of the risks, the qualitative aspects of likelihood and impact are briefly discussed.

Aspects of risk management such as risk appetite, risk mitigation, risk acceptance, and risk avoidance, have been left out of this chapter because these aspects vary between organizations. One organization has a higher risk appetite than others and therefor may accept more risks.

5.4.1 Visualization of the 3rd risk scenario

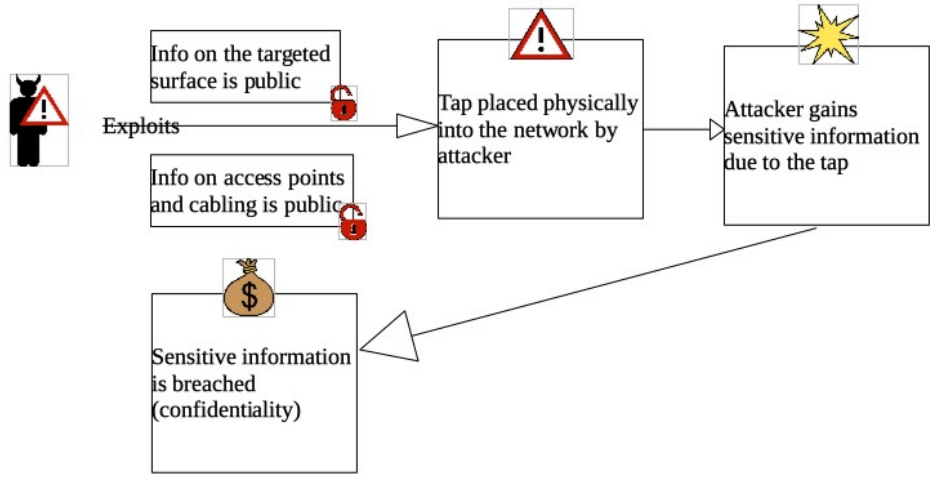

Figure 12: CORAS model risk scenario 3

The model above shows that an adversary performs an attack on an organization. The adversary has information on the physical surface due to maps found of the location. These maps also show the access points in this location and the wiring. This information is found from public tender documents. The adversary visits the location and places a rogue device in the network. The adversary can tap network traffic and therefore the risk exists that sensitive information is breached. Also, the access into the network is also another step for further attacks.

The fact that the maps and the information about access points are accessible to everyone increases the likelihood of this incident. The impact of the incident depends on what is being tapped and what subsequent steps are being taken by the attacker. The incident can therefore have major consequences for the confidentiality of information (tapping of data), integrity of information (when tapping is followed by damage to information) and also availability of information (when information is deleted).

5.4.2 Visualization of the 8th risk scenario

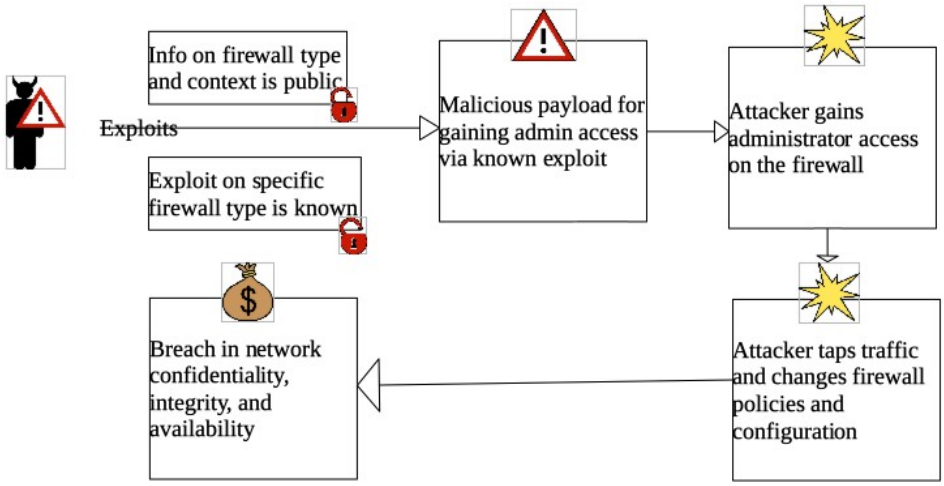

Figure 13: CORAS model risk scenario 8

The model in figure 13 shows that the adversary has information on the firewall type and context on the organization. This information is found in tender documents. The adversary can find a known exploit on the internet for this specific firewall type. The adversary uses this malicious

payload to gain admin access. After gaining admin access, the adversary has major opportunities for further attacks.

Here too, the likelihood of such an incident is greatly increased because the information in the tenders about the firewall types is easily accessible for attackers. As with the previous incident, the impact can also be large in terms of confidentiality, integrity and availability. Depending on what the follow-up actions are of the attacker and the type of organization being attacked.

5.4.3 Visualization of the 9th risk scenario

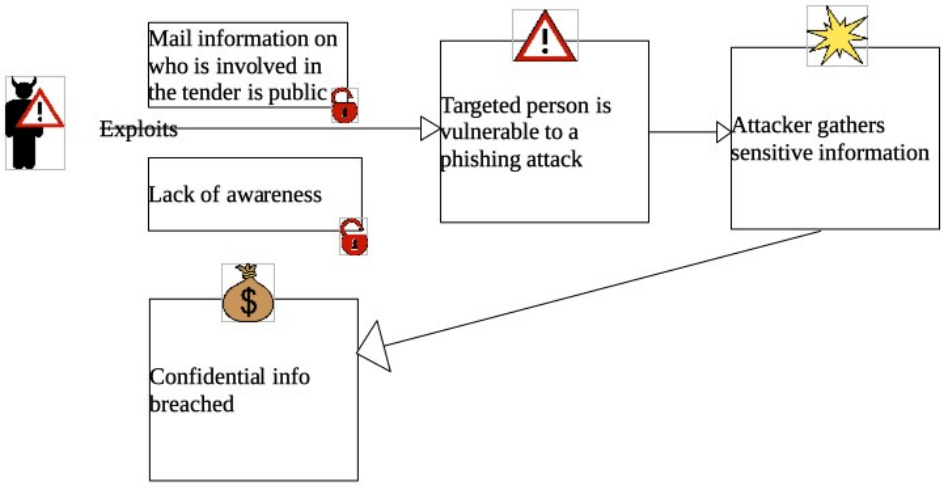

Figure 14: CORAS model risk scenario 9

The model in figure 14 shows that an adversary is able to perform phishing attacks. The adversary gains information on the mail format due to this information being present in public tenders. Furthermore, a lack of awareness is a big vulnerability that exists in multiple organizations. A lack of awareness combined with fishing attacks is a possibility for a breach. The attacker can phish targets on a big scale to gather sensitive information or credentials.

The likelihood of the described incident is increased because information about mail formats and mail addresses are themselves in tenders. This information is therefore easy to access. This makes it easier for attackers to perform targeted phishing. The impact of this incident focuses on the confidentiality of information, but can extend to integrity and availability when a target gives its username and password through phishing.

5.5 Summary of 'Risks for tendering organizations'

The purpose of this chapter was to identify the risks that tendering organizations face by publishing tenders that contain information that is relevant to hackers. A consequence of this publication is that malicious parties collect this information for their own benefit. This results in potential risks for the tendering organizations.

The identification of the risks for these organizations is done by interviewing security professionals. Through the interviews, the security professionals gathered information from existing tenders that they consider relevant to malicious parties. This information is related to the information types.

The interviewees then identified risks for the tendering organizations based on their knowledge and experience. From this it can be concluded that the identified risks have potential impact on the confidentiality, integrity and availability of organizational assets. The risks are visualized on the basis of the CORAS model.

We can conclude that:

- Tenders contain sensitive information that is relevant for adversaries as a result of which tendering organizations face risks in terms of confidentiality, integrity, and availability.
- Adversaries can misuse this data in order to perform a cyber-attack on the tendering organizations. Because this information is easy to access for attackers, the likelihood of cyber-attacks taking place is increased with all the associated risks.
- The attacks have possible impact on the confidentiality, integrity and availability of organizational assets. The impact of the risks that may occur strongly depend on the type of organization.

The interviews and the results have limitations. In particular that the results are purely theoretical. The information collected is not actually misused to demonstrate the risks. The interviewees identified the potential risks purely on the basis of their knowledge and experience. In addition, the visualisations of the risks have been kept simple. This is because they aim to show potential consequences, and not to show the entire attack path.

5.5.1 Research methodology: identify problem & motivate phase

In this section a short reflection is done on the Design Science Research Methodology phase by Peffers et al. (2008).

After answering research questions one, two and three the DSRM phase 'Identify Problem & Motivate' is completed. As described in chapter two, 'Research Design Methodology', the aim of this phase was to identify and substantiate the problem and the justification of a solution. This goal has been successfully achieved. The problem that tenders contain sensitive information that can be misused by malicious parties is clearly described. The motivation to solve the problem is thereby also given.

In the next research phase, 'Define Objectives of a Solution', the input from the answers on research questions one, two, and three is used. The purpose of the next phase is to reason the 'objectives of a solution' from the problem identification. The objectives that are determined are of a qualitative nature: "*a description of how a new artefact is expected to support solutions to the identified problem*" (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2008). This creates the guidelines that help to develop a tool that reduces sensitive information in tenders, as indicated in the introduction of the study.

6 Guidelines to Reduce Sensitive Information in Tenders

6.1 Opening

This chapter determines the guidelines that provide input for the development of a tool that addresses the identified problem; that tendering organizations face risks due to the fact that public tenders contain sensitive information that can be misused by malicious parties. The guidelines are descriptions for a solution to the problem identified in the previous chapters.

The following two paragraphs mention the research question that is answered in this chapter and how this answering is approached. The guidelines are extracted from multiple sources. After describing the different sources, a paragraph is dedicated per source in which the guidelines are determined that are derived from that specific source. The final paragraph summarizes this chapter.

6.1.1 Research question

In this chapter the fourth research question is answered:

4. Which guidelines can be determined that aim to prevent the risks that occur if adversaries collect sensitive information from tenders during the preparation of cyber-attacks?

This research question aims to determine guidelines that describe how to prevent sensitive information from being included in tenders. If the information is not included in the tenders, the organizations do not face the identified risks. The guidelines deal with the identification and reduction of sensitive information, but also with the practical aspects that a final tool must meet, such as the context in which the tool should be used.

6.1.2 Approach

To answer the fourth research question and therefore to establish the guidelines, use was made of the information from the previous research questions, information on the tendering process, interviews with senior purchasers, and other existing tools that may yield relevant practical guidelines. The information from previous research questions focuses in particular on how the problem arises that sensitive information is included in tenders, how malicious parties can collect and misuse this information, and what risks arise for the tendering organizations. By analysing this information from the perspective of solution-oriented thinking, solutions can also be filtered out of the answers as guidelines. A concrete example is the information types that malicious parties search for, while the information types can also be used as a guide for what information should not be included in tenders. The tendering process provides guidelines that an ultimate tool must conform to. The tool must be used in an existing situation: organizations, people, and processes. Another existing tool that is being analysed provides guidelines that focus on the practical side of a solution. The interviews with senior purchasers are used to evaluate the already defined guidelines and also to establish new guidelines that concern, for example, who uses a final solution. In addition, these interviews also provide information that is separate from the guidelines, but that is relevant to the research.

In terms of design science research, the interview results and the guidelines that are filtered from the previous chapters stem from the 'Environment'. The 'Knowledge base' offers already existing artefacts (Meta-Artefacts) that can serve as examples and can be analysed from which guidelines can also be filtered. Answering the fourth research question contributes to the DSRM phase 'Define Objectives of a Solution'.

6.2 Sources from which the guidelines arise

As is described in the previous paragraph 'Approach', there are several aspects – sources – from which guidelines can be derived. Four sources can be distinguished. They are explained below:

- The first source is ‘the identified problem’, this is the umbrella term for the answers to the previous research questions. The identified problem shows a number of specific information types that are relevant to malicious parties. These information types can therefore also be used to solve the problem. For example, by identifying sensitive information in tenders and then reducing it. The guidelines that are derived from the source ‘problem identification’ primarily focus on the reduction of sensitive information.
- The second source is the context in which the problem occurs. After all, tendering is bound by laws and regulations, but also by processes that are already embedded in organizations. The tenders must be made public, but process steps may be added or adjusted. If a final tool is designed it has to be embedded in the existing context; the processes, the organizations, and also the people who have to use the tool. The guidelines that are derived from the source ‘tender context’ primarily focus on the implementation within existing organizations, processes, and people.
- The third source for establishing guidelines are similar artefacts. Other tools that already exist in practice that have similar functions are analysed. By means of this analysis, guidelines can be established that, for example, improve the practical side of the final tool. It is also possible that guidelines may be established that ensure better implementation in the process. The guidelines that are derived from the source ‘similar artefacts’ primarily focus on the practical side of a final tool.
- The fourth and final source is the interviews with senior purchasers. The senior purchasers are familiar with the organizations that are tendering, with the regulations, and also with the tender process. The purpose of the interviews is to evaluate the already established guidelines with the senior buyers. The interviews can also result in new artefacts. The guidelines that are derived from the source 'interviews' primarily focus on implementation within existing organizations, processes, and people, because the senior purchasers are familiar with this. But the guidelines may also be focused on the practical side or equally on the reduction of sensitive information. This depends on the knowledge of the interviewees on the identified problem.

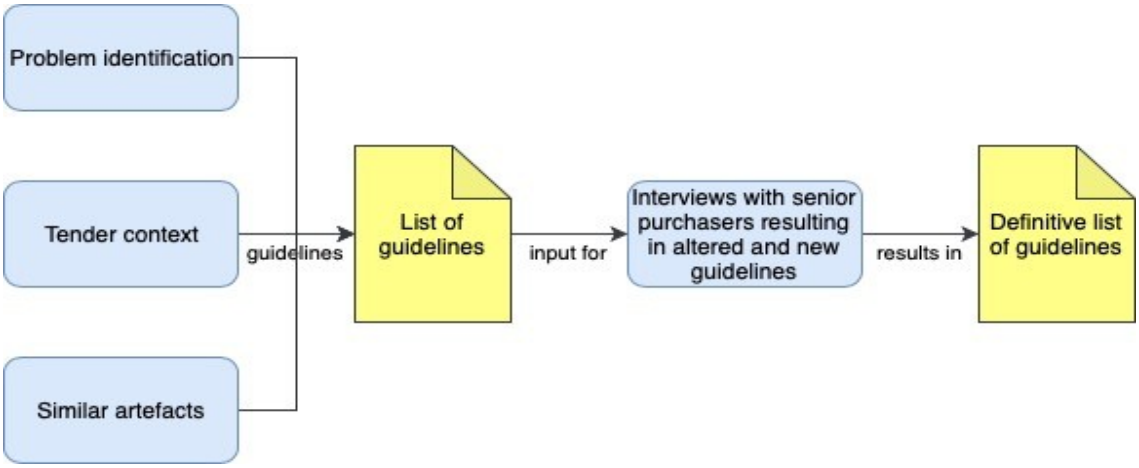

Figure 15: Visualization of how the guidelines are determined. The first three sources result in an initial list of guidelines. Thereafter, the interviews are used to reflect on these guidelines and also to establish additional guidelines.

