CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG

A Methodology to Validate Compliance to the GDPR

Master's thesis in Software Engineering

Axel Ekdahl
Lídia Nyman

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2018

Master's thesis 2018

A Methodology to Validate Compliance to the GDPR

AXEL EKDAHL
LÍDIA NYMAN

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2018

A Methodology to Validate Compliance to the GDPR
AXEL EKDAHL
LÍDIA NYMAN

© AXEL EKDAHL, LÍDIA NYMAN, 2018.

Supervisor: Riccardo Scandariato, Computer Science and Engineering.
Examiner: Jan-Philipp Steghöfer, Computer Science and Engineering.

Master's Thesis 2018
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by [Chalmers University of Technology and University of Gothenburg]
Gothenburg, Sweden 2018

A Methodology to Validate Compliance to the GDPR
AXEL EKDAHL
LÍDIA NYMAN
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

This study analyses two state-of-the-art methodologies for eliciting privacy threats in software contexts, LINDDUN and PIA. A first goal is to understand the limitations of these methodologies in terms of compliance to the provisions of the robust General Data Protection Regulation (GDPR). A second goal is to improve the first methodology by addressing its limitations and proving a more complete coverage with regards to the regulation. The study is divided into two phases; an analysis of the current coverage of the two methodologies and the development of an extended version of LINDDUN. The extended LINDDUN includes a privacy-aware Data Flow Diagram and extensions of the Content Unawareness and Policy and Non-compliance threat trees, as well as developed rules for defining where in a software design a privacy threat commonly exists. It was observed that PIA was considered more effective than LINDDUN in identifying design issues related to GDPR. While the extended version of LINDDUN showed to provide a more complete coverage than the original LINDDUN.

Keywords: Privacy, Privacy Threat Modeling, GDPR, LINDDUN, PIA, GDPR compliance, Privacy Impact Assessment

Acknowledgements

We would like to thank our academic supervisor Riccardo Scandariato at Chalmers University of Technology for his expertise, valuable insight and guidance along this study. Additionally, a prodigious thank you to our industry supervisors Christian Sandberg and Atul Yadav for providing us with information, operational arrangements and other related areas needed to conduct our work at Volvo Group Trucks Technology. Also a grateful thanks to the participants in this study, both the engineers at Volvo Group Trucks Technology and the students from Chalmers University of Technology and University of Gothenburg.

Axel Ekdahl and Lídia Nyman, Gothenburg, June 2018

Contents

List of Figures

List of Tables

1

Introduction

Technology has become an essential part of today's society. More and more individuals are dependent on software systems and tools such as websites, mobiles and computer applications, or any other system within a computer environment, to perform day-to-day tasks. Each time an individual makes use of any of these applications or information systems, commonly, they are required to provide some type of personal information, such as name, email address, or other information that can be used to identify that specific individual in a digital form. By providing too much personal information an individual can become vulnerable to attacks that can violate the individual's privacy [3].

Privacy can be defined as the right of the individual to decide which kind of information regarding him or her is going to be revealed [4]. Thus, thinking about users' privacy and to ensure a higher level of privacy within software systems the European Union (EU) has introduced new provisions into the General Data Protection Regulation (GDPR) [5]. The primary focus of the regulation is to enforce the rights of individuals to privacy, as well to ensure that organizations are being compliant with the proposed privacy principles and user rights. A suggestion to this is that organizations start addressing privacy in the early stages of the software life-cycle, a practice known as Privacy by Design (PbD) [6].

Prior the regulation, that has become into place since May 2018, organizations were empowered over the information representing other individuals. The GDPR harness organizations that collects information about individuals and instead empower the individual to be able to make a choice over his or her data. In recent GDPR lawsuits made against two of the world's largest software companies, Google and Facebook, the fines for not providing compliance to the regulation resulted in a combined fine of 8.8 billion dollars [7]. These high fees show to organizations the importance of finding ways to become compliant with the GDPR.

Currently, there are different state-of-the-art threat analysis methodologies used to address privacy concerns within software systems [8] [9] [10]. This study will provide an overview of the most common methodologies to address PbD and investigate its effectiveness in identifying design issues related to GDPR compliance. The goal of this study is to provide a methodology to validate compliance with GDPR provisions in order to help organizations to achieve PbD.

Due to limited resources only two methodologies were included in this study, which are LINDDUN and PIA. The choice of these two methods is out of convenience and also because they are well regarded in the privacy community. The LINDDUN methodology is a *"threat modeling technique that encourages analysts to consider privacy issues in a systematic fashion"* [1]. Further, LINDDUN elicits privacy from the perspective of a given software's architectural design with a provided and structured step-by-step strategy. Privacy Impact Assessment (PIA) is another methodology for finding privacy threats in software systems. Further, privacy regulations such as the GDPR specifically advocates to use PIA to reach sufficient privacy enforcement in order to become compliant with the regulations [5].

With GDPR organizations are now obligated to review its processes and how private information is retrieved and used. The goal of this study is to choose two of the well regarded methodologies, presented above, and analyze their possible gaps with relation to compliance with GDPR. Furthermore, this study aims to address these issues and propose an extended methodology that can be used by organizations to analyze design issues related to compliance with the regulations, and thus provide privacy by design to its users.

This study was performed in collaboration with Volvo Group Trucks Technology (Volvo GTT) under the HoliSec (Holistic Approach to Improve Data Security) project [11]. A project that aims to address data security in development processes within the automotive domain. The purpose of the project is to develop a bank of security solutions which prohibit data security problems from violating vehicle safety. Until today, the degree of privacy-related work in the HoliSec project is limited and the outcome of the work of this study is considered an entry point for introducing privacy in the project.

1.1 Research Questions

The research questions defined for this study are:

- **RQ1:** *How effective are state-of-the-art threat analysis techniques like LINDDUN and PIA in identifying design issues related to GDPR compliance?*

- **RQ2:** *Does an extended version of LINDDUN provide a more complete coverage of said issues?*

1.2 Scientific Contribution

This study propose to analyze two well regarded methodologies in the privacy community for eliciting privacy threats. Specifically, to contribute with knowledge of current coverage of compliance to a robust data protection regulation of two methodologies, LINDDUN and PIA, as well as presenting extensions to one of the privacy threat analysis methodologies. The methodologies are used to identify and elicit privacy threats during the early stages of the software development life-cycle. The proposed extensions can be used to help practitioners to ensure compliance to privacy regulations when developing software systems, and thus achieve PbD The extended work for the methodology aims to be used not only by experienced privacy practitioners but also for non-experienced privacy practitioners. The proposed changes to the methodology address the new GDPR guidelines, in place since May 2018, and together provide a step-by-step approach to ensure PbD. The system domain used to evaluate the extended methodology also adds significance to this study due to its context, since the methodology had not been previously applied within the automotive domain. Furthermore, the results from the evaluation of this study provide evidence of the effectiveness of the extended framework in identifying privacy threats with a correlation to the GDPR provisions.