In the paragraphs that follow the guidelines are determined. A paragraph is dedicated per source in which the guidelines are determined that are derived from that specific source. The exception here are the paragraphs on the interviews conducted with senior purchasers because in the interviews the already determined guidelines are reflected on and other information has been gained that do not relate to guidelines. The paragraph on the interviews describe the interview approach, the results that do not relate to the guidelines, and the results that relate to the guidelines.

6.3 Guidelines from the problem identification

The table below contains the guidelines that are derived from the problem identification. The primary focus of the guideline is indicated. The guidelines mainly stem from the MITRE techniques, translated into the information types. These information types can be used to search for in tenders and remove the identified information.

Table 5: Initial guidelines from the problem identification

ID	Guideline	Focus
1	Tenders must be analysed to identify and remove information related to 3rd party infrastructures services. Infrastructure services includes the hardware, software, and network resources required to operate a communications environment. This infrastructure can be managed by a 3rd party rather than being managed by the owning organization.”	Reduction of sensitive information
2	Tenders must be analysed to identify and remove information related to domain and IP address space. Domain Names are the human readable names used to represent one or more IP addresses. IP addresses are the unique identifier of computing devices on a network. Both pieces of information are valuable to an adversary who is looking to understand the structure of a network.	Reduction of sensitive information
3	Tenders must be analysed to identify and remove information related to external network trust dependencies. Network trusts enable communications between different networks with specific accesses and permissions. Network trusts could include the implementation of domain trusts or the use of virtual private networks (VPNs).	Reduction of sensitive information
4	Tenders must be analysed to identify and remove information related to firmware version. Firmware is permanent software programmed	Reduction of sensitive

	into the read-only memory of a device. As with other types of software, firmware may be updated over time and have multiple versions. Apart from firmware, versions of other components are also relevant.	information
5	Tenders must be analysed to identify and remove information related to target logon/email address format. Email addresses, logon credentials, and other forms of online identification typically share a common format. This makes guessing other credentials within the same domain easier. For example, if a known email address is first.last@company.com it is likely that others in the company will have an email in the same format.	Reduction of sensitive information
6	Tenders must be analysed to identify and remove information related to externally facing software applications technologies, languages, and dependencies. Software applications will be built using different technologies, languages, and dependencies. This information may reveal vulnerabilities or opportunities to an adversary.	Reduction of sensitive information
7	Tenders must be analysed to identify and remove information related to security defensive capabilities. Security defensive capabilities are designed to stop or limit unauthorized network traffic or other types of accesses.	Reduction of sensitive information
8	Tenders must be analysed to identify and remove information related to web defensive capabilities. An adversary can attempt to identify web defensive services as CloudFlare, IPBan, and Snort. This may be done by passively detecting services, like CloudFlare routing, or actively, such as by purposefully tripping security defences.	Reduction of sensitive information
9	Tenders must be analysed to identify and remove information related to network topology. A network topology is the arrangement of the various elements of a network (e.g., servers, workstations, printers, routers, firewalls, etc.). Mapping a network allows an adversary to understand how the elements are connected or related.	Reduction of sensitive information
10	Tenders must be analysed to identify and remove information related to groups and roles. Personnel internally to a company may belong to a group or maintain a role with electronic specialized access, authorities, or privilege that make them an attractive target for an adversary. One example of this is a system administrator.	Reduction of sensitive information
11	Tenders must be analysed to identify and remove information related to centralization of IT management. Determining if a "corporate" help desk exists, the degree of access and control it has, and whether there are "edge" units that may have different support processes and standards.	Reduction of sensitive information
12	Tenders must be analysed to identify and remove information related to physical locations. Physical locality information may be used by an adversary to shape social engineering attempts (language, culture, events, weather, etc.) or to plan for physical actions such as dumpster diving or attempting to access a facility.	Reduction of sensitive information
13	Tenders must be analysed to identify and remove information related to business processes/tempo. Understanding an organizations business processes and tempo may allow an adversary to more effectively craft social engineering attempts or to better hide technical actions, such as those that generate network traffic.	Reduction of sensitive information
14	Before the analysis starts where information related to the above guidelines is identified and removed, it is important that new types of	Embedding in the contextual

	information can be added. The specific information types that hackers look for may change in the future. It is possible that with a new cyber attack it appears that hackers have used a different information type when preparing for the attack.	Practicality

6.4 Guidelines from the tender context

The table below contains the guidelines that are derived from the tender context. Each guideline is described together with the primary focus of the guideline. The guidelines stem from the information on tenders, the tender process, and the laws and regulations on tenders.

Table 6: Initial guidelines from the tender context

ID	Guideline	Focus
15	After analysing the information in tenders, tenders must still be usable. Information that is identified and removed must not result in a tender that does not comply with the applicable laws and regulations.	Embedding in the context
16	After analysing the information in tenders, tenders must still be usable. Information that is identified and removed must not result in a tender that is no longer usable in the sense that the tender does not provide sufficient information to contractors who want to register for the assignment.	Embedding in the context
17	The analysis where the tenders are examined to identify sensitive information must be embedded in the tendering process prior to the step where the documents are made public. After performing the analysis based on the information types and processing the results, the tenders must be published.	Embedding in the context
18	The analysis that is performed where tenders are examined must be executed by a professional in the tendering process, for example a buyer. Nor should it be expected that the user has explicit security knowledge.	Embedding in the context

6.5 Guidelines from similar artefacts

An example of an a-like artefact as a foreseeable solution that can be made on the basis of the guidelines is a plagiarism scanner. The plagiarism scanner is used by researchers, teachers, and professors, to determine whether fraud has been committed in the sense that text has been improperly copied in a publication. The scanner takes the document that is being published as input. The scanner uses a large database containing all kinds of previous publications and analyses the input, whereby a comparison is made between the input and the text in the publications in the database. The output of the scanner is the document (the input) containing highlighted text which is also found in publications in the database and a percentage of the amount of highlighted text in relation to the total document. The researcher, teacher, or professor must then determine from his or her own professionalism whether plagiarism has been committed or sources have been handled incorrectly.

The table below contains the guidelines that are derived from the similar artefact; the plagiarism scanner. Each guideline is described together with the primary focus of the guideline.

Table 7: Initial guidelines from similar artefacts

ID	Guideline	Focus
19	Prior to the analysis on the tender information, the user, for example a buyer, must provide the documents as input by simply selecting or dragging and dropping them.	Practicality
20	After the analysis, the results are clearly displayed to the user. In this way the user can see in an overview how much sensitive information has been identified and removed.	Practicality
21	A database is kept with identified and deleted results. The objects that the database contains are included in the analysis. In this way, analysis is not only based on the information types, but an analysis can also be carried out whereby a comparison is made between the information in the tenders and the objects in the database – the previously identified and deleted information in other tender documents.	Practicality

6.6 Guidelines from the interviews

As described earlier in this chapter the interviews influence this research differently than just as a source for guidelines. The results of the interviews influence this research in three different ways.

1. First of all, information was gained through the interviews which has nothing to do with the guidelines but which is relevant to mention in relation to the research.
2. Secondly, guidelines have been discussed with the interviewees that are not (entirely) correct or cannot be implemented for example. These guidelines must be adjusted or removed.
3. Finally, new guidelines emerged from the interviews.

The following sections describe the interview approach and the three different ways in which the interview result has impact and the details thereof.

6.6.1 Interview approach

Two senior purchasers are interviewed in person. These purchasers are part of the researchers network. This is done on the basis of the structure and interview questions in annex D. The interviews had a structured character. Interview reports have been made of both interviews. The final reports of both interviews are included in annex E. The reports are shared with the interviewees via e-mail for comments and feedback. The names of the interviewees and of companies in answers by the interviewees have been left out of the interview results. The interviews had multiple goals. First of all to get clarification about the tender context. Secondly, the identified problem is discussed to ascertain to what extent the interviewees are familiar with this. Perhaps security is already been taken into account implicitly in the process. Finally, the purpose of the interview was also to evaluate the already defined guidelines. New guidelines can also arise from these results. The interviewees work in the same organization, although not in the same team. This means that they do not work on the same tenders and also not on tenders for the same organization. It is a limitation that no purchasers are interviewed from other organizations. The interviewees however did mention that all organizations have to comply with the procurement act. The process is the same. Apart from that it has been taken into account to interview senior buyers, as they have the most knowledge on the tendering process.

The results of the interviews were analysed by keeping them against the existing guidelines. In addition, the research results thus far have been compared with the results of the interviews to

either invalidate or confirm them. The research results, the answers to the research questions, have been confirmed by the interview results.

6.6.2 Relevant information from the interviews

Below are the results from the interview that are not related to the guidelines. However, these results are relevant to mention.

- Both interviewees indicated that a check on sensitive information would be good. Both interviewees indicated that a check on sensitive information is now sort of implicit. Depending on the people involved in the tender process, attention is paid to what information is disclosed. Someone from security is not always involved. This depends on what the tender is about. An interviewee indicated that he occasionally sends documents to his security officer and the security officer of the client to have the content scanned for sensitivity. This is not a structural step in the process. In addition, this step therefore depends on the knowledge, experience and wilfulness of the security officers. A check by the security officer is also not based on a baseline. One security officer defines sensitive information differently than the other security officer. One of the purchasers also indicated that, based on knowledge and experience, they sometimes implicitly indicate that information should not be included, but that this check is not based on a certain baseline. The above steps do indicate that there is a need for a check on sensitive information. The senior purchasers agree. The example of the plagiarism scanner was then explained to the interviewees. The interviewees responded enthusiastically to this.
- An interviewee also indicated that not all tenders are made public. Some tenders are for a product that should be kept secret for example. But the documents that are drawn up are still shared with various parties. According to the interviewee, it would therefore be good to also scan these documents. Even though there is cooperation with a party, it is good not to reveal too much information.

6.6.3 Modification in established guidelines

Below are the results from the interviews that relate to already established guidelines. This is shown in the table below. The first column contains the ID of the guideline in which a change takes place. The second column describes what will be changed in the guideline and why.

Table 8: Modifications to the initial guidelines

ID	Guideline
1-16	The guidelines 1 up to and including 16 prescribe that the identified information is also deleted immediately. This is not possible. Both interviews clearly showed that the purchaser involved in the tender process is not ultimately responsible for the tender documents and the content of the tender documents. The role of the purchaser is to provide advice during the tender process. The client of the purchaser is responsible for the tender documents and their contents. The client therefor has to determine which identified information shall be deleted. The guidelines are modified to only prescribe 'identify'.
14	This guideline prescribes that new information types can be added to the analysis because adversaries may discover new information types they can use to prepare cyber-attacks. The interviewees both indicated that they themselves do not have the security knowledge to determine what types of information should be sought in the tenders. For this reason, when there are new information types, it is an option to automatically add new information types from the MITRE website to the final tool. This is added to the guideline.
17	This guideline prescribes that the tool must be embedded in the tendering process prior to the step where the documents are made public. The interviewees indicated that a step

	whereby the tender document are scanned can be included in the tender process. This must be done before the tender documents are made public. The documents must also be provided with information to the extent that they contain all information and that new information is not added after scanning. This would ensure that new information is not scanned. One of the interviewees indicated to have the '0.9' versions scanned, so actually just before they are uploaded. The guideline is modified to prescribe 0.9 documents.
18	This guideline prescribes that a professional related to the tender process must use the tool. The interviewees indicate that they are the logical choice of who uses the tool. Because thy are the persons that upload all the documents. This has been modified in the guideline.
19	This guideline prescribes that a professional related to the tender process must use the tool. The interviewees indicate that they are the logical choice of who uses the tool. This has been modified in the guideline.
20	This guideline prescribes that the user of the tool sees an overview indicating which information has been identified and deleted. This is not possible. Both interviews clearly showed that the purchaser involved in the tender process is not ultimately responsible for the tender documents and the content of the tender documents. The role of the purchaser is to provide advice during the tender process. The client of the purchaser is responsible for the tender documents and their contents. The client therefor has to determine which identified information shall be deleted. The guideline is modified to only prescribes an overview of identified information.
21	This guideline prescribes that the database is kept with identified and deleted results. It is possible however that the analysis resulted in more identified results than that the client actually deleted. It is the decision of the client to delete the information. The guideline is modified to keep track of all identified results.

6.6.4 New guidelines from the interviews

The table below contains the new guidelines that are derived from the interviews. Each guideline is described together with the source. The primary focus of the guideline is also indicated.

Table 9: Guidelines as a result of the interviews

ID	Guideline	Focus
22	The purchasers have clearly indicated that the analysis must take place automatically. It is not desirable to use a manual checklist that purchasers must use.	Embedding in the context & Practicality
23	When sensitive information related to the information type that the analysis is based on is found it should be made clear why the information is identified as sensitive. The purchaser may not know why this information might be relevant for adversaries.	Embedding in the context
24	The purchasers involved can perform the automated analysis to identify information. As a result of guideline 20 – the overview with the result of the analysis must be saved in a visual format.	Embedding in the context
25	The purchasers have clearly indicated that they want to perform the automated analysis. As a result of guideline 23, the results must be sent to the client in the tendering process. The client must then decide which information must be removed.	Embedding in the context
26	Tenders should be analysed to identify information related to privacy sensitive information. Information about persons is sensitive to privacy and, therefore, may not be included in tender documents that are made public according to legislation and regulations (General Data	Reduction of sensitive information

	Protection Regulation). In addition, malicious parties can use such information for social engineering activities.	

As a result of the above paragraphs the complete and final list of guidelines is included in Annex F. Note that the above paragraphs contain guidelines that are removed or altered as a result of the interviews.

6.7 Summary of 'Guidelines to reduce sensitive information in tenders'

The purpose of this chapter was to define qualitative objectives for a solution to the problem of this research. The objectives that have been determined are the guidelines. The guidelines are qualitative descriptions that can be used as input when creating a solution to the problem that tenders contain information that is relevant for adversaries and can be collected by them during the reconnaissance phase. The guidelines are therefore part of the recommendations for further research into the creation of the solution.

This chapter described the four different sources from which the guidelines have been defined. The sources are: the identified problem, the tender context, similar artefacts, and the interviews with senior purchasers. In addition to the guidelines that are derived from the interviews, the guidelines from other sources were also evaluated during the interviews with senior purchasers to check whether they stood firm in the practice of the procurement world. It can be concluded that senior purchasers see added value in a solution based on the established guidelines.

We can conclude that:

- The guidelines have been established from various sources (the identified problem, the tender context, a-like artefacts, and the interviews) which have been extensively researched.
- Guidelines have been defined with the aim of preventing risks for tendering organizations by reducing sensitive information in tenders, embedding a solution in the context of the procurement world, and creating a solution that can be used in practice by people involved in the tender process.
- The guidelines were evaluated through the interviews with senior purchasers. In addition, they have also been tested to see whether they stand in the practice of tenders and whether they offer added value. The guidelines stand and are of added value to reduce sensitive information from tenders according to the senior purchasers.
- The risks that tendering organizations face are mainly prevented due to the fact that the likelihood of the risks decreases because the information is excluded from tenders before being made public.
- The guidelines are defined in order to support follow-up research in which a solution can be build on the basis of the guidelines.

The main goal of this research is to establish guidelines aimed at identifying and reducing sensitive information in tenders, in order to prevent that malicious parties gather and use this information in the preparation of cyber attacks against tendering organizations. The guidelines that focus on the practical side and embedding in the process are secondary to the goal. Therefor, the following chapter discusses the guidelines with the primary focus: 'Reduction of sensitive information'. Further research is done on which techniques can be used to implement the guidelines.

7 Implementation of Guidelines through Automation

7.1 Opening

This chapter further examines the guidelines that primarily focus on reducing sensitive information in tenders. The purpose of this chapter is to describe more specifically for these guidelines which techniques can be applied in practice. For example, the use of image recognition or text mining. This allows for more concrete recommendations for further research.

The following two paragraphs mention the research question that is answered in this chapter and how this answering is approached. The paragraphs that follow describe multiple techniques that can be researched further to implement the guidelines. The last paragraph summarizes the important aspects of this chapter.

7.1.1 Research question

In this chapter the fifth research question is answered:

5. How can the established guidelines be implemented to create a solution that can be used to identify and reduce sensitive information in tenders in order to prevent that adversaries gather sensitive information during the preparation of cyber-attacks?

This research question aims to describe concrete techniques to the guidelines that focus on reducing sensitive information. This is done by doing literature research.

7.1.2 Approach

The guidelines in the previous chapter have different goals that they focus on: reducing sensitive information, embedding in the context, and practicality. Since the aim of this research is to identify and reduce information from tenders that is relevant for adversaries in preparing cyber-attacks, the guidelines with the scope of 'reduction of sensitive information' are further researched. The purpose of this is to describe more concrete techniques as to how the guidelines can be implemented. This is purely done through literature research. For the collection of literature, a start was made to search in Google Scholar for the specific terms: "text mining" and "image recognition". After reading literature regarding these terms, other search terms were searched for more specifically, namely: "regular expressions", "text mining" "images", "intrusion detection technique", "data comparison", "optical character recognition ", and "image recognition". The literature on how these techniques work is analysed in particular and translated into how to apply them in the context of this research. Literature on examples of the techniques is used in particular to clarify how the techniques work. The guidelines are then subdivided into the techniques that can be applied, depending on the type of identification that must take place. Consider, for example, the differences between text recognition and the recognition of patterns in images. The different ways of identification are discussed in the following paragraph. Literature research is then carried out per technique to describe how the identification of sensitive information can take place. A section is dedicated to this per technique.

In terms of design science research, both the 'Environment' and the 'Knowledge base' are consulted in this chapter. The 'Environment' supports this research question with the determined guidelines. The 'Knowledge base' is consulted for literature on existing techniques that can possibly be implemented to identify sensitive information from tenders. The outcome of the chapter, the answer to the research question, adds new information to the 'Knowledge base' and enhances the guidelines in the 'Environment'. Answering the fifth research question contributing to the DSRM

phase 'Define Objectives of a Solution', as it gives more detailed suggestions for the design and development phase.

7.2 Dividing the guidelines per technique

The guidelines that focus on reducing information can be distinguished in five different techniques. The techniques are divided in such a way that each technique is characterized by the type of identification that must take place[2]. The techniques are:

- Regular Expressions: a regular expression is a sequence of characters that define a search pattern. The tender to be analysed must be analysed on the basis of fixed expressions. Examples are IP numbers, names, e-mail addresses, etc. These objects have a fixed format so that regular expressions can be used. This technique can also be used to search for specific words.
- Text mining: text mining is the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning. The purpose of using text mining is to make connections between different information in the tender that is analysed. It is possible that information is not relevant in itself, but in combination with information elsewhere in the tender it is.
- Comparison with known information: by this technique information in tenders is compared with sources on, for example, the internet about this information. For example, when a tender includes information about version numbers of information systems to compare it with information about these version numbers on the Internet. It is possible that vulnerabilities are known for the specific version numbers. The aim is to gather additional information in this way about information that is included in tenders. Just as with text mining, it is possible that information itself is not relevant to malicious parties, but together with other information it is.
- Optical character recognition: optical character recognition is the conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo, or another image. The purpose of using this technique is to convert images in tenders into text so that other techniques that focus on text analysis can then be applied. Tenders can contain images such as floor plans, IP plans, architectural diagrams, et cetera. These images regularly contain text. Without converting this text in images into machine-encoded text, it cannot be included in further analysis.
- Image recognition: image recognition (or computer vision) is an interdisciplinary scientific field that deals with how computers can be made to gain high-level understanding from digital images or videos. The purpose of using this technique is, in particular, to be able to analyse images containing symbols, process diagrams, and other forms of symbolism other than text. Tenders can contain images such as process diagrams. These are not always provided with text. In order to be able to analyse these, technology must be applied using artificial intelligence theory.

The last two techniques aim to convert images to text, so that the first three techniques can be applied to analyse this text.

2 Perhaps there are more techniques related to artificial intelligence and machine learning. The purpose of this research, however, is not to examine all the possible techniques in detail.

The table below contains the guidelines in scope of this chapter. An estimate has been made for each guideline which techniques can be applied.

Table 10: Guidelines correlated to the techniques. Multiple techniques can be related to a single guideline.

ID	Guideline	Technique
1	Tenders must be analysed to identify information related to 3rd party infrastructures services. Infrastructure services includes the hardware, software, and network resources required to operate a communications environment. This infrastructure can be managed by a 3rd party rather than being managed by the owning organization.”	Regular expressions Image recognition Optical character recognition
2	Tenders must be analysed to identify information related to domain and IP address space. Domain Names are the human readable names used to represent one or more IP addresses. IP addresses are the unique identifier of computing devices on a network. Both pieces of information are valuable to an adversary who is looking to understand the structure of a network.	Regular expressions Optical character recognition
3	Tenders must be analysed to identify information related to external network trust dependencies. Network trusts enable communications between different networks with specific accesses and permissions. Network trusts could include the implementation of domain trusts or the use of virtual private networks (VPNs).	Regular expressions Image recognition Optical character recognition
4	Tenders must be analysed to identify information related to firmware version. Firmware is permanent software programmed into the read-only memory of a device. As with other types of software, firmware may be updated over time and have multiple versions. Apart from firmware, versions of other components are also relevant.	Regular expressions Comparison with known information
5	Tenders must be analysed to identify information related to target logon/email address format. Email addresses, logon credentials, and other forms of online identification typically share a common format. This makes guessing other credentials within the same domain easier. For example, if a known email address is first.last@company.com it is likely that others in the company will have an email in the same format.	Regular expressions
6	Tenders must be analysed to identify information related to externally facing software applications technologies, languages, and dependencies. Software applications will be built using different technologies, languages, and dependencies. This information may reveal vulnerabilities or opportunities to an adversary.	Regular expressions Optical character recognition Comparison with known information
7	Tenders must be analysed to identify information related to security defensive capabilities. Security defensive capabilities are designed to stop or limit unauthorized network traffic or other types of accesses.	Regular expressions Image recognition Optical character recognition
8	Tenders must be analysed to identify information related to web	Regular expressions

	defensive capabilities. An adversary can attempt to identify web defensive services as CloudFlare, IPBan, and Snort. This may be done by passively detecting services, like CloudFlare routing, or actively, such as by purposefully tripping security defences.	Image recognition Optical character recognition
9	Tenders must be analysed to identify information related to network topology. A network topology is the arrangement of the various elements of a network (e.g., servers, workstations, printers, routers, firewalls, etc.). Mapping a network allows an adversary to understand how the elements are connected or related.	Regular expressions Image recognition Optical character recognition
10	Tenders must be analysed to identify information related to groups and roles. Personnel internally to a company may belong to a group or maintain a role with electronic specialized access, authorities, or privilege that make them an attractive target for an adversary. One example of this is a system administrator.	Regular expressions
11	Tenders must be analysed to identify information related to centralization of IT management. Determining if a "corporate" help desk exists, the degree of access and control it has, and whether there are "edge" units that may have different support processes and standards.	Regular expressions Image recognition Optical character recognition
12	Tenders must be analysed to identify information related to physical locations. Physical locality information may be used by an adversary to shape social engineering attempts (language, culture, events, weather, etc.) or to plan for physical actions such as dumpster diving or attempting to access a facility.	Regular expressions Image recognition Optical character recognition
13	Tenders must be analysed to identify information related to business processes/tempo. Understanding an organizations business processes and tempo may allow an adversary to more effectively craft social engineering attempts or to better hide technical actions, such as those that generate network traffic.	Regular expressions Image recognition Optical character recognition
26	Tenders should be analysed to identify information related to privacy sensitive information. Information about persons is sensitive to privacy and, therefore, may not be included in tender documents that are made public according to legislation and regulations (General Data Protection Regulation). In addition, malicious parties can use such information for social engineering activities.	Regular expressions

Note that the technique ‘Text mining’ is not included because when implementing this technique, all existing information in the tender is linked to each other. The technology therefore does not focus on one specific information type but on several together. Where multiple information types are abused at different steps in the attack process (for example, the multiple phases in the cyber kill chain (Lockheed Martin, 2015)).

7.3 Applying the techniques

The different techniques are described in the following sections. Examples are given on the application of the technique. First the three techniques are described that take raw text as input.

The two techniques that aim to convert images to text are discussed last, as these techniques result in raw text that can be inputted in the previously described techniques.

7.3.1 Regular expressions

Regular expressions are included as a technique in all guidelines. Regular expressions can be used for two purposes.

- Searching for specific words: this applies to all guidelines. Each guideline contains relevant keywords. Examples related to the guidelines are: 'outsourced to', 'IP address', 'VPN', 'version', 'e-mail', 'protocol', specific protocols such as TCP or UPD, 'firewall', 'IDS', 'IPS', 'proxy', 'server', 'topology', 'map', 'administrator', 'active directory', 'username', 'help-desk', 'ITIL', 'process', 'surname', and 'last name'.
- In addition, specific expressions can be searched for. The examples for searching for IP addresses and e-mail addresses is explained below.

  "*Matching an IP address is another good example of a trade-off between regex complexity and exactness.* `\b\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}\b` *will match any IP address just fine. But will also match 999.999.999.999 as if it were a valid IP address*" (Goyvaerts, 2018). This regular expression searches for a string that contains four numbers from 0 to 999, divided by a dot ('.').

  The regular expression is: "`\b[A-Z0-9._%+-]+@[A-Z0-9.-]+\.[A-Z]{2,}\b`" (Goyvaerts, 2019). This regular expression searches for a string that contains characters, followed by a '@', followed by more characters, followed by a dot '.', followed by more characters.

The text that is identified by regular expressions can be highlighted as 'sensitive'. Due to the specific keywords and expressions it can also be indicated to which information type the identified text belongs to.

7.3.2 Text mining

Text mining is an artificial intelligence technology that uses natural language processing to transform the free text in documents and databases into normalized, structured data suitable for analysis or to drive other techniques for further analysis (Linguamatics, 2019).

In a paper by (Sharma, Kumar, & Chand, 2018) text mining was used to analyse and extract useful information from a large number of research articles quickly and automatically. "*Text mining is the method of defining innovative, and unseen knowledge from unstructured, semi-structured and structured textual data" (Sharma, Kumar, & Chand, 2018).* Implementing text mining in the context of this research has the same goal: extracting useful information – sensitive information – from large amounts of data – the tenders.

The figure below shows the different steps when conducting text mining. These steps contain the following (Foster Open Science, 2018):

1. "*Gathering: Collecting data from different resources, such as website, emails, customer comments, document file. Depending on the application, this process can be completely automated or guided by the text miner*." In the context of this research the documents come from the tender process.
2. "*Pre-processing, such as content identification / extraction of representative features:*"

- *"Text clean-up: removing of any unnecessary or unwanted information such as ads from pages."*
- *"Tokenisation: a computer only 'sees' a string of characters, without for example being able to identify paragraphs, sentences or words. Tokenisation splits the text in meaningful entities (words, sentences, etc.) given present white spaces and punctuations."*
- *"Feature extraction (also called attribute selection): it is the process of characterizing the text to obtain a set a quantitative measurements. For example frequency of words in a text, type of words, syntactic information. These features can then be used for further processing."*

3. "*Index: Creating an index of certain terms, their locations and numbers. This allows quick access to and structuring of the processed data.*"
4. "*Mining: At this step, the text has been properly pre-processed and can now be 'mined'. For that, we apply different data exploration techniques to reveal new knowledge. These can for example include identifying mention of specific terms, linking of these terms with a dictionary or thesaurus for disambiguation, and identifying relationships between different terms.*"
5. "*Analysis: The mining steps produce raw results. These need to be evaluated and visualized so that they can be interpreted with respect to the questions the text-miner wants to investigate.*"

Figure 16: The steps in text mining (Foster Open Science, 2018)

According to (Shinde & Gill, 2014) the fundamental methods for conducting text mining are natural language processing (NLP) and information extraction (IE) techniques. "*The former technique focuses on text processing while the latter focuses on extracting information from actual texts*" (Shinde & Gill, 2014). IE can be used to create relations from different sources of unstructured text. This aligns with step 4 of the above described process. Where specific terms can be linked with a dictionary, for example a dictionary aimed at the different steps in the Cyber Kill Chain; which information is needed to execute each step? This way, correlations between multiple information types are made clear.