1.3 Thesis Outline

The report is structured as follows. Chapter 2 explains the research methodology used under this study. The Chapter three introduces privacy, privacy-related activities, and provides an overview of the terms related to privacy and used during the development of this study; i.e. privacy terminologies, threats to privacy, privacy properties, privacy regulations, privacy solutions, privacy threat analysis methodologies, and privacy in the automotive domain. Chapter 4 presents the work performed in the Iteration 1 of this work, where a comparison of two selected methodologies was performed in terms of compliance with the provisions stated by GDPR. An introduction of the case study used to evaluate the methodologies is also included in Chapter 4, together with the results of the evaluation. Chapter 5 presents the proposed extensions of LINDDUN resulting in a new version named LINDDUN+. Chapter 6 is used to validate the LINDDUN+. It describes how the evaluation was performed and the obtained results. Additionally, a discussion of the results, threats to validity and other factors that might have influenced this study are also presented in Chapter 7. Finally, Chapter 8 provides insights for future work and summarizes the findings of this study.

2

Research Methodology

The research methodology used in this study is based on the design research methodology. It contains six stages, where four stages forms an iterative cycle. From an overview, the methodology addresses a literature review; a comparison of LINDDUN and PIA; a proposal for an extended version of LINDDUN, the LINDDUN+; and an evaluation of the proposed extended version. An overview of the methodology can be seen in Figure 2.1.

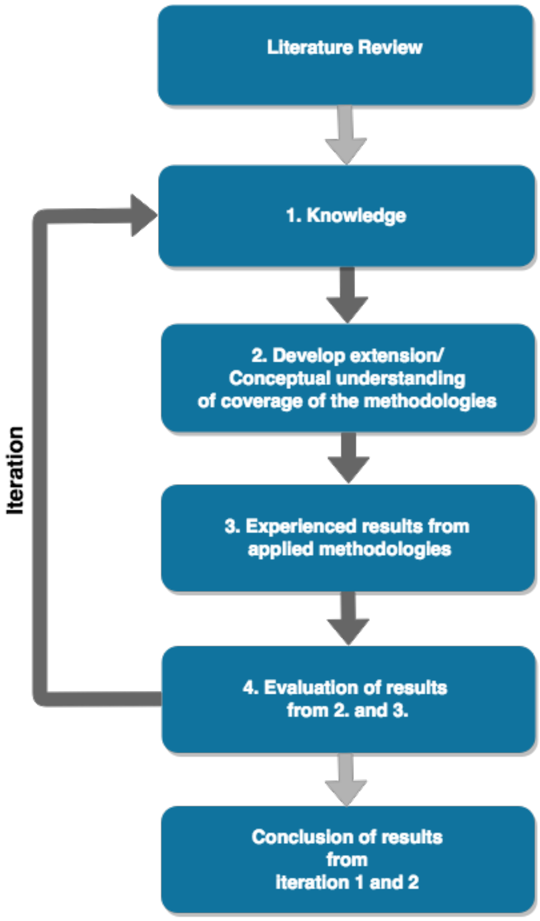


Figure 2.1: General View of the Methodology of the Study.

2.1 Literature Review

As the aim of this study is to analyze the performance of two state-of-the-art methodologies for eliciting compliance to privacy regulations, such as the GDPR, in a software context, it is necessary to gain an understanding of crucial aspects that relates to this context. Hence, in order to successfully perform such analysis knowledge of correlated areas that affects privacy was needed to be gained. For this reason, the first stage of this work consisted of a literature review used to understand general concepts of privacy, focusing on threat analysis methodologies and also compliance to the regulation GDPR. The literature review thus has a connection to the **RQ1:** *How effective are state-of-the-art threat analysis techniques like LINDDUN and PIA in identifying design issues related to GDPR compliance?*, since the need to understand the GDPR (what are the requirements how does one comply with them) in order to be able to conduct a coverage-analysis of the two methodologies. The literature review has also a connection to the **RQ2:** *Does an extended version of LINDDUN provide a more complete coverage of said issues?*. For the same reason as for the first research question, an understanding of the provisions of the GDPR was needed to be gained in order to be able to conduct the coverage-analysis of the extended methodology. Also, the found work in the literature review also helped from the aspect of providing insights on what has an impact of privacy. Without such knowledge gained, and in combination with the knowledge of the GDPR provisions, an extension of the LINDDUN+ would not be possible to be conducted. The result of the literature review can be seen in the related work in Chapter 3.

2.2 Design of the study

As can be seen from the Figure 2.1 the core of this study concerns for steps forming an iterative design; knowledge, conceptual understanding of the methodologies, developed extensions and empirical sessions of the methodologies as well as an evaluation of the generated results of the iteration. Two iterations has been conducted in this study. This section provides a description of what each iteration covered with an explanations of the performed work in each iterative step.

2.2.1 Iteration 1: LINDDUN versus PIA

The first iteration of this work consisted of an analysis of two state-of-the-art threat analysis techniques, LINDDUN and PIA. The analysis was used to identify gaps with relation to compliance to regulations, and work was divided in four parts, as presented below.

2.2.1.1 Part 1 - Knowledge from Literature Review

With the extensive literature review conducted a rich knowledge base was gained. However, the authors experienced difficulties in the interpretation of the GDPR. This due to the ambiguity and non-precision of the provisions of the regulation. For this reason semi-structured interviews with two privacy and security engineers at Volvo

Group Trucks Technology was held. The interviews served the purpose to clarify the ambiguities of the regulation through having a set of questions. The questions represented the authors interpretation of the provisions and an discussion was formed around these questions. Further, the interviews was held at two different occasions, where the interviews took place with the engineers individually. They addressed privacy concerns provisioned of the GDPR both from a general perspective but also how Volvo Group Trucks Technology (as a representative from the automotive domain) do act upon the provisions. After the interviews useful insight of how to interpret the principles and user rights of the GDPR as well as the implications that follows was gained. To conclude, after the conducted extensive literature review in combination with the the knowledge gained from the semi-structured interviews sufficient knowledge had been gained to start the exploratory work of understanding the coverage of the two methodologies in this study, in terms of compliance to the GDPR. This is further explained in the following sections.

2.2.1.2 Part 2 - LINDDUN and PIA comparison

After the literature review and the semi-structured interviews were conducted work was initiated for making of a conceptual effort of mapping the two methodologies with the provisions stated by the GDPR regulation. By going through the published material regarding the two methodologies and comparing this to the principles and user rights, a conceptualized understanding was developed regarding how well the two methodologies do sufficiently eliciting design issues related to the provisions of the GDPR. This attempt resulted in the Table 4.1 provided in Chapter 4 and was a reason for further investigation of the effectiveness of methodologies in identifying design issues related to GDPR compliance. As was explained in the previous section, some of the principles and user rights can be considered ambiguous and vague. An attempt to do a quantitative coverage-analysis some principles and user right were needed to be more concrete. For this reason, some provisions were divided into smaller and more concrete provisions. For those provisions that were assigned sub-principles, as seen in Table 4.1, all sub-principles of that principle were needed to be covered in order for the methodology to be considered to cover the principle. The conceptual mapping was done without any external stakeholders, thus the engineers participating in the previously explained semi-structured interviews were not involved in this process.

2.2.1.3 Part 3 - Empirical Evaluation of LINDDUN and PIA

From the conceptualized mapping explained in the previous section, an initial understanding of how well the methodologies perform in terms of GDPR-compliance. Hence, the authors gained an expectation regarding to what extent a practitioner would be able to find privacy threats that relates to the provisions of the regulation. However, this expectation might not represent all scenarios and all stakeholders where external factors can affect the performance of the elicitation. For this reason, the third part of the iteration consisted of an empirical evaluation of the methodologies through a case study provided by Volvo Group Trucks Technology. This evaluation was conducted for further comparison in order to identify how well the methodologies identify design issues related to the provisions stated in the GDPR, but from empirical results. The two authors were assigned to one methodology each and did apply them individually and isolated, without communicating or interacting with each other or any other external interference. Both sessions took one workday to be conducted. For LINDDUN only the three first steps was performed since these are the only steps of the methodology that focuses on identifying and documenting privacy threats, thus the scope of this study. Regarding PIA, the guideline developed by Oetzel et. al [10] was used. This guideline is applicable since it is developed for ensuring compliance to regulations similar to GDPR for systems in the automotive domain.

2.2.1.4 Part 4 - Evaluation of Results from 2 and 3

Ones the conceptual mapping in the second part and the empirical evaluation in the third part of the first iteration had been conducted, results were assembled. The results from the two parts, were compared. Hence, an evaluation was made for analyzing if the empirical results from part three did align with the estimated results from part two. This comparison can be found in section 4.3.

2.2.2 Iteration 2: LINDDUN+

The second iteration consists of work for developing an extended version of LINDDUN. This for the purpose to analyze if the extended version of the LINDDUN methodology covers the provisions of the GDPR to a higher extent compared to the current version of the methodology.

2.2.2.1 Part 1 - Knowledge from Iteration 1

The work in the first part of the second iteration focused at making a proposal for an extended version of the LINDDUN methodology, named LINDDUN+. Knowledge was gathered from the part 4 of the first iteration, to address the gaps observed of LINDDUN with regards to compliance to the GDPR. Specifically, it was clear that LINDDUN did posses improvement potential. This since it only covered 2 principles and user rights from the GDPR, implying a coverage of 14 percent. For instance, the right to object or the right to erasure was not addressed at all by the methodology. The PIA did cover 11 principles and user rights, making PIA covering 71 percent of the GDPR provisions. Since LINDDUN only covered 14 percent of the GDPR provisions a set of extensions are proposed in this study. The extensions are; introduction of a privacy-aware DFD (PA-DFD), extension of the two threat trees Content Unawareness and Policy and Consent Non-Compliance as well as development of threat tree rules.

Since it was clear that LINDDUN suffered improvement potential work started for finding literature that address privacy from a design modeling perspective. The intention was to see if work could be found that would help LINDDUN filling the gaps LINDDUN suffered. Literature was found that was considered being helpful, namely the work proposed by Antignac et al. [12] [13]. The purpose behind the PA-DFD was to make the practitioner in a more extensive way being able to understand sensitive parts of the analyzed software system. This by extending the the first step of LINDDUN, development of DFD, to include a privacy-aware DFD instead of a business-oriented DFD. With the PA-DFD the practitioner get a more explicit view over how data flows and in this way he or she can easier understand sensitive parts of the software system in terms of privacy. Further, from the literature review, conducted prior the first iteration, the importance of understanding what data, used by the software system, can be classified as Personally Identifiable Information (PII) was learned. Also, from the semi-structured interviews in the first iteration it was clear that the phenomenon of Data Classification could be of use for analyzing data in a privacy context. For this reason, in LINDDUN+, the practitioner is advocated to use a data classification specification during the first step of the methodology. It defines what data, in the software system analyzed, is evaluated as PII.

Besides the proposal of the PA-DFD another improvement potential of LINDDUN was two of its threat trees. After thoroughly reading the documentation of LINDDUN, and thus a gained knowledge of the threat tree catalogue, it was clear that the threat trees of LINDDUN, which has should a connection to the provisions to GDPR, did not address such provisions sufficiently. Also, the threat trees of LINDDUN aim at addressing the most common attack paths to a software system, in terms of privacy. Hence, the relevance for extending the two threat trees Policy and Consent Non-Compliance and Content Unawareness.

After the empirical evaluation of LINDDUN in the first iteration a belief by the practitioner was that the LINDDUN methodology was hard to learn and also time consuming. Especially for a novice practitioner in both privacy and the LINDDUN methodology. A belief, of the authors, was that with an automated tool for the LINDDUN methodology the future practitioner would more easier and faster familiarize himself/herself with the methodology. Thus, affecting the practitioners ability, in a positive direction, to eliciting privacy threats in a software design context. For this reason, a proposal for taking an initial step towards an automated LINDDUN was made. This initial steps would consist of a set of threat tree rules, specifying where, in a software system, a threat from a threat tree node most commonly arises.

2.2.2.2 Part 2 - LINDDUN+ mapping to the GDPR

When the proposal for the improvements of LINDDUN were defined, work for developing such improvements was initiated. All the developed work were done by the authors of this study without any external involvement. First, an assessment of how a PA-DFD could be included in LINDDUN was made. By having the work proposed by Antignac et al. [12] [13] serving as inspiration, a meta model was created. This since the inspiration taken from their work was the more specific DFD elements proposed. In order for a practitioner to be able to use such elements in a future LINDDUN it was clear that a meta-model that explains how the new DFD elements could be used and applied was needed. When the meta model was created work focused instead to understand how the LINDDUN methodology could improve its potential for its practitioner to have a better understanding of what privacy sensitive data is at what parts of the applied software system. As was known from the semi-structured interviews in the first interview an idea was to use a data classification.