7.3.3 Comparison with known information

Although this technique can only be used to realise two guidelines, as far as known, it is very relevant. With this technique, known vulnerabilities, and therefore also known attacks, can be prevented. Several websites on the internet contain databases containing information about known vulnerabilities, for example based on applications and version numbers. An example of this is https://www.cvedetails.com.

The comparison can be done by combining multiple techniques. First of all, regular expressions must be used to filter version numbers from the tenders. This regular expression is already known and has been described. The expression is "`\eg?[0-9]+\.[0-9]+\.[0-9]+(?:\.[0-9]+)?\B`" (Jan Goyvaerts, 2019). Briefly said, this expression identifies numbers split by a dot ".". Examples of formats of version numbers that can be identified are 6.0.516, 1.1.20, 10.1.1.100. In addition, regular expressions should be used to search for specific application names in the tender documents. Determining which applications to look for can be done as follows:

1. First, the version numbers are filtered from the tender by means of the aforementioned regular expression.
2. A comparison is then made between the filtered version numbers and databases with known vulnerabilities specific to those version numbers. An attempt is made to filter the application names that belong to the known vulnerabilities. The screenshot below shows that by searching on version numbers in https://www.cvedetails.com gives lots of results.

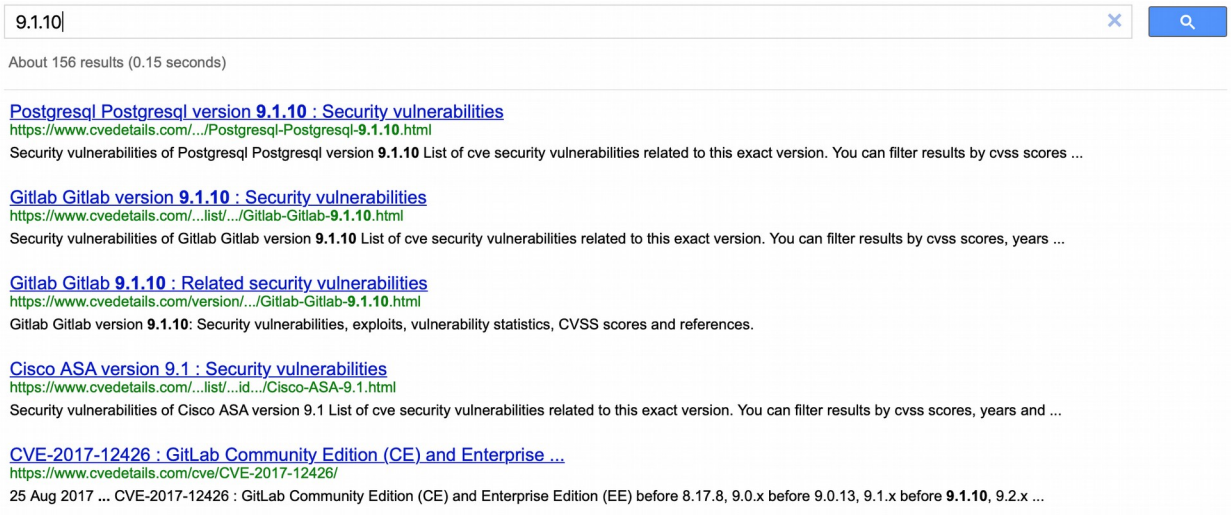

Figure 17: Random version number search results in multiple application names

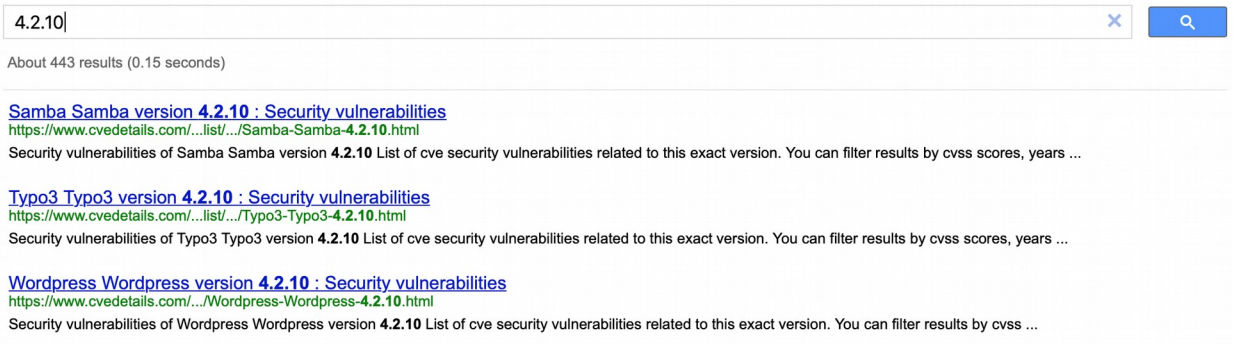

Figure 18: Random version number search results in application names

3. These specific application names are then searched for in the tender documents together with the version numbers. In this way, specific applications and version numbers can be identified that contain known vulnerabilities.

This process is shown visually below.

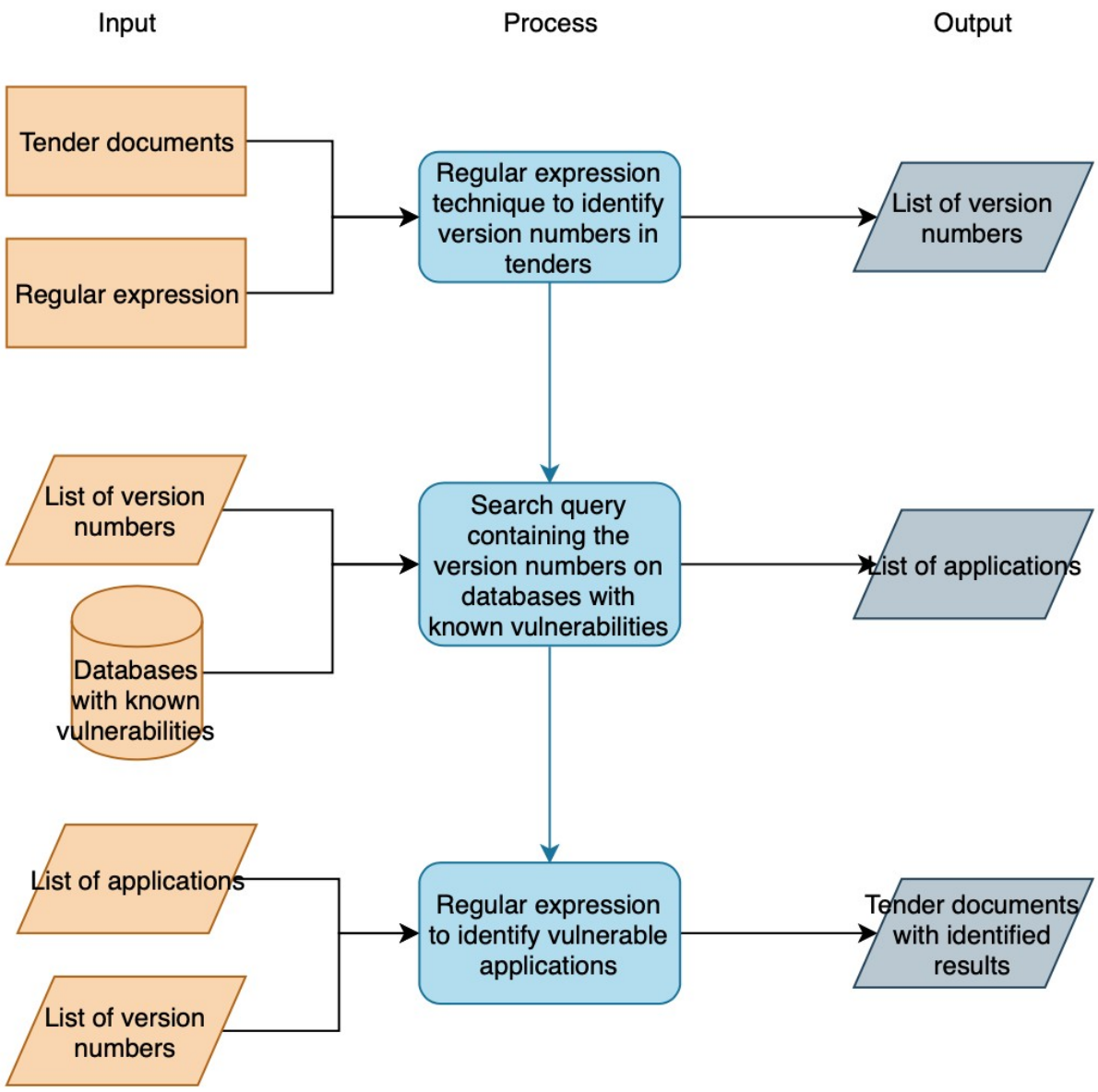

Figure 19: Visualisation of the technique 'Comparison with Known Information'

7.3.4 Optical character recognition

Images in tenders can be provided with text. These texts may contain sensitive information. This text cannot be read by using regular expressions because they are part of an image and are not included as separate text. The images must be analysed using optical character recognition (OCR). This technique ensures that images are scanned for the purpose of identifying the text in images. This text can then be analysed by means of regular expressions.

The techniques for OCR are quite advanced. There are various tools available that apply OCR. In a study by (S & A, 2015) different OCR tools were compared with each other. Google also has an API – the Google Cloud Vision API – that promises high accuracy OCR for multiple languages (Walker, Fujii, & Popat, 2018). The process how OCR works is visualized in the image below. Below the figure it is explained in general how it works.

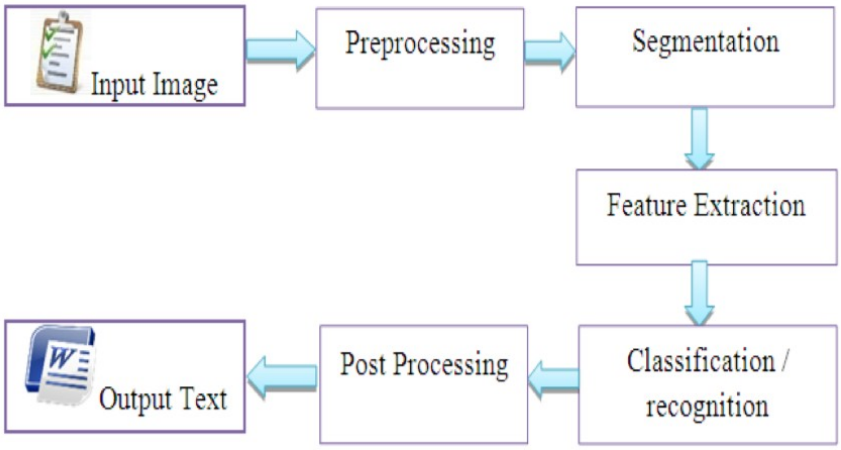

Figure 20: Visualization of OCR (S & A, 2015)

1. *"Input Image: Input image is digitalized images like a scanned or captured text image. It may be of different formats, i.e. JPG, PNG, BMP, GIF, TIFF and multi-page PDF files."* (S & A, 2015)
2. "*Preprocessing: Preprocessing techniques are important and essential for OCR system for image handling. These techniques are used to add or remove noises from the images, maintaining the correct contrast of the image, background removal which contains any scenes or watermarks. These are applied into images which enhance the image quality. This step is essential for OCR systems.*" (S & A, 2015)
3. "*Segmentation: The accuracy of OCR system mainly depends on the segmentation algorithm being used. Segmentation extracts pages, lines, words and then finally into characters from the text document images. Page segmentation separates graphics from text, a line segment is a part of a line that is bounded by two distinct end points and Word segmentation is the problem of dividing a string of written language into its component words. Character segmentation separates characters from others.*" (S & A, 2015)
4. "*Feature Extraction: Feature Extraction stage analyses a text segment and select a set of features that can be used to uniquely identify the text segment. This stage is used to extract the most relevant information from the text image which helps to recognize the characters in the text.*" (S & A, 2015)
5. "*Classification / recognition: Optical character Recognition is a most significant application. The main objective of Optical Character Recognition (OCR) is to classify the optical patterns like alphanumeric and other characters. The OCR is required when the information should be readable to both human and machine. Recognition has become essential for performing classification task.*" (S & A, 2015)
6. "*Post Processing: The post processing stage is used to increase recognition. The goal of post processing is to detect and correct grammatical misspellings in the OCR output text after the input image has been scanned and completely processed.*" (S & A, 2015)
7. "*Output Text: The result of the input images is displayed in the output text.*" (S & A, 2015)

In the perspective of this study, the input are in particular word, excel, and pdf documents. These documents contain images that have to be analysed. These images have to be filtered out of the documents before they are analysed by a OCR-tool. The output of the OCR-tool are text documents

which can be analysed using regular expressions, text mining, and comparison with known information.

7.3.5 Image recognition

Tenders frequently contain images such as network topologies, maps of physical locations, and process workflows. The previous technique does not ensure that all images are analysed. Images do not always contain text that can be recognized by means of optical character recognition. Images may also contain symbols (take the CORAS-symbols as an example) that have meaning. In order to interpret this, image recognition (also called symbol recognition) is required (US10068156B2, 2015). This technique has similarities with optical character recognition. The difference is that with this technique an attempt is made to recognize symbols that mean something that can be expressed in text, with the previous technique an attempt was made to recognize text. The biggest challenge with this technique is that symbols have different meanings.

According to a study by (Lladós, Valveny, Sánchez, & Martí, 2002) "*many symbol recognition techniques are available, but it is difficult to see a dominant one. The influence of the domain knowledge and the diagrammatic notation properties makes each family of applications to develop its own methods.*" "*Because of this wide range of graphic documents, each one with its own characteristic set of symbols, it is not easy to find a precise definition of a symbol. In a very general way, a symbol can be defined as a graphical entity with a particular meaning in the context of a specific application domain. Thus, and depending on the application, we can find different kinds of symbols according to their visual properties: simple 2D binary shapes composed of line segments (engineering, electronics, utility maps, architecture),a combination of line segments and solid shapes (musical scores), complex grey level or colour shapes (logos), silhouettes (geographic symbols), etc.*" (Lladós, Valveny, Sánchez, & Martí, 2002).

The differences in the meanings of symbols also occur in tenders. Tender information is provided by the contracting parties. They are therefore the ones who determine the context of the symbols. The purchaser may also have a role here to realize a centralization of which symbols are used and their meaning, for example in process diagrams, floor plans, but also network topologies. This makes the process of image recognition easier.

After the process of image recognition is done and the meaning of the images have been converted to text, the techniques regular expression, text mining, and comparison with known information can be used for further analyses.

7.4 Summary of 'Implementation of guidelines through automation'

The purpose of this chapter was to describe more specifically which techniques can be applied for the guidelines that focus on the reduction of sensitive information in tenders through automation. This has been done by first dividing the guidelines in techniques based on the type of identification of information that has to take place. Five different techniques were defined: regular expressions, comparison with known information, text mining, image recognition, and optical character recognition. The latter two techniques aim to convert images to text, in order to be able to analyse the images in tenders with the first three techniques. For each technique it was briefly described what the technique entails, how it works (conceptually), and how the technique fits in the context of reducing sensitive information in tenders.

We can conclude that:

- The guidelines can be implemented in a final product through automation. Multiple techniques are applicable for this, at least the five investigated in this chapter.
- At least the five techniques mentioned can be implemented in creating the ultimate solution to identify and reduce sensitive information in tenders. The techniques are concrete examples of how guidelines can be realised through automation.
- The technique regular expression can be used in every guideline in searching for specific keywords as well as specific formats.
- The technique comparison with known information can be used in at least two guidelines, to compare known information on vulnerabilities and information in the tenders.
- The technique text mining is an overall technique and can be used for ever every guideline. Text mining aims to correlate information in tenders. Information on its own may not be sensitive, but together with other information in the tenders it can be highly relevant for malicious parties to misuse.
- The technique image recognition can be used on several guidelines to analyse images. Images can contain symbols that have meaning. This meaning can be expressed in text by using this technique. This way, the text can be analysed using the previous three techniques.
- The technique optical character recognition can be used on several guidelines to analyse images.
- These techniques that are described in this chapter are recommendations for further research into developing a final solution.

By automation using the described techniques the established guidelines can be implemented to create a solution that can be used to identify and reduce sensitive information in tenders in order to prevent that adversaries gather sensitive information during the preparation of cyber-attacks.

7.4.1 Research methodology: define objectives of a solution

In this section a short reflection is done on the Design Science Research Methodology phase by Peffers et al. (2008).

After answering research questions four and five the DSRM phase 'Define Objectives of a Solution' is completed. The goal of this phase was to "*infer the objectives of a solution from the problem definition and knowledge of what is possible and feasible. The objectives can be quantitative, e.g., terms in which a desirable solution would be better than current ones, or qualitative, e.g., a description of how a new artefact is expected to support solutions to problems not hitherto addressed. The objectives should be inferred rationally from the problem specification. Resources required for this include knowledge of the state of problems and current solutions, if any, and their efficacy*" (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2008). The objectives that are described are the guidelines together with described techniques; a description of how a new artefact is expected to support a solution to the problem identified in the previous phase.

This research ends here. The next phase, 'Design & Development', focusses on the design and realisation of a final solution, where the guidelines should be taken into account. The results of this research serve as input for follow-up research, initiating the next DSRM phase.

In terms of the Three Cycle View by Hevner (2007) the guidelines together with the descriptions of the techniques are the requirements that the relevance cycles passes to the Design Cycle.

8 Conclusion & Recommendations

8.1 Opening

In this chapter the conclusions are drawn from the conducted research. The goal of the research is reflected on:

The main goal of this research is to establish guidelines aimed at identifying and reducing sensitive information in tenders, in order to prevent that malicious parties gather and use this information in the preparation of cyber attacks against tendering organizations.

Reflecting on the main goal of the research is done by giving concise answers to the five research questions in the next paragraph and by giving recommendations for follow-up research in the subsequent paragraph. The paragraphs that follow contain the limitations of the research and a personal reflection.

8.2 Conclusion

Each of the research questions defined in the first chapter of this research is included below as a section. In these sections the research questions are answered. In this way the main research goal is reflected on and the overall conclusion is formed.

8.2.1 What is the relationship between the information in tenders and cyber-attacks?

The intention of making more information public is to offer transparency, share knowledge, and deliver other benefits to society. This is also the case with information in tenders. Information in tenders on https://www.tenderned.nl is also publicly available. It can be accessed by anyone. As such it benefits organizations in providing fair chances in winning tenders. But although the intention of making information available to the public is positive, there are also disadvantages such as misuse of information.

According to the Cyber Kill Chain of Lockheed Martin (2000), adversaries gather information on their targets during the reconnaissance phase. The information that is gathered is used to prepare cyber-attacks or to enhance cyber-attacks. Open data is an important source of information for adversaries. Without alarming their targets, they collect relevant public information on their targets. The information that adversaries collect can be categorized in organizational-, human-, and technical information.

Because the information on TenderNed is also publicly available due to the principles on which the procurement law is based, adversaries are able to gather information from tenders and use it to prepare their cyber-attacks.

8.2.2 Which information that is relevant for adversaries can be collected from public tenders during the preparation of cyber-attacks?

The PRE-ATT&CK framework of MITRE (n.d.) describes how adversaries gather information during the reconnaissance activities. MITRE calls this the ‘techniques’ that attackers use. MITRE derives these techniques from forensic reports of attacks, red-team activities, and contributions from other people. The information types that are determined that malicious parties are looking for are derived from these techniques. This is done because the techniques clearly indicate what kind of information is needed to execute the technique. The information types therefore are: third party infrastructure services, domain and IP address space, external network trust dependencies,

firmware version, target logon/email address format, externally facing software applications technologies, languages, and dependencies, security defensive capabilities, web defensive services, network topology, groups/roles, centralization of IT management, physical locations, and business processes/tempo.

After determining which information types adversaries search for, a document analysis is done on existing tenders. This analysis examined whether these tenders contained information that could be related to the information types (the MITRE techniques). The results of the analysis showed that all types of information could be gathered from existing tenders. This means that adversaries can collect sensitive information from tenders during their reconnaissance activities that they can use in preparing cyber-attacks against the tendering organizations.

8.2.3 What are the risks that tendering organizations face if adversaries collect this information from tenders during the preparation of cyber-attacks?

In order to identify the risks that tendering organizations face, five interviews were held with security professionals and other professionals who have a job related to security. The professionals have analysed existing tenders based on their knowledge and expertise, with the aim of collecting information that they think is relevant for adversaries. They were then asked what risks they can identify when malicious parties have collected this information about their target.

The results of the interviews show that the interviewees were all able to collect information that they deem relevant for adversaries. The information could all be related to the determined information types. The interviewees were then able to identify multiple risks on the basis of the collected information. The risks that occur due to the sensitive information in tenders have an adverse effect on the confidentiality, integrity, and/ or availability of important organizational assets. Mainly the likelihood of these risks increases due to the fact that it becomes easier to perform cyber-attacks when information that is needed in the preparation is easy accessible in tenders. The impact of the risks differs, depending on the organizational assets that are damaged. However, severe damage may be caused.

8.2.4 Which guidelines can be determined that aim to prevent the risks that occur if adversaries collect sensitive information from tenders during the preparation of cyber-attacks?

In order to prevent the risks due to that sensitive information is included in tenders that are made public a solution must be created. This research aims to propose guidelines that can be used to develop a final solution to this problem. 26 guidelines have been determined which are qualitative descriptions of functionalities/requirements a final solution should meet. The guidelines have different goals they try to achieve: reducing the sensitive information in tenders, embedding a final solution in the context of the tender process and the people involved, and creating a solution such that it can be used in practice by the people involved in the tendering process. The guidelines are derived from four different sources: the information that stems from the previous research questions, the tender context, similar artefacts, and the interviews with senior purchasers. Besides that guidelines are derived from the interviews, the guidelines from the other three sources were evaluated during the interviews.

From the interviews it can be concluded that senior purchasers see added value in a solution based on the established guidelines that *identifies* sensitive information. The *reduction* however of the identified information is the responsibility of the client, according to the senior purchasers. A solution based on the established guidelines should therefor not automatically *reduce* identified information. All together, the guidelines successfully aim to prevent the risks that occur due to that adversaries can gather sensitive information from tenders during their reconnaissance activities.

The risks are mainly prevented due to the fact that the likelihood of the risks decreases because the information is excluded from tenders before being made public.

8.2.5 How can the established guidelines be implemented to create a solution that can be used to identify and reduce sensitive information in tenders in order to prevent that adversaries gather sensitive information during the preparation of cyber-attacks?

The guidelines that have the goal 'reducing the sensitive information in tenders' have been further examined. The purpose of this thorough investigation into these specific guidelines is to describe more specifically how these guidelines can be implemented in a final solution. Before doing research into the details the guidelines have been divided based on which type of identification of information that has to take place. Five different techniques are distinguished: regular expressions, comparison with known information, text mining, image recognition, and optical character recognition. The determined information types that malicious parties search for in the reconnaissance phase have been taken into account in dividing the guidelines. The first three techniques aim to identify sensitive information in tenders by analysing the text in tenders. The latter two techniques aim to convert the images in tenders to text. By doing this, this text from the images can be analysed by the first three techniques.

Applying these techniques results in identified information from tenders, based on the information types, which are relevant to adversaries. With this result, the goal of this research – to identify sensitive information in tenders that adversaries misuse in preparing their cyber-attacks against tendering organizations – has been achieved. By proposing the guidelines that help develop a solution that can do it in practice.

The goal of reducing the identified information has partly been achieved. The interviews showed that the removal of the identified information was not devoted to an automated solution, but to the person responsible for tendering.

8.3 Recommendations

As already explained in earlier chapters, this research initiates the follow-up research in which an end product is made based on the established guidelines. When creating this product, it is recommended to take into account all the established guidelines.

The five techniques proposed must be further investigated in order to realize a final solution. The technique text mining, in particular, needs further study because this technique ensures correlations between multiple information types and, for example, the various attack phases (Cyber Kill Chain) that the adversaries go through. Information not deemed sensitive might be identified by text mining in combination with other information in the tender.

In addition, research can be done into a different way how the identified problem can be solved. For example by not publishing tenders at all that contain sensitive information. The step to remove the identified information is then not necessary. This proposed research will have to be done by a researcher with a background in law and cyber security, because the principles why the tender is made public are based on legislation and regulations.

Lastly, more research is needed into the types of information that hackers abuse when preparing cyber-attacks. This research is done by various security organizations and also academics. However, it is difficult to determine what information has been misused and where it comes from. To reduce the digital footprint of organizations it is necessary to determine which information is misused.

8.4 Limitations of the research

Although the research has been conducted in a structured manner and decisions have been considered, there are certain limitations to the research. The limitations, other than the limitations mentioned throughout the earlier chapters, are listed below:

- Possibly tools already exist that can analyse documents based on certain types of information, which may or may not be used by adversaries. Research has been done into this, but after some time without results this has been omitted with the assumption that these do not exist.
- The research results in proposed guidelines for further research. The guidelines help in creating the final tool. Due to this the research is more conceptual and less concrete.

8.5 Personal reflection

This section reflects on the research. No separate reflection report has been drawn up.

Writing the thesis in English was a small challenge. Because it is not my native language it was sometimes difficult to describe something such that it was clear what was meant. Google Translate and the search for example sentences have helped with this. I am satisfied with the textual end result, although there are certainly still a number of language errors.

The subject 'tenders' was completely new to me, in contrast to open data and hacker activities. Conversations with several purchasers prior to the execution of this study were enlightening and very informative. Eventually I was able to gain new knowledge and to combine it with my own knowledge and experience. This has led to the interesting research question that has been thoroughly investigated.

Time management was the biggest challenge during the research process. It was difficult for me to combine work, study and private life. Starting a new job (in security) in the middle of the process and other private circumstances also did not help. The interim deadlines set by the CSA were therefore not met once. In the final phase of the research, extra days off of work ensured that the final deadline was met. Whether the number of prescribed hours (15 EC = 420 hours) has been exceeded by this I am not sure, but it definitely feels like it. A number of aspects of the research have taken a lot of time. The proposal for this study described that OSINT-tooling would be used during the study and that literature would be collected through systematic literature review instead of 'snowballing'. Both were ultimately not used in the study. This took a lot of time in the first phase of the research.

Learning and applying a new research method, design science, has taken a lot of effort. Although it has been very useful, I was also stuck several times by it. I struggled for a while with the phases proposed by Peffers et al. (2008). I desperately tried to plot the execution of my research on all phases of the DSRM methodology. This while I actually did not go through all phases of DSRM. Together with my supervisors I decided to restructure the research in such a way that not all phases are conducted, which is not a bad thing. The restructuring took a while.

Another challenge was the transition from the conceptual level (design science research) to making the research results concrete. Both supervisors mentioned this during progress meetings. The information provided by them helped in the realization of the result.

A limited number of interviews (2) were held with purchasers about the practical side of the solution. No interviews were held with purchasers from non-central government purchasing organizations. This partly questions the conclusion of the study. However, the purchasers

interviewed indicated that the non-central government purchasing organizations do work according to the same process and that the maturity of these organizations is probably lower than that of the central government purchasing organizations. Taking this into account, it can be stated that the conclusion of the study is correct.

As intended with Design Science, the final tool must be developed. After conducting the interviews and talking about the further development of the artefact based on the guidelines I proposed, I get eager to continue and to seek cooperation with others to realize a prototype. Many months can still be spent on this. For the writing of this thesis, however, this is not required. I am satisfied with the end result.

Bibliography

Bazzell, M. (2018). Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information. N/A: CreateSpace Independent Publishing Platform.

Chulinin, Y., & Vatlin, Y. (2015). US10068156B2. Rusland: ABBYY Production LLC.