Once the scope of the PA-DFD was defined work for extending the two threat trees, Content Unawareness and Policy and Consent Non-Compliance, started. By thoroughly analyzing the details of the provisions from the GDPR, more threat tree nodes were added. Hence, work focused at understanding, from a design modelling perspective, where in a software system could violations to the provisions occur.

Ones the extensions of the threat trees were accomplished, work for defining between which DFD elements a certain threat commonly exists. This was done by writing threat tree rules. Hence, the intention was accompanying each threat tree node with a rule. This for the purpose of taking an initial step towards automation of LINDDUN and hence a believed ease of use of the methodology.
When the extensions had been developed the same type of conceptualized mapping, as was done in the part two of the preceding iteration, was performed. This so an understanding and an expectation could be gained of how well the extended work of LINDDUN would perform, in terms of coverage of the provisions of the GDPR. The results from the conceptualized mapping can be seen in Table 6.1 in section 6.2.

2.2.2.3 Part 3 - Empirical Evaluation of LINDDUN+

When the extensions, which forming LINDDUN+, was developed an empirical evaluation of LINDDUN+ was made, similar to the one conducted in part 3 in the first iteration. The LINDDUN+ was applied on the same case provided by Volvo Group Trucks Technology as was used to evaluate LINDDUN and PIA. The proposed extensions developed in this study are defined in section 5.2.

Pilot Workshop

The pilot workshop was used to validate the architectural view extracted from the system documentation provided by Volvo and to validate LINDDUN+. It was performed with a cyber-security specialist at Volvo Group Truck Technology. The workshop was divided in two parts and was performed in two different days. The first part consisted of a brief introduction of the LINDDUN, LINDDUN+ and the GDPR principles and user rights. The introduction session lasted approximately an hour and at the end a material guide was handed to the participant. In this way, the participant would have some time to assimilate the concept and be familiar with the methodology before performing the second part of the pilot workshop.

The second part of the pilot workshop was performed two days after the introduction session and consisted of a brief recap of the methodology and the validation of the developed DFD. In total the second part of the workshop lasted four hours. After the recap, the system under analysis was presented to the participant. In the second session, a system description extracted from the documentation provided by

Volvo Group Trucks Technology, was handed to the participant. This documentation consisted of a system description containing four use cases and the an already developed DFD. The DFD was developed by the authors of this study.

Workshop with Students

The workshop with students was conducted after the pilot study and was used to validate LINDDUN+. It was divided in two parts as the pilot study and was performed with three software engineering master's students in their final year of from both Chalmers University of Technology and the University of Gothenburg. The participants were selected according to a criterion considered of relevance to the authors of this study. In order to participate in the workshop the students had to have successfully completed the Advanced Software Architecture course provided by both universities. During the course the students become familiar with different architectural views of a system and are given the opportunity to apply STRIDE, as similar methodology used to identify security threats in software systems, developed by Microsoft.

The workshop consisted of an introduction session of an hour, where LINDDUN, LINDDUN+ and the GDPR provisions were presented to the participants. The same material guides handed to the participant of the pilot workshop were given to the students. The actual workshop was also performed a day after the introduction session and also included a brief recap of the methodology and the presentation of the system used for the analysis.

2.2.2.4 Part 4 - Evaluation of Results from 2 and 3

The final part of this iteration was used to assemble the generated results from the conceptual mapping performed in the second part of this iteration and the results gathered from the pilot workshop and the workshop with the students. These results were analyzed and used to draw conclusions of the effectiveness and performance of the proposed extensions in relation to the GDPR provisions. All the results are presented in Chapter 6.

2.3 Conclusion

After have conducted two iterations, results were generated providing evidence to answer the two research questions of this study. Hence, conclusions are made for answering the coverage of the two methodologies, LINDDUN and PIA, regarding the GDPR provisions. Also, possibility to answer the question if an extended version of LINDDUN provide a higher coverage to the same regulation compared to the original version exists. The conclusion was used to provide an overview of the purpose of the study and to provide suggestions for possible future work, presented in Chapter 8.

3

Understanding Privacy

3.1 Introduction to Privacy

To conceptualize and explain a generic definition to privacy can be very hard, and the significance of the meaning of privacy can vary between different contexts [14]. One of the most referred definitions for privacy is given by Warren and Brandeis from 1890 [15]. They describe privacy as *"the right to be left alone"*. Clearly, privacy here is not referred to the context of information technology but instead of a much more analogue and social context. The authors discuss violation to privacy as a form of defamation and exposure. More specifically, when public readers (citizens) of a newspaper have access to personal information disclosed by journalists, through articles containing slander news regarding these citizens.

Another way to think of privacy is through *"the right to select what personal information about me is known to what people"* [4]. This definition fits more into the context of today's information technology environments, where data often refers to personal information in the form of a digital nature. In this specific context, privacy refers to which personal information a person is willing to share and provide to organizations when using their products or services.

The remaining parts of this section are used to present different sensitive scenarios regarding privacy.

3.1.1 Privacy-related Activities

Solove [16] developed a taxonomy for privacy where he thoroughly explains common activities that infringe personal privacy. The taxonomy was developed by investigating social violations to privacy and comparing it with the American law's perspective. His viewpoint has become well established and accepted in literature. Thus, this taxonomy of privacy will be used in this work to explain key aspects and concerns regarding privacy. The taxonomy is divided into four activity groups; information collection, information processing, information dissemination and invasions. Each group consists of subgroups that aims to identify and understand privacy violations from different perspectives and scenarios. The different activity groups are explained below and were all extracted from the work created by Solove.

3.1.1.1 Information Collection

Not all collections of information have to be considered harmful. However, data must be collected before it can be considered to possess any harm. Once, this collected data can be misused and disseminated. The two activities that represent how information is collected, and thus can create a threat during information collection, are explained bellow.

- *Surveillance:* means any intruder or unauthorized party continuously listens or observes another individual to get access to any desired information. Such intrusion can happen either with or without the individual's consent. The desired outcome from surveillance is the increased control over the subject. The harm can escalate, if surveillance being abused, and impact the subject's freedom and creativity.

- *Interrogation:* refers to the act of pressuring individuals to reveal information that others want to know. Since the interrogator possess the control over what information to obtain and can interpret and form personal impressions from the obtained information, wrong assumptions and conclusions can be drawn. When these wrong assumptions are spread it can result in a harm to the individual.

3.1.1.2 Information Processing

When the data that has been collected is used, stored, or manipulated in any form, this is categorized as processing of information. The five activities that characterize information processing are explained below:

- *Aggregation:* occurs when small sections of information are assembled together and stored in a distinct place. By placing together small parts of information from an individual, more accurate inferences can be drawn. The inference accuracy is possible due to the increased information value provided by the combination of information. This can create a harm when additional information from an individual is revealed. Moreover, aggregation can violate privacy when it significantly inflates others knowledge regarding an individual, even if such knowledge is derived from public sources.