Europa decentraal. (2000, January 1). Aanbesteden voor decentrale overheden - Europa decentraal. Retrieved 23 August 2019, from https://europadecentraal.nl/onderwerp/aanbestedingen

European Data Portal. (2000). What is open data? Retrieved 25 August 2019, from https://www.europeandataportal.eu/elearning/en/module1/

European Data Portal. (2016, November 5). Wat we doen - Europees Data Portaal - European Data Portal. Retrieved 25 August 2019, from https://www.europeandataportal.eu/nl/what-we-do/our-activities

fosteropenscience. (2018). Text mining 101,. Retrieved 24 November 2019, from https://www.fosteropenscience.eu/content/text-mining-101

Goyvaerts, J. (2018, March 20). How to Find or Validate an IP Address. Retrieved 20 November 2019, from https://www.regular-expressions.info/ip.html

Goyvaerts, J. (2019, October 28). How to Find or Validate an Email Address. Retrieved 20 November 2019, from https://www.regular-expressions.info/email.html

Goyvaerts, Jan. (2019, September 25). Full Documentation to RegexMagic: The Regular Expression Generator. Retrieved 22 November 2019, from http://www.regexmagic.com/manual.html#xmpversion

Hevner, A. R. (2007). A Three Cycle View of Design Science Research. Scandinavian Journal of Information Systems, 19(2), 1–6. Retrieved from https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1017&context=sjis

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). DESIGN SCIENCE IN INFORMATION SYSTEMS RESEARCH. MIS Quarterly, 28(1), 75–105. Retrieved from https://wise.vub.ac.be/sites/default/files/thesis_info/design_science.pdf

Internet World Stats. (2019). Internet Growth Statistics 1995 to 2019 - the Global Village Online. Retrieved 7 September 2019, from https://www.internetworldstats.com/emarketing.htm

Investopedia. (2019, April 9). Tender. Retrieved 23 August 2019, from https://www.investopedia.com/terms/t/tender.asp

Labohm, M. A. J. (2018). De effecten van open overheidsdata op social engineering gericht op de Nederlandse overheid (1st ed.). Groningen: Open Universiteit, faculteit Management, Science & Technology Masteropleiding Business Process Management & IT.

Leiden University. (2000, January 1). Thesis and paper writing - Leiden University. Retrieved 25 April 2019, from https://www.student.universiteitleiden.nl/en/study--studying/study/educational-information/thesis-and-paper-writing/humanities/museums-and-collections-ma?cd=guest

linguamatics. (2019, September 12). What is Text Mining, Text Analytics and Natural Language Processing? Retrieved 24 November 2019, from https://www.linguamatics.com/what-text-mining-text-analytics-and-natural-language-processing

Lladós, J., Valveny, E., Sánchez, G., & Martí, E. (2002). Symbol Recognition: Current Advances and Perspectives. Lecture Notes in Computer Science, 104–128. https://doi.org/10.1007/3-540-45868-9_9

Lockheed Martin. (2015a). Cyber Kill Chain [Afbeelding]. Retrieved from https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/photo/cyber/THE-CYBER-KILL-CHAIN-body.png.pc-adaptive.full.medium.png

Lockheed Martin. (2015b). Gaining the Advantage Cyber Kill Chain. Retrieved 25 August 2019, from https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/Gaining_the_Advantage_Cyber_Kill_Chain.pdf

Løkke, A., & Dissing Sørensen, P. (2014). Theory Testing Using Case Studies. The Electronic Journal of Business Research Methods, 12(1), 66–74. Retrieved from www.ejbrm.com

Mauri, M., Mulas, A., & Ariu, D. (2000, January 1). The Dark Side of Open Data. Retrieved 15 May 2019, from http://ceur-ws.org/Vol-1748/paper-19.pdf

Merriam-Webster. (2000, January 1). Dictionary. Retrieved 30 April 2019, from https://www.merriam-webster.com/dictionary/double-edged%20sword

MITRE. (2000a). MITRE ATT&CKTM. Retrieved 26 September 2019, from https://attack.mitre.org

MITRE. (2000b). PRE-ATT&CK | MITRE ATT&CKTM. Retrieved 26 August 2019, from https://attack.mitre.org/resources/pre-introduction/

MITRE. (2000c). Tactics: PRE-ATT&CK | MITRE ATT&CKTM. Retrieved 26 August 2019, from https://attack.mitre.org/tactics/pre/

MITRE. (2000d). Techniques: PRE-ATT&CK | MITRE ATT&CKTM. Retrieved 26 August 2019, from https://attack.mitre.org/techniques/pre/

Nunamaker, J. F., Briggs, R. O., Derrick, D. C., & Schwabe, G. (2015). The Last Research Mile: Achieving Both Rigor and Relevance in Information Systems Research. Journal of Management Information Systems, 32(3), 10–47. https://doi.org/10.1080/07421222.2015.1094961

Open Knowledge Foundation. (2000a). About. Retrieved 24 August 2019, from https://okfn.org/about/

Open Knowledge Foundation. (2000b). What is Open Data? Retrieved 24 August 2019, from http://opendatahandbook.org/guide/en/what-is-open-data/

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2008). A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems, Volume 24(Issue 3, Winter 2007-8, pp. 45-78.), 45–78. https://doi.org/DesignScienceResearchMethodology(DSRM)byPfeffers

PIANOo. (2000a). Wat zijn de beginselen van Europees aanbesteden? Retrieved 24 August 2019, from https://www.pianoo.nl/nl/metrokaart/wat-zijn-beginselen-van-europees-aanbesteden

PIANOo. (2000b, January 1). Aankondigen. Retrieved 24 August 2019, from https://www.pianoo.nl/nl/inkoopproces/fase-2-doorlopen/aankondigen

PIANOo. (2000c, January 1). Aankondigen. Retrieved 24 August 2019, from https://www.pianoo.nl/nl/inkoopproces/fase-2-doorlopen/aankondigen

PIANOo. (2000d, January 1). Fase 1: Voorbereiden inkoopopdracht. Retrieved 21 August 2019, from https://www.pianoo.nl/nl/inkoopproces/fase-1-voorbereiden-inkoopopdracht

PIANOo. (2000e, January 1). Government Procurement Agreement (GPA). Retrieved 20 August 2019, from https://www.pianoo.nl/nl/regelgeving/europese-richtlijnen/government-procurement-agreement-gpa

Pols, P. (2017). The Unified Kill Chain (1st ed.). Retrieved from https://www.csacademy.nl/images/scripties/2018/Paul-Pols---The-Unified-Kill-Chain.pdf

Rijksoverheid. (2018, August 6). Aanbestedingsregels. Retrieved 15 October 2019, from https://www.rijksoverheid.nl/onderwerpen/aanbesteden/aanbestedingsregels

RSA. (2015). Reconnaissance. Retrieved from https://drive.google.com/file/d/0B3tdhdmrVDEwQ3ptdHJKb3N1NjA/view

S, V., & A, S. (2015). Performance Comparison of OCR Tools. International Journal of UbiComp, 6(3), 19–30. https://doi.org/10.5121/iju.2015.6303

Sanghvi, H. P., & Dahiya, M. S. (2013). Cyber Reconnaissance: An Alarm before Cyber Attack. International Journal of Computer Applications, 63(6). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.278.5965&rep=rep1&type=pdf

Sharma, D., Kumar, B., & Chand, S. (2018). Trend Analysis in Machine Learning Research Using Text Mining. 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN). https://doi.org/10.1109/icacccn.2018.8748686

Shinde, R., & Gill, P. C. (2014). Pattern Discovery Techniques for the Text Mining and its Applications. International Journal of Science and Research (IJSR), 3(5). Retrieved from https://pdfs.semanticscholar.org/26f3/5cced5e66f90a68712d301526de455a37ada.pdf

Sourceforge. (2015, November 16). The CORAS Method. Retrieved 11 October 2019, from http://coras.sourceforge.net

Triad 3. (2016, March 9). An Introduction to Document Analysis. Retrieved 15 September 2019, from https://lled500.trubox.ca/2016/244

Tuominen, S. (2019). OPEN SOURCE INTELLIGENCE AND OSINT APPLICATIONS (Thesis). Retrieved from Oulu

Walker, J., Fujii, Y., & Popat, A. C. (2018). A Web-Based OCR Service for Documents. Short Papers Booklet DAS 2018, 21–22. Retrieved from https://das2018.cvl.tuwien.ac.at/media/filer_public/85/fd/85fd4698-040f-45f4-8fcc-56d66533b82d/das2018_short_papers.pdf#page=23

Wired. (2017, June 3). An Unprecedented Look at Stuxnet, the World's First Digital Weapon. Retrieved 26 August 2019, from https://www.wired.com/2014/11/countdown-to-zero-day-stuxnet/

Zuiderwijk, A., & Janssen, M. F. W. H. A. (2014). The negative effects of open government data. Proceedings of the 15th Annual International Conference on Digital Government Research, 147–152. Retrieved from https://www.researchgate.net/publication/264434508_The_negative_effects_of_open_government_data_-_Investigating_the_dark_side_of_open_data

Appendix

Annex A

Table 11: Relevancy of the PRE-ATT&CK Tactics

PRE-ATT&CK TACTIC	Description (MITRE, n.d.)	Relevant	Motivation
Priority Definition Planning	*"Priority definition planning consists of the process of determining the set of Key Intelligence Topics (KIT) or Key Intelligence Questions (KIQ) required for meeting key strategic, operational, or tactical goals. Leadership outlines the priority definition (may be considered a goal) around which the adversary designs target selection and a plan to achieve. An analyst may outline the priority definition when in the course of determining gaps in existing KITs or KIQs."*	No	The study focusses on the collection of open data. This happens after setting the priorities.
Priority Definition Direction	*"Priority definition direction consists of the process of collecting and assigning requirements for meeting Key Intelligence Topics (KIT) or Key Intelligence Questions (KIQ) as determined by leadership."*	No	The study focusses on the collection of open data. This happens after setting the priorities.
Target Selection	*"Target selection consists of an iterative process in which an adversary determines a target by first beginning at the strategic level and then narrowing down operationally and tactically until a specific target is chosen. A target may be defined as an entity or object that performs a function considered for possible engagement or other action."*	No	The study focusses on the collection of open data. This happens after or prior the target selection.
Technical Information Gathering	*"Technical information gathering consists of the process of identifying critical technical elements of intelligence an adversary will need about a target in order to best attack. Technical intelligence gathering includes, but is not limited to, understanding the target's network architecture, IP space, network*	Yes	The study focusses on the collection and gathering of open data. This includes technical information.

	services, email format, and security procedures."		
People Information Gathering	*"People Information Gathering consists of the process of identifying critical personnel elements of intelligence an adversary will need about a target in order to best attack. People intelligence gathering focuses on identifying key personnel or individuals with critical accesses in order to best approach a target for attack. It may involve aspects of social engineering, elicitation, mining social media sources, or be thought of as understanding the personnel element of competitive intelligence."*	Yes	The study focusses on the collection and gathering of open data. This includes people information.
Organizational Information Gathering	*"Organizational information gathering consists of the process of identifying critical organizational elements of intelligence an adversary will need about a target in order to best attack. Similar to competitive intelligence, organizational intelligence gathering focuses on understanding the operational tempo of an organization and gathering a deep understanding of the organization and how it operates, in order to best develop a strategy to target it."*	Yes	The study focusses on the collection and gathering of open data. This includes organization information.
Technical Weakness Identification	*"Technical weakness identification consists of identifying and analysing weaknesses and vulnerabilities collected during the intelligence gathering phases to determine best approach based on technical complexity and adversary priorities (e.g., expediency, stealthiness)."*	No	The identification of weaknesses is based on the information that is gathered. The less information gathered by hackers, the less weaknesses they identify.
People Weakness Identification	*"People weakness identification consists of identifying and analysing weaknesses and vulnerabilities from the intelligence gathering phases which can be leveraged to gain access to target or intermediate target persons of interest or social trust relationships."*	No	The identification of weaknesses is based on the information that is gathered. The less information gathered by hackers, the less weaknesses they identify.

Organizational Weakness Identification	*"Organizational weakness identification consists of identifying and analysing weaknesses and vulnerabilities from the intelligence gathering phases which can be leveraged to gain access to target or intermediate target organizations of interest."*	No	The identification of weaknesses is based on the information that is gathered. The less information gathered by hackers, the less weaknesses they identify.
Adversary OPSEC	*"Adversary OPSEC consists of the use of various technologies or 3rd party services to obfuscate, hide, or blend in with accepted network traffic or system behaviour. The adversary may use these techniques to evade defences, reduce attribution, minimize discovery, and/or increase the time and effort required to analyse."*	No	This tactic is not about the collection of data. It focusses on entering the targeted network.
Establish & Maintain Infrastructure	*"Establishing and maintaining infrastructure consists of building, purchasing, co-opting, and maintaining systems and services used to conduct cyber operations. An adversary will need to establish infrastructure used to communicate with and control assets used throughout the course of their operations."*	No	This tactic is about the establishment and maintenance of the attacker's own attack infrastructure.
Persona Development	*"Persona development consists of the development of public information, presence, history and appropriate affiliations. This development could be applied to social media, website, or other publicly available information that could be referenced and scrutinized for legitimacy over the course of an operation using that persona or identity."*	No	This tactic consists of activities to impersonate a persona.
Build Capabilities	*"Building capabilities consists of developing and/or acquiring the software, data and techniques used at different phases of an operation. This is the process of identifying development requirements and implementing solutions such as malware, delivery mechanisms,*	No	This is a clear weaponize-tactic as it is about the development the building blocks needed for the attacks.

	obfuscation/cryptographic protections, and call back and O&M functions."		
Test Capabilities	*"Testing capabilities takes place when adversaries may need to test capabilities externally to refine development goals and criteria and to ensure success during an operation. Certain testing may be done after a capability is staged."*	No	This is a clear weaponize-tactic as it is about the testing the building blocks needed for the attacks.
Stage Capabilities	*"Staging capabilities consists of preparing operational environment required to conduct the operation. This includes activities such as deploying software, uploading data, enabling command and control infrastructure."*	No	This is a clear weaponize-tactic as it is about the staging the building blocks needed for the attacks.