- *Identification:* is the association of information and individuals, and it has a central role in the scope of privacy. One benefit provided by identification is the possibility to verify an individual's identity or whom an individual claims to be. Identification relates to disclosure by means of revealing true information about an individual, while identification specifically reveals true information regarding ones identity. Additionally, it also relates to aggregation since both involves combinations of pieces of information, where one piece is the identity of the individual. The identification's difference to aggregation is the link to the individuals physical space which identification uncover.

- *Insecurity:* concerns the problematic regarding how information is processed and kept protected. It refers to the damage created for a weakened state, or the possibility of being prone to a range of future harms. Insecurity relates to aggregation in scenarios when insufficient protection mechanisms of the processed personal data are applied. The implications of such risk could be an adversary that possess the ability to access personal data due to insufficient protection mechanisms. Insecurity is further related to identification. Previously, it has been explained that the privacy concerns regarding identification is when information, which represents an individual, is disclosed. The relation between insecurity and identification on the contrary comes down to the inability to sufficiently and correctly identify an individual. A common scenario is during identity theft, where the adversary can take over a stricken individual's personal data. Such over taken data can later be used by the adversary to perform actions where he/she claims to be the individual which the stolen data represents.

- *Secondary Use:* refers to data that is used for a purpose that is contrary to its initial purpose. Additionally, it can include data that is used without the data subject's consent. The initial purpose is defined as the reason why the data is collected in the first place. While the secondary use can also be defined as the usage of the collected information in contexts where the data subject does not consent or desire. Secondary use is considered to be similar to breach of confidentiality, due the betrayal of the data subject's expectations when submitting its data to the data receiver.

- *Exclusion:* is the failure to provide data subjects with notice and input regarding their information. Implications of exclusion include the reduced accountability provided by responsible parties that maintain individual's information. Commonly, exclusion is not a harm that is present due to the lack of protection mechanisms from data leakage or contamination. Instead, its presence rely on the data subject's unawareness of how the data is used. Consequently, not allowing the individual to make decisions regarding the usage of its data.

3.1.1.3 Information Dissemination

The threat categories included in this grouping consist of ways to spread or reveal personal data. The seven harms discussed in this category are described below:

- *Disclosure:* refers to the damage caused to an individual's reputation due to the dissemination of certain true information regarding this individual to others. This can be considered to be a harm even if the information was disclosed by a stranger. The fear of disclosure can inhibit people from interacting with others, from engaging in activities that can improve their self-development, and from express themselves freely. Moreover, disclosure can be a threat for a person's security and make a person a *"prisoner of [her] recorded past"*.

- *Breach of Confidentiality:* can, just as disclosure, cause harm to an individual when secrets regarding this individual are leaked. The difference is that breach of confidentiality infringes trust in a specific relationship. In other words, the focus is not in the information that has been but instead the betrayal suffered by the individual.

- *Exposure:* involves the act of exposing information to others related to a person's physical and emotional state. Although exposure can be similar to disclosure, the difference relies on the fact that the information disclosed does not affect our judgment of a person's character or personality. Exposure, however, can also affect a person's ability to participate in society by striping a person's dignity.

- *Increased Accessibility:* is when personal information that is already available to the public is made easier to be accessed. Increased accessibility can increase the risk of harms to disclosure by allowing the available information to be exploited for malicious purposes.

- *Blackmail:* is the action of coercing an individual by threatening to expose his or her personal secrets if he or she does not comply to the demands of the blackmailer, often involving the payment of hush money. By prohibiting the payment to a blackmailer the treat for disclosure increases. However, the harm caused by blackmail is not the disclosure of information, but the power and control the blackmailer exercises over the data subject. Moreover, blackmail can also be related to exposure and breach of confidentiality.

- *Appropriation:* is the use of an individual's identity or personality in order to fulfill the purposes or goals of another individual. Appropriation can be related to disclosure and distortion, causing privacy disruptions and involving the way an individual wishes to present himself or herself to the society.

- *Distortion:* is the inaccurately exposure of a person to the public and involves the manipulation of the way an individual is perceived and judged by others. It not only affects the offended individual but also the way society judges this individual. Distortion is similar to disclosure, since both affect the way an individual is seen by the society. However, with distortion the information is false and deceptive.

3.1.1.4 Invasion

Invasion harms are different from information collection, processing and dissemination threats because they do not always involve information. They are divided in two types as described below.

- *Intrusion:* is the invasion or incursions into a individual's life, disturbing the individual's daily activities and making the individual feel uncomfortable and uneasy. Intrusion can be related to disclosure, once the disclosure of information is made possible by intrusive activities in one's life. With the same intrusive activities, it can also be related to surveillance and interrogation.

- *Decisional Interference:* is when the government interferes with individual's decisions regarding certain aspects of the individual's personal life. This can include decisions related to sex, sexuality and upbringing of children. Decisional interference can be related to insecurity, secondary use, and exclusion, since all the threats related to these three can affect the individual's decisions when it comes to his or her health and body.

3.2 Privacy Terminology

This section describes the common terms used when referring to data privacy in software systems. The terms include the definition of personal data, personal identifiable information, data subject, data controller, data processor and the privacy related operations.

3.2.1 Personal Data and Personally Identifiable Information

Data can be seen and structured in various different ways. Commonly, in privacy literature two types of data recur in the discussions namely *personal data* and *sensitive data*. The GDPR [17] defines personal data as any type of information that can be used to identify a person directly or indirectly. This includes different types of identifiers such as name, identification number, location data, online identifier and etc. Sensitive personal data is defined by the Article 9 of the GDPR as *"special categories of personal data"*. This category includes sensitive data such as genetic and biometric data which can be used to directly identify an individual.

Personal Identifiable Information (PII) is a central factor in privacy. It is a classification of sensitive and unique information that makes an individual's identity possess the risk of being distinguished or traceable and which further is linkable to the individual itself [18]. In other words, PII is a set of information properties that alone or together can make an individual becoming uniquely identified. Explicitly, properties that can be considered as PII are; full name, alias, social security number, driver´s license number, street address, email address or personal characteristics such as handwriting or photographic image of an individual [19]. However, these are only a generic exemplification of PII properties where more exists depending on the context. The concept of data theft is a big reason to the concern of PII and hence privacy. This is also one of the reasons to the new upcoming regulations of GDPR coming into place in May 2018 [5].

The health-care is a domain where privacy has been a concern for many years. Specifically, the privacy of the patients. The Health Insurance Portability and Accountability Act (HIPAA) [20] states a privacy rule which has been established as a guideline to be followed. The privacy rule contains a methodology called "Safe Harbor" which defines eighteen privacy attributes which all should be considered as personally identifiable information. The attributes are shown in table 3.1. Therefore, the importance of protecting an individual's PII from the tremendous volume of available information that cannot be classified as personal. In today's cyber-environments, computer engineers can even turn non-PII it into PII [21]. This can be related to the incident of the America Online's (AOL) disclosure from 2006.

In 2006, AOL assembled twenty million search queries made by various individuals and determined to make them public. Individuals were supposed to be anonymous by being given a pseudonym instead of their real identity. The authors to the article in New York Times however demonstrated how it was possible to re-identify individuals by mapping search queries performed by the same pseudonymous identity and hence for instance understand life patterns by the individuals [22].

Today re-identification mechanisms exists that can in a relatively easy way re-identify individuals from analyzing disclosed Non-PII. This even though privacy protection technologies is present in a system. However, alternatives still persist which can in a structured way enforce privacy. Instead of applying technologies which serves re-actively to an attack a better solution is to apply pro-active alternatives. Such alternatives would include differential privacy (explained in section 3.6.3.5) data minimization, data access control and user consent compliance [23].

Table 3.1: Table explaining privacy attributes defined in HIPAA Privacy Rule

Names	Street address, city, county, precinct, ZIP code
Birth date, admission date, discharge date, death date	Telephone numbers
Fax numbers	Email addresses
Social security numbers	Medical record numbers
Health plan beneficiary numbers	Account numbers
Certificate/license numbers	Vehicle identifiers and serial numbers, including license plate numbers
Device identifiers and serial numbers	Web Universal Resource Locators (URLs)
Internet Protocol (IP) addresses	Biometric identifiers, including finger and voice prints
Full-face photographs and any comparable images	Any other unique identifying number, characteristic, or code except dates

3.2.2 Data Subject, Data Controller and Data Processor

As previously explained, privacy tends to relate to how information regarding individuals is processed, stored, shared, or in other means, used by a software system. In order for software engineers, architects, and other parties with a relation to a privacy law context, to counter privacy threats at such activities it is important to look at the information flows in a software system from different view points. Another important aspect to look from is who are the affected individuals upon a potential privacy attack as well as who are the responsible parties to counter such attacks. This section explains three keywords which are frequently used in privacy contexts of software systems to address just this. [24].

Data Subject

In privacy it is important to understand who is the owner of the data a system store, process, share or forward to external parties. Such owner is the physical person who the information represents or in any way is related to. Often this information is provided by the individual itself by input, while using the software system. This owner is commonly referred to as the data subject. In scenarios where privacy threats being discussed or elicited the central concern is to enforce the privacy mechanisms of the data related to the individuals of the system, the data subject. [24].

Data Controller

A data controller is one (individual, organization or corporate person) who is responsible for determining the purposes of which personal data are currently, or in the future, processed. Commonly, the data controller refers to an organization, with exemption to small companies (for instance where the personnel being self-employed). In other words, the data controller holds the highest responsibility when personal information, that they are responsible for, is being processed. Further, the data controller overall determines the "why" and "how" of processing of data subjects information. The responsibility of the data controller extends to ensuring all data processing complies with current legislations and regulations. If a system fails to comply with such laws this falls on the controller's responsibility. [17]. Finally, the data controller determines the purpose for a processing, what personal data to store, which individuals to store information about and how long the retention time for the stored data should be. [24]

Data Processor

Just as the data controller, the data processor is any entity that processes data. The difference between them two is the latter being one (individual, organization, or corporate person) other than an employee of the controller itself who perform processing activities. Hence, a data processor is not the highest responsible for the data. Instead, the data processor acts on behalf of the controller. [17].

Data processors may still be considered as controllers in one way. This since, data is processed and hence the responsibility of the actual processing shall be performed in accordance to the law. Specifically, a data processor has the right to determine, how to store the personal information, how to delete and retrieve personal data about individuals. Additionally a data processor has the right to decide the methodology which ensuring the retention time scheme. Also, the data processor needs to ensure the processing of the data made in a secure way. Finally, such responsibilities of the processor needs to be agreed with the data controller and established through a contract. [24].

3.2.3 Privacy by Design

Robust and well established methodologies consisting of steps for building software systems have been available for many years. The Waterfall, V-Model and RUP are examples of methodologies software manufacturers can follow in order to, in a systematic way, build software for creating satisfaction to their customers. Further, these methodologies all focuses at defining a strategy for maintaining the so called Software Development Life Cycle (SDLC). Common steps in these methodologies include; requirements definition, solution building, testing and deployment [25]. Furthermore, crucial areas in the context of developing software such as *"innovation, creativity and competitiveness must be approached from a "design-thinking" perspective"* [6].

Privacy by Design (PbD) is a software development approach used by the software industry to provide privacy aware systems. The aim is to address privacy requirements from the early stages of the SDLC. This for the reason same design-thinking reason stated above. Cavokian mean, *"privacy must become integral to organizational priorities, project objectives, design processes, and planning operations."* and thus needs to be addressed by default [6]. Hence, PbD implies a proactive solution by addressing potential threats a system posses with sufficient counter measures. Software elicitation methodologies, technical implementations as well as other means to accomplish PbD into a software system will be explained from section 3.6 and further.

Cavokian [6] explains seven principles which need to be satisfied in order to develop a software architecture through PbD. The principles are; Proactive not Reactive, Privacy as default Setting, Privacy Embedded into Design, Positive-Sum not Zero-Sum, Full Lifecycle Protection, Keep it User-Centric. The first definition has previously been defined in the this section. Privacy as default setting, as the name implies, focuses on providing the maximum degree of privacy in a system. Privacy embedded into design propose to include privacy in technologies, operations and information architectures. Positive-Sum not Zero-Sum relates to the problematic privacy has been connected to prior the existence of PbD. Namely, in order to provide privacy into a software system, other non-privacy goals had to be compromised, referred to as the Zero-Sum approach. Instead, PbD treats such non-privacy goals with the same importance as the privacy goals. In other words, PbD propose to meet both privacy and non-privacy goals with a positive and embracing manner, referred to as Positive-Sum. The fifth principle embraces the need for security in order to establish privacy. This meaning, if sufficient security protection is non-present the privacy protection will be affected. The last principle advocates the need for a user to posses the ability to self determine whether he or she comply with how its data being used. A key way to establish such user empowerment is to ensure user consent. Hence, in order for a system to process and store user sensitive information the user has been informed about such procedures and additionally made an active choice to this. This choice prove the user's understanding of its possibility for retrieval of its information, as well as his or her compliance to how the organization store and process such data.