Annex B

Introduction mail send to the interviewees

Hello [name],

Before completing my Cyber Security program at the Cyber Security Academy (collaboration between Delft University of Technology, Leiden University and The Hague University of Applied Sciences), I am investigating whether information in tenders is relevant for malicious parties in preparing cyber-attacks on government organizations. If this can be demonstrated, I provide a heads-up for a mechanism that prevents this.

Because of your expertise in the field of security or procurement, I approach you to participate in this research. I have already received the confirmation of your participation by email or verbally. Officials within the Dutch government (also external people) and people from outside the government have also been approached.

Your participation is a kind of interview / case study. Your participation contributes to determining whether information is included in tenders that is relevant to malicious parties. Participation takes about half an hour and can be performed independently at a place of your choice. The questions are to be answered independently.

You will find tender documents and a questionnaire attached to this mail. A number of questions must be answered on the basis of the tender documents that you have received. The structure of the questionnaire speaks for itself. I look forward to receiving your response before 4 October. If this is not possible then please October 11 at the latest.

The answers to the questions are treated confidentially and are included anonymously in my thesis.

I can be reached via this email address or on [phone number].

Thank you very much for your participation and kind regards,

Arjan Sibma

Mail of gratitude send to the interviewees

Hello [name],

You have participated in my interview and thereby have contributed to my research. The results are processed in my thesis. If you are interested in the results of the research, please let me know. I will send you the final version when ready (mid-January).

If you have any comments or questions, please let me know. I can be reached via this email address.

Thank you again for your participation and kind regards,

Arjan Sibma

Annex C

Questionnaire

The text below was added as an attachment to the first mail in Annex B.

Questions 1 to 4 are for the introduction of yourself.

Questions 5 to 7 are about the tender documents in the attachment of the e-mail that also included this questionnaire.

The answers can be given below the questions.

1. ***What is the position that you currently hold?***
2. ***How many years have you held this position?***
3. ***Have you held other positions related to security? If yes which one?***
4. ***Have you held other positions related to procurement? If yes which one?***

The research focuses on "passive reconnaissance": the gathering of information about an object without making contact with the object in any way. The information about tenders is public. This information can be read and related documents can be downloaded for subsequent reading. These documents may contain information that is used by malicious parties to prepare cyber-attacks.

You will find tender documents in the attachment of the e-mail you have received. The questions below concern these documents. These documents come from https://www.tenderned.nl and can be accessed by anyone with internet access. Assume that an adversary is targeting the organization concerning this tender.

5. ***Can you find information in the tender documents that may be relevant for malicious parties when preparing cyber-attacks?***
6. ***If so, can you provide specific cases (maximum 5) of information in the documents below? This can be done by taking a screenshot, describing of what you have found, copying text and / or listing the page numbers.***
7. ***If you have found specific cases: can you give a description of possible risks of 1 case if a malicious person has access to this information?***

Thanks for your participation.

Arjan Sibma

Annex D

Interview structure

1. Introduction interviewer and interviewee
    a) Confidentiality of the interview
2. Explanation of the thesis goal and interview goal
    a) Design science
    b) Problem statement
    c) Goal of the research
    d) The guidelines
3. Questions on the tender process
4. Questions on the implementation of a 'scanning' tool within the process
5. Questions on the content of the guidelines

No invitation mail and mail of gratitude have been sent because the interviewers have been approached in person.

Interview questions

The questions are not set in stone. Depending on the answers of the interviewee other questions can be asked and questions can be altered.

Questions on the tender process

1. *Can you describe the tender process shortly, till the point the tender documents are made public on TenderNed?*
    1. *Follow-up: What is your opinion on making all this information public?*
    2. *Follow-up: Is it correct that this process does not include an activity that checks the 'to be published' tender documents on sensitivity of information?*
    3. *Follow-up: Are there activities that determine if the tenders conclude privacy sensitive information?*
    4. *Follow-up: Are there security people involved in the tendering process of IT-components?*
    5. *Follow-up: Is it correct that all tender documents pass the purchaser involved in the tender before being made public?*
    6. *Follow-up: Who uploads the documents on TenderNed?*

Questions on the implementation of a 'scanning' tool within the process

2. *Are you familiar with the fact that tenders can conclude sensitive information (relevant for adversaries)? Explain to the interviewee if the answer is No.*
3. *The plagiarism scanner is used by researchers, teachers, and professors, to determine whether fraud has been committed in the sense that text has been improperly copied in a publication. The scanner takes the document that is being published as input. The scanner*

uses a large database containing all kinds of previous publications and analyses the input, whereby a comparison is made between the input and the text in the publications in the database. The output of the scanner is the document (the input) containing highlighted text which is also found in publications in the database and a percentage of the amount of highlighted text in relation to the total document. The researcher, teacher, or professor must then determine from his or her own professionalism whether plagiarism has been committed or sources have been handled incorrectly, for example. What do you think of an a-like scanning tool to scan tender documents on sensitive information?

1. *Follow-up: What do you think would be the best step in the process for such a tool?*
2. *Follow-up: The best thing would be if someone without a security background could use such a tool. Do you think purchasers can use this tool in the tendering process? Because buyers are the central person in the process.*

Questions on the content of the guidelines

4. *My research until now has identified certain information types that hackers search for. What do you think that information types are that tender documents should not contain? F.e. personal information.*
    1. *Follow-up: Which information is this?*
    2. *Follow-up: Why should a tender not contain this information?*

Closing questions

5. *Is it possible to receive from you tender documents (IT-related) that are not yet published? I will use them in a case study with the goal to identify and reduce sensitive information. After the analysis, together we can determine if the tender document is still usable.*
6. *Do you have any questions, comments, or suggestions for the research and the outcome of this research?*

Annex E

The interview reports below are written in Dutch. This is done because the interviewees are Dutch.

Interview with Senior Purchaser number 1

Datum en tijd interview: 06-11-2019 – 11:00 tot 12:00

Aanwezig: Arjan Sibma (interviewer) en [...] (geïnterviewde, senior inkoper)

Onderwerp: thesis A. Sibma

Dit document beschrijft het verloop van het interview en dient als gespreksverslag. Dit gespreksverslag wordt voor hoor en wederhoor afgestemd met de geïnterviewde. Het gespreksverslag wordt opgenomen als bijlage (in het Engels) bij het onderzoeksrapport waarbij de naam van de geïnterviewde wordt weggelaten.

Introductie

Het interview begon met een korte introductie van beide personen.

De geïnterviewde houdt zich bezig met Europese Aanbestedingen waarbij raamovereenkomsten worden gemaakt. De organisaties waarop dit betrekking heeft zijn weggelaten. Deze raamovereenkomsten bevatten de randvoorwaarden voor vervolg trajecten waarbij binnen de raamovereenkomsten opdrachten worden verstrekt.

Arjan lichtte vervolgens toe dat dat de resultaten van dit interview vertrouwelijk zijn. Namen en bedrijfsnamen worden weggelaten.

Vervolgens is het onderzoek toegelicht:

1. *Methodologie Design Science: op welke wijze het probleem is geïdentificeerd en hoe naar een oplossing is toegewerkt.*
2. *Probleemstelling: wat is de probleemstelling van het onderzoek. Uit het literatuur onderzoek, de document analyse en interviews met security specialisten is gebleken dat aanbestedingen (TenderNed) relevante informatie bevatten voor kwaadwillenden.*
3. *De oplossing: de oplossing is een opzet voor een eindproduct. De getoonde oplossing is een soort 'checklist' die gebruikt kan worden door een inkoper. Het doel is echter wel om dit te automatiseren. Het voorbeeld van de plagiaat scanner is toegelicht.*

Vraag 1: Kan je het proces beschrijven van 'aanbesteden' van het begin tot het moment dat aanbestedingsinformatie openbaar wordt gemaakt?

Bij complexe Europese aanbestedingen worden een stuurgroep en een werkgroep opgezet. De stuurgroep bestaat uit een afvaardiging van de klant(en) die beslissing bevoegd zijn. Zij zijn betrokken bij strategievorming en sturen op het proces. De werkgroep bestaat uit een inkoopadviseur, een projectleider en materie deskundigen. De werkgroep is uitvoerend. Wie de materie deskundigen zijn is afhankelijk van wat wordt aanbesteed. De stuurgroep is verantwoordelijk voor het helder formuleren van de opdracht. De projectleider is verantwoordelijk voor de uitvoering van de opdracht en dient daarbij de juiste personen te betrekken. Bij Europese aanbestedingen op gebied van IT wordt standaard een security medewerker betrokken.

Er vindt een intake plaats met de opdrachtgever om de behoefte helder te krijgen. Ook wordt er een startformulier (A en B) ingevuld. Formulier A wordt door de klant ingevuld en formulier B

gezamenlijk. In dit startformulier staan onder andere aandachtspunten voor het traject waar op gelet moet worden. Voorbeelden zijn: is er sprake van privacygevoelige informatie verwerking en of het een risicovol traject is ***(vraag 1.3)****.*

*De opdracht gevende partij voorziet de inkoper van informatie aangaande de aanbesteding. De inkoper verwerkt deze informatie in documenten die bedoeld zijn om gepubliceerd te worden. Dit wordt gedaan met behulp van templates. De geïnterviewde geeft aan dat hij regelmatig de documenten afstemt met de Security Officer van de opdracht gevende partij en die van zijn eigen organisatie (****vraag 1.2****). Met name zodat de juiste eisen worden opgenomen in de aanbestedingsdocumenten, maar ook om een controle uit te laten voeren of de documenten gevoelige informatie bevatten. De geïnterviewde is zich bewust van het feit dat bepaalde informatie niet openbaar dient te zijn (****vraag 1.1*** *en* ***vraag 2****). Dit doet de geïnterviewde voordat de documenten openbaar worden gemaakt. De geïnterviewde geeft wel aan dat er niet een standaard een controle wordt gedaan op gevoelige informatie in de aanbestedingsdocumenten in het proces (vraag 1.2 en 1.4).Hierbij wordt uitgegaan van de professionaliteit van de klant en de inkoper. De klant blijft verantwoordelijk voor de informatie die hij aanlevert. Dit geldt ook voor privacygevoelige informatie – er is niet een specifieke controle op de documenten die openbaar gemaakt worden of ze privacygevoelige informatie bevatten (****vraag 1.3****).*

*Documenten die de geïnterviewde wil uploaden naar TenderNed gaan via CTM (Complete Tender Management). CTM is een portal waarop de inkoper de documenten upload. De documenten worden daarmee gelijk ook op TenderNed openbaar gemaakt. Vervolgens kunnen organisaties zich oriënteren en inschrijven op de aanbesteding. Alleen de inkopers uploaden documenten op CTM → TenderNed (****vraag 1.5*** *en* ***vraag 1.6****).*

De geïnterviewde geeft verder aan dat er verschillen zitten in volwassenheid tussen de diverse inkooporganisaties. Aanbestedingen op TenderNed zijn ook afkomstig van niet-Rijksinkoop organisaties. Qua volwassenheid lopen deze wellicht achter op de Rijksinkoop organisaties.

Vraag 2: *zie alinea 3 onder vraag 1.*

Vraag 3 na toelichting guidelines en werking plagiaat scanner: Wat vindt jij van een dergelijke tool om documenten te scannen op gevoelige informatie?

De geïnterviewde geeft aan dat hij een dergelijke tool op prijs zou stellen. De tool zou een middel zijn om effectiever en efficiënter documenten door te nemen op gevoelige informatie. Nu gebeurt dit impliciet door de personen die betrokken zijn bij het proces en mogelijk door een security officer. Dit betekent ook dat de impliciete controle gedaan wordt aan de hand van de kennis en ervaring van de betrokken personen. Er is geen baseline. De geïnterviewde geeft ook aan dat de security officer niet altijd bekend is met de specifieke aanbesteding wat het lastig maakt om informatie te schatten zonder context.

*De geïnterviewde geeft aan dat een dergelijke tool gebruikt zou kunnen worden mits deze geautomatiseerd is. Manueel zal het veel tijd in beslag nemen. De tool kan in het proces worden opgenomen voordat de documenten worden geüpload naar CTM en TenderNed (****vraag 3.1****). De geïnterviewde geeft ook aan dat een inkoper de tool zou kunnen hanteren (****vraag 3.2****). Het resultaat wat wordt teruggegeven door de tool (de documenten met gearceerde tekst dat dus mogelijk relevant is voor kwaadwillenden) zou dan naar de opdrachtgever gestuurd worden. De opdrachtgever is namelijk verantwoordelijk voor de documenten en de inhoud daarvan. De opdrachtgever bepaald vervolgens welke informatie uit de documenten gehaald wordt. De inkoper upload de documenten.*

Vraag 4: Wat zijn volgens jou nog specifieke informatie typen waarop de tool kan scannen?

De geïnterviewde geeft aan dat het lastig in te schatten is voor hem welke informatie dit moet zijn. Privacygevoelige informatie zoals namen zouden opgenomen kunnen worden omdat deze niet in aanbestedingsdocumenten mogen staan in verband met de AVG (vraag 4.1 en 4.2). De geïnterviewde geeft aan dat zijn privé mobiele telefoonnummer per ongeluk op TenderNed terecht is gekomen.

Vraag 5: Is het mogelijk om aanbestedingsdocumenten te ontvangen die nog niet gepubliceerd zijn om deze te analyseren met behulp van de tool?

Nee. Momenteel geen documenten die in het voorstadium zijn van publiceren.

Vraag 6: Heb je nog vragen, commentaar of suggesties voor dit onderzoek?

Nee. Wel ontvangt de geïnterviewde graag het resultaat.

Interview with Senior Purchaser number 2

Datum en tijd interview: 07-11-2019 – 10:00 tot 11:00

Aanwezig: Arjan Sibma (interviewer) en [...] (geïnterviewde, senior inkoper)

Onderwerp: thesis A. Sibma

Dit document beschrijft het verloop van het interview en dient als gespreksverslag. Dit gespreksverslag wordt voor hoor en wederhoor afgestemd met de geïnterviewde. Het gespreksverslag wordt opgenomen als bijlage (in het Engels) bij het onderzoeksrapport waarbij de naam van de geïnterviewde wordt weggelaten.

Introductie

Het interview begon met een korte introductie van beide personen.

De geïnterviewde houdt zich bezig met aanbestedingen binnen raamovereenkomsten – minicompetities. De raamovereenkomsten vormen voor de geïnterviewde de randvoorwaarden voor zijn opdrachten. De organisaties waarvoor de geïnterviewde dit doet zijn weggelaten in dit verslag.