3.3 Privacy Properties

In order to help understand privacy and how to preserve it in software systems, it is necessary to explain its relevant properties. Pfitzmann et al. [3] have created a well established terminology for privacy. They explicitly define and explain key privacy properties that has become a well established outline for privacy engineers. Thus, their definitions are also used in this section to clarify the privacy properties and their implications.

3.3.1 Anonymity

Anonymity is described by Pfitzmann et al. as *"Anonymity of a subject from an attacker's perspective means that the attacker cannot sufficiently identify the subject within a set of subjects, the anonymity set"* [3]. The definition clearly states anonymity refers to when an individual's personal information remains hidden and thus safe from a malicious attacker. Anonymity can also be described by means of unlinkability when one consider transactions of messages as attributes, as explained by Deng et al. [8]. They explain that sender anonymity of an individual means the messages sent being unlinkable to the individual.

3.3.2 Unlinkability

"Unlinkability of two or more items of interest (IOIs, e.g., subjects, messages, actions, etc...) from an attacker's perspective means that within the system (comprising these and possibly other items), the attacker cannot sufficiently distinguish whether these IOIs are related or not" [3]. Hence, it is not the concerns of the actual IOIs being addressed here but instead the relationship between them. This means, in order for a risk of violation to unlinkability being present in a software system it requires not only one but two individuals, units or actions. Additionally, the IOIs

do in some form interact with each other.

3.3.3 Undetectability and Unobservability

Undetectability and unobservability are two privacy properties that share the same fundamental meaning but differs in who or what it aims to address. The definitions by Pfitzmann et al. explains*"Undetectability of an item of interest (IOI) from an attacker's perspective means that the attacker cannot sufficiently distinguish whether it exists or not"*. Also, *"Unobservability of an item of interest (IOI) means undetectability of the IOI against all subjects uninvolved in it and anonymity of the subject(s) involved in the IOI even against the other subject(s) involved in that IOI"* [3].

As been explained earlier, in unlinkability the focus is at the relationtionship between two IOIs. Now, the actual IOI being of focus. The first definition explains the former privacy attribute's meaning of hiding an entity, action (processing unit) or any sensitive data from an attacker. Additionally the benefit of undetectability is an adversary can unsuccessfully understand whether such entity, processing unit, individual, or other information exist or not. The second definition explains the latter privacy attribute being the same as the former privacy attribute, but from an uninvolved entity's, individual's, action's point of view, even if the IOIs are detected. Specifically, unobservability is similar to undetectability since the purpose being the intention to hide an entity, processing unit, individual or other information from an adversary. The difference lies within unobservability targeting the items with a connection or relation to the item of interest with the exclusion of the item of interest itself. Finally, undetectability entails anonymity and unobservability is a combination of undetectability and anonymity [8].

3.3.4 Pseudonymity

Pseudonymity is a concept where one, most commonly a user of a software service, is assigned a "false name", such as a nickname on a blog post. Pseudonyms can be used in any scenario where an individual exploring a network or system. Further, the network or system needs the possibility to connect actions to the related individual, but still has an interest to provide anonymity. According to Pfitzmann et al. *"A pseudonym is an identifier of a subject other than one of the subject's real names"* [3].

3.3.5 Plausive Deniability

According to Deng et al. [8] plausible deniability refers to when one (for instance a user of a software system or an individual participating in an election) can deny having performed an action that other parties can neither confirm nor contradict. From an attackers perspective it means that the attacker cannot prove that a user knows, has done or has said something.

3.3.6 Confidentiality

Confidentiality, itself, is a not a privacy property. But literature imply it does have an impact on privacy [8]. ISO/IEC 27001 [26] defines confidentiality as *"the property that information is not made available or disclosed to unauthorized individuals, entities, or processes"*. Confidentiality can also be seen as the privacy enforcement of an individual's information which has been disclosed due to trust to the data receiver by the individual itself. In other words, unauthorized parties are not able to understand, and hence unable to draw inferences based on, the content of such shared information. Finally, the previously explained trust, that the the sender makes, imply the expectation of the disclosure of such information will not be spread beyond the data receivers borders. [27]

3.3.7 Content awareness

The preceded and discussed properties are well established concerns in terms of privacy. Although, Deng et al. [8] advocates two more attributes being important in order to reach appropriate privacy levels in software systems. The first attribute being content awareness and alludes an individual, that using a software system, is aware of its personal data. When individuals suffering from content unawareness unnecessary or unintended dissemination of the individual's personal data might occur. By ensuring the individual understand how such data is used as well as what information is sent to other parties, by the system the individual can itself decide to take actions to prevent or mitigate potential privacy threats. Clearly, such knowledge should be provided to the individual before the individual does send such information to the system. Finally, as a golden rule a software system shall only inquire for and use the absolute minimal information about its individual.

3.3.8 Policy and consent compliance

Policy and consent compliance is the final attribute. By ensuring the system's policy and the user's consent, in the form of a textual representation, are indeed implemented and enforced in the system this can be considered to be an attribute that address the demands from privacy legislations and regulations [8]. This privacy property has a connection to the content awareness property explained previously. This, since one way to implement content awareness is to ensure the individuals of a system knows about the system's policy. A policy which explains how the individual's information being stored, processed and potentially disseminated by the system. The individual then confirms such policy has been received, read through and additionally agree to it through a consent to the system.

3.4 Threats to Privacy

Threats to privacy are potential violations that can jeopardize the data subjects right to privacy. Deng et al. [8] has defined seven privacy threat types derived from the eight properties discussed in the previous section. Further, the threats have a direct correlation to the previously mentioned privacy properties. The seven privacy threats types are linkability, identifiability, non-repudiation, detectability, disclosure of information, content awareness, and policy and consent non-compliance. They are further explained in this section.

1. **Linkability:** linkability threats violate the unlinkability privacy property. It allows an attacker identify if two or more IOIs are related to each other, even not knowing the real identity of the data subject. Examples of linkability threats are: *"anonymous letters written by the same person, web pages visits by the same user, entries in two databases related to the same person, people related by a friendship link"*. [1]

2. **Identifiability:** these threats violate the anonymity and pseudonymity properties. It allows an attacker identify, from an identifiability set, the data subject related to an IOI. Examples of identifiability threats are: *"identifying the reader of a web page, the sender of an email, the person to whom an entry in a database relates"* [1]. This means that an individual's identity being exposed against its will and hence the attacker can use this information to establish harm to the individual.

3. **Non-repudiation:** non-repudiation threats violate the plausive deniability property. The attacker gathers information to prove that a user knows, has done or has said something and the user are not able to deny it. Examples of non-repudiation threats are: *"anonymous online voting systems, and whistle-blowing systems where plausible deniability is required"* [1].

4. **Detectability:** these threats violate the undetectability and unobservability privacy properties. It allows an attacker identify weather an IOI exists or not. Example of detectability threat are: messages that are identifiable from random noise [1].

5. **Disclosure of information:** disclosure of information threats violates the confidentiality privacy property. It allows an attacker expose personal information that are not supposed to be shared with other unauthorized individuals [1].

6. **Content awareness:** these threats violate the content awareness privacy property. When users are not aware of the consequences of disclosing personal information to a system, they provide an attacker easy access to their identity or provides wrongful information that can led to incorrect decisions or actions [1].

7. **Policy and consent non-compliance:** policy and consent non-compliance threats violate the policy and consent compliance property. It means, even if the system provides its privacy polices for the user, these policies might not be compliant with regulations, legislation and corporate policies. Thus, threatening the user's personal data [1].

3.5 Privacy Regulation - GDPR

With the new General Data Protection Regulation (GDPR) [5], provided by the European Union Parliament, coming into place in May 2018 organizations are being pressured to establish a planned and strategic road map to maintain user's data privacy. For organizations that has not already established managerial procedures to comply with already existing legislations similar to GDPR, such as the UK DPA from 1998 [28], this will not be an easy fix that will be accomplished over a night shift. Organizations are forced to comply with procedures and audits to demonstrate how personal data is being processed and stored. Furthermore, the focus is at the importance of the accountability of organizations, EU citizens rights to their own data as well as an increased understanding for the users how their personal data is processed by the organization [29]. From May 2018, the user of a software system shall have the rights to be informed regarding how the system is processing its data. This will allow the individual to feel empowered to have an impact of how the system will use the individual's personal data [30]. According to Article 84 (5)(a) in the GDPR [5], the implications of not sufficiently supporting the privacy provisions of the regulation will make the organization be a victim of a fine with a sum of 20 000 000 EUR or four percent of the total volume of sales of the preceding year.

Reviewers of GDPR do not only look from the perspective of the user's rights to privacy. Also, the regulation can be seen as a mutual benefit for both organizations and its users. For organizations, the biggest benefit is the possibility to increase their reputation and trust towards its users due to the accountability demands from the regulation. A factor which can be considered a positive outcome for a company or organization in the long end. This because of a believed increased user satisfaction due to the user's empowerment over its own data that the GDPR comes with. Moreover, there are alternatives for an organization to be put in practice in order to side step potential fines invoked due to the inability to reach sufficient compliance for using consents. However, this is considered to be outside of scope for this study and thus further investigation of such alternatives will not be conducted.

As explained earlier, the new privacy regulation impose more strict policies during data processing. Due to this, the importance of Data Controllers and Data Processors and their responsibilities are increased. Their difference has previously been explain in section 3.2.2. Since the controllers and processors are responsible for data processing the GDPR impose to perform a Data Protection Impact Assessment (DPIA) at these locations, Article 33 [5]. This, not only due to security concerns but also to address privacy concerns [31]. The imposed Data Protection Impact Assessment can be considered as being the same category of threat and risk assessment as the Privacy Impact Assessment (PIA) explained in section 3.7.2.

The new GDPR guidelines are based on eight data protection principles that define the organizations main responsibilities when processing personal data. The principles are defined under the Article 5 of the GDPR [32] and are described as follows:

1. **Lawfulness, fairness and transparency:**
   This principle requires that personal data shall be processed lawfully, fairly and in a transparent manner in relation to individuals. The main goal of this principle is *"to protect the interests of the individuals whose data is being processed"*. This principle applies to everything concerning personal data, unless if granted an exemption. In order to collect and use personal data, it is necessary to have purpose, be transparent regarding the intended usage of the data, provide adequate privacy notices to the individuals which data is being collected, and be compliant with the law.

2. **Purpose:**
   Data shall be collected for specified, explicit and legitimate purposes and can not be disseminated in any other matter than for what it was first specified. This principle states that organizations must be clear regarding why they are collecting data and what they intend to do with the data. Moreover, it is important to ensure that data is not being used for any other purpose than for what it was originally specified.

3. **Adequacy and data minimization:**
   The third principle impose that the collected personal data shall be adequate, relevant and limited. Organizations need to ensure they are collecting the minimum amount of data necessary to their purposes. This means that the data that is collected should be just sufficient to their specific purpose and that they are not collecting more information than they actually need. Hence, the regulation suggests to perform generalization and data minimization techniques. This, for the aim to reduce the risk of malicious inferences at scenarios where too detailed data is used in during a processing.

4. **Accurate and up-to-date processing:**
   This principle instructs organizations to keep an accurate and up-to-date data processing in the system. Any personal data that are inaccurate, in relation to its purpose, shall be erased or rectified with no delays. In order to comply with this principle, organizations must take measures to ensure the accuracy of any personal data they are collecting, that the source from which they are collecting personal data is clear, and consider if this personal data needs to be updated in the system.

5. **Storage limitation and retention time:**

This principle is closely related to both principle three and four, and states that organizations shall keep personal data in a form which allows identification of data subjects for no longer than is necessary in relation to its purposes. By ensuring that personal data is erased when no longer needed will decrease the risk of this data becoming inaccurate, out of date, or irrelevant. Organizations need to review the amount of time they need to keep personal data. To do so, they need to consider the purpose for what they are collecting data, delete data that is not longer required for this purpose, and update, archive or safely delete personal data that is out of date.

6. **Individual rights:**
   The sixth principle gives individuals the rights over their personal data. It states that *"personal data shall be processed in accordance with the rights of data subjects under this Act"*. The rights which the Data Protection Act refers to are:

   - Right to be informed – individuals have the right to be informed regarding the collection and usage of their personal data. This right is also related to the first principle, lawfulness, fairness and transparency. The right to be informed is a way to accomplish transparency, where organizations are required to provide privacy notices containing the purpose for collecting and processing their personal data, the retention time, and if this data will be shared. This privacy notice shall be provided at the time of the collection. However, in case the individual already has the information, this become unnecessary. Moreover, the information provided must be easily accessible and contain a clear and plain language.

   - Right to access – this states that individuals have the right to access their personal data and allow them to be aware and verify if the processes are lawfully compliant. Individuals have the right to obtain confirmation that their personal data is being processed, and any other supplementary information. This information must be provided with no delay (latest within one month), and free of charge.

   - Right to rectification – this gives individuals the right to have their information rectified if this data is inaccurate and incomplete. In case that personal data has been forwarded to another party, it is necessary to communicate the rectification, unless if this is, for a certain reason, impossible.

   - Right to erasure – is also known as "the right to be forgotten". This allows individuals to request the erasure and removal of their personal data. This applies to personal data that is no longer necessary in accordance to its purpose, when individuals withdraw the consent given, when individuals contest the processing or the personal data was unlawfully processed