Arjan lichtte vervolgens toe dat de resultaten van dit interview vertrouwelijk zijn. Namen en bedrijfsnamen worden weggelaten.

Vervolgens is het onderzoek toegelicht:

1. *Methodologie Design Science: op welke wijze het probleem is geïdentificeerd en hoe naar een oplossing is toegewerkt.*
2. *Probleemstelling: hoe het probleem is onderzocht. Uit het literatuur onderzoek, de document analyse en interviews met security specialisten is gebleken dat aanbestedingen (TenderNed) relevante informatie bevatten voor kwaadwillenden.*
3. *De oplossing: de oplossing is een opzet voor een eindproduct. De getoonde oplossing is een soort 'checklist' die gebruikt kan worden door een inkoper. Het doel is echter om dit te automatiseren. Het voorbeeld van de plagiaat scanner is toegelicht.*

Vraag 1: Kan je het proces beschrijven van 'aanbesteden' van het begin tot het moment dat aanbestedingsinformatie openbaar wordt gemaakt?

*Aan het begin heeft de geïnterviewde eerst het verschil duidelijk gemaakt tussen de aanbestedingen waar hij zich mee bezig houdt en de aanbestedingen waar de eerder geïnterviewde persoon zich mee bezig houdt. De geïnterviewde houdt zich bezig met 'minicompetities'. Dit zijn kleinere opdrachten die zich afspelen binnen grotere raamovereenkomsten. De eerder geïnterviewde persoon houdt zich bezig met de raamovereenkomsten. De raamovereenkomsten scheppen de randvoorwaarden voor de geïnterviewde om binnen te blijven, om andere opdrachten af te roepen. De raamovereenkomsten worden afgesloten met bijvoorbeeld 3 of 4 partijen. Vervolgens worden de minicompetities binnen de raamovereenkomsten uitgezet, waarbij de behoefte van de opdracht gever duidelijker is. In de raamovereenkomst is de behoefte nog 'vaag' beschreven. De partijen waarmee de raamovereenkomst is afgesloten kunnen zich inschrijven op de minicompetities. De geïnterviewde geeft daarbij nog een ander belangrijk (in relatie met dit onderzoek) verschil aan, namelijk dat de documenten horende bij de minicompetities niet openbaar worden gemaakt op TenderNed. Deze worden gepubliceerd op CTM. De partijen die inschrijven hebben ook toegang tot CTM. Deze documenten zijn dus niet openbaar. De geïnterviewde geeft aan dat deze documenten dus wel verstrekt worden aan andere partijen (de inschrijvers): het is dus nog steeds belangrijk om geen gevoelige informatie hierin op te nemen (**vraag 2**).*

Naast de minicompetities voert de geïnterviewde ook marktconsultaties uit. Hierbij worden documenten wel op TenderNed gezet. Het doel van een marktconsultatie is om de markt, de geïnteresseerde partijen, te consulteren over een aankomende opdracht. Hierbij is het dus wel van belang om rekening te houden met welke informatie wordt gepubliceerd. De geïnterviewde geeft aan dat door middel van vraag en antwoord (Nota van inlichtingen) de markt de inkoper kan consulteren.

Bij de minicompetities die de geïnterviewde uitvoert wordt bij de start een Startformulier ingevuld. Dit formulier bestaat uit twee delen: A en B. Deel A wordt ingevuld door de klant en deel B gezamenlijk met de inkoper. In deze formulieren worden aandachtspunten benoemd waar rekening mee gehouden moet worden bij het traject. Denk hierbij aan:

- *worden er persoonsgegevens verwerkt?*
- *is het een risicovol traject?*
- *rechtmatigheid.*

*Specifieke aandachtspunten met betrekking tot security worden hier niet benoemd (**vraag 1.2**). De geïnterviewde geeft aan dat dit startformulier wordt ingevuld op basis van de kennis en expertise van de betrokken personen. Als iemand bijvoorbeeld niet weet dat er persoonsgegevens verwerkt worden dan wordt dit niet aangegeven.*

*De geïnterviewde verteld dat bij het traject een werkgroep wordt opgericht. Deze groep bestaat uit een projectleider, de inkoper en andere materie deskundigen. De projectleider is verantwoordelijk voor het betrekken van de juiste personen bij het traject. Dit betekent dat er niet standaard een security medewerker betrokken is bij het traject (**vraag 1.4**). Dit is afhankelijk van wat er wordt aanbesteed.*

De opdracht gevende partij voorziet de inkoper van informatie over de aanbesteding. De inkoper verwerkt deze informatie in documenten die gepubliceerd worden via CTM. De geïnterviewde geeft aan dat een veel gebruikte werkwijze het hergebruiken is van documentatie en informatie binnen het netwerk van inkopers. Een positief effect hiervan is dat door het raadplegen van documentatie van andere inkopers er mogelijk nieuwe aandachtspunten aan bod komen. De

*geïnterviewde geeft het voorbeeld dat bij organisatie […] een aanbesteding loopt voor een bepaald product. Vanuit zijn netwerk heeft hij meegekregen dat een andere organisatie ook een aanbesteding heeft gehad voor hetzelfde type product. De inkoper vraagt de gestelde eisen op die ze daar hebben gehanteerd. Vanuit zijn eigen kennis en ervaring en die van ander personen betrokken bij het traject vult hij deze aan (**vraag 1.1**). De geïnterviewde geeft verder aan dat er niet specifiek wordt gekeken naar gevoelige informatie in de documenten. Dit wordt impliciet gedaan op basis van kennis en ervaring (**vraag 1.2**). Dit geldt ook voor privacygevoelige informatie – er is niet een specifieke controle op de documenten die gepubliceerd worden op CTM of ze privacygevoelige informatie bevatten (**vraag 1.3**).*

*Alle documenten die worden geüpload via CTM passeren de inkoper. De inkoper upload namelijk zelf de documenten (**vraag 1.5 en 1.6**).*

Vraag 2: zie alinea 1 onder vraag 1.

Vraag 3 na toelichting guidelines en werking plagiaat scanner: Wat vindt jij van een dergelijke tool om documenten te scannen op gevoelige informatie?

De geïnterviewde geeft aan dat inkopers al gebruik maken van vele checklists en richtlijnen. Een extra checklist is niet waar inkopers op staan wachten. De geïnterviewde zegt dat een geautomatiseerde oplossing, zoals de omschreven plagiaat scanner, wel mogelijkheden biedt. De tool zou een middel zijn om effectiever en efficiënter documenten door te nemen op gevoelige informatie. Nu gebeurt dit impliciet door de personen die betrokken zijn bij het proces.

*De geïnterviewde geeft aan dat de tool in het proces kan worden opgenomen voordat de documenten worden geüpload naar CTM (**vraag 3.1**) – bij de 0.9 versie. De geïnterviewde geeft ook aan dat een inkoper de tool zou kunnen hanteren (**vraag 3.2**). Het resultaat wat wordt teruggegeven door de tool (de documenten met gearceerde tekst dat dus mogelijk relevant is voor kwaadwillenden) zou dan naar de opdrachtgever gestuurd worden. De opdrachtgever is namelijk verantwoordelijk voor de documenten en de inhoud daarvan. De opdrachtgever bepaald vervolgens welke informatie uit de documenten gehaald wordt. De inkoper upload de documenten naar CTM.*

Vraag 4: Wat zijn volgens jouw specifieke informatie typen waarop de tool kan scannen?

De geïnterviewde geeft aan dat het lastig in te schatten is voor hem welke informatie dit moet zijn. Een nice-to-have zou zijn dat de tool niet kijkt naar wat uit documenten gehaald moet worden, maar dat de tool ook kijkt op wat nog mist in de documenten (vraag 4.1 en 4.2).

Vraag 5: Is het mogelijk om aanbestedingsdocumenten te ontvangen die nog niet gepubliceerd zijn om deze te analyseren met behulp van de tool?

Ja. Worden verzonden via de e-mail.

Vraag 6: Heb je nog vragen, commentaar of suggesties voor dit onderzoek?

Nee.

Annex F – The guidelines

The guidelines that are determined in chapter 6 are listed in the table below. This is the definitive list of guidelines.

Table 12: The definitive list of guidelines as a result of the interviews

ID	Guideline	Source	Focus
1	Tenders must be analysed to identify information related to 3rd party infrastructures services. Infrastructure services includes the hardware, software, and network resources required to operate a communications environment. This infrastructure can be managed by a 3rd party rather than being managed by the owning organization."	Problem identification	Reduction of sensitive information
2	Tenders must be analysed to identify information related to domain and IP address space. Domain Names are the human readable names used to represent one or more IP addresses. IP addresses are the unique identifier of computing devices on a network. Both pieces of information are valuable to an adversary who is looking to understand the structure of a network.	Problem identification	Reduction of sensitive information
3	Tenders must be analysed to identify information related to external network trust dependencies. Network trusts enable communications between different networks with specific accesses and permissions. Network trusts could include the implementation of domain trusts or the use of virtual private networks (VPNs).	Problem identification	Reduction of sensitive information
4	Tenders must be analysed to identify information related to firmware version. Firmware is permanent software programmed into the read-only memory of a device. As with other types of software, firmware may be updated over time and have multiple versions. Apart from firmware, versions of other components are also relevant.	Problem identification	Reduction of sensitive information
5	Tenders must be analysed to identify information related to target logon/email address format. Email addresses, logon credentials, and other forms of online identification typically share a common format. This makes guessing other credentials within the same domain easier. For example, if a known email address is first.last@company.com it is likely that others in the company will have an email in the same format.	Problem identification	Reduction of sensitive information
6	Tenders must be analysed to identify information related to externally facing software applications technologies, languages, and dependencies. Software applications will be built using different technologies, languages, and dependencies. This information may reveal vulnerabilities or opportunities to an adversary.	Problem identification	Reduction of sensitive information
7	Tenders must be analysed to identify information related to security defensive capabilities. Security defensive capabilities are designed to stop or limit unauthorized network traffic or other types of accesses.	Problem identification	Reduction of sensitive information
8	Tenders must be analysed to identify information related to web defensive capabilities. An adversary can attempt to	Problem identification	Reduction of sensitive

	identify web defensive services as CloudFlare, IPBan, and Snort. This may be done by passively detecting services, like CloudFlare routing, or actively, such as by purposefully tripping security defences.		information
9	Tenders must be analysed to identify information related to network topology. A network topology is the arrangement of the various elements of a network (e.g., servers, workstations, printers, routers, firewalls, etc.). Mapping a network allows an adversary to understand how the elements are connected or related.	Problem identification	Reduction of sensitive information
10	Tenders must be analysed to identify information related to groups and roles. Personnel internally to a company may belong to a group or maintain a role with electronic specialized access, authorities, or privilege that make them an attractive target for an adversary. One example of this is a system administrator.	Problem identification	Reduction of sensitive information
11	Tenders must be analysed to identify information related to centralization of IT management. Determining if a "corporate" help desk exists, the degree of access and control it has, and whether there are "edge" units that may have different support processes and standards.	Problem identification	Reduction of sensitive information
12	Tenders must be analysed to identify information related to physical locations. Physical locality information may be used by an adversary to shape social engineering attempts (language, culture, events, weather, etc.) or to plan for physical actions such as dumpster diving or attempting to access a facility.	Problem identification	Reduction of sensitive information
13	Tenders must be analysed to identify information related to business processes/tempo. Understanding an organizations business processes and tempo may allow an adversary to more effectively craft social engineering attempts or to better hide technical actions, such as those that generate network traffic.	Problem identification	Reduction of sensitive information
14	Before the analysis starts where information related to the above guidelines is identified, it is important that new information types can be added. This must be done automatically, for example by pulling new information types from the MITRE website. The specific information types that hackers look for may change in the future. It is possible that with a new cyber attack it appears that hackers have used a different information type when preparing for the attack.	Problem identification	Embedding in the contextual Practicality
15	After analysing the information in tenders, tenders must still be usable. Information that is identified and removed by the client in the tender process must not result in a tender that does not comply with the applicable laws and regulations.	Tender context	Embedding in the context
16	After analysing the information in tenders, tenders must still be usable. Information that is identified and removed by the client in the tender process must not result in a tender that is no longer usable in the sense that the tender does not provide sufficient information to contractors who want to register for the assignment.	Tender context	Embedding in the context
17	The analysis where the tenders are examined to identify	Tender	Embedding

	sensitive information must be embedded in the tendering process prior to the step where the documents are made public. The analysis must be done on 0.9 versions. No new information can be added after the analysis. After performing the analysis based on the information types and processing the results, the tenders must be published.	context	in the context
18	The analysis that is performed where tenders are examined must be executed by the purchaser involved in the tendering process. The analysis therefore must be easy to do. Nor should it be expected that the user has explicit security knowledge.	Tender context	Embedding in the context
19	Prior to the analysis on the tender information the purchaser must provide the documents as input by simply selecting or dragging and dropping them.	Similar artefacts	Practicality
20	After the analysis, the results are clearly displayed to the purchaser. In this way the purchaser can see in an overview how much sensitive information has been identified.	Similar artefacts	Practicality
21	A database is kept with identified results. The objects that the database contains are included in the analysis. In this way, analysis is not only based on the information types, but an analysis can also be carried out whereby a comparison is made between the information in the tenders and the objects in the database – the previously identified information in other tender documents.	Similar artefacts	Practicality
22	The purchasers have clearly indicated that the analysis must take place automatically. It is not desirable to use a manual checklist that purchasers must use.	Interviews	Embedding in the context
23	When sensitive information related to the information type that the analysis is based on is found it should be made clear why the information is identified as sensitive. The purchaser may not know why this information might be relevant for adversaries.	Interviews	Embedding in the context
24	The purchasers involved can perform the automated analysis to identify information. As a result of guideline 20 – the overview with the result of the analysis must be saved in a visual format.	Interviews	Embedding in the context
25	The purchasers have clearly indicated that they want to perform the automated analysis. As a result of guideline 23, the results must be sent to the client in the tendering process. The client must then decide which information must be removed.	Interviews	Embedding in the context
26	Tenders should be analysed to identify information related to privacy sensitive information. Information about persons is sensitive to privacy and, therefore, may not be included in tender documents that are made public according to legislation and regulations (General Data Protection Regulation). In addition, malicious parties can use such information for social engineering activities.	Interviews	Reduction of sensitive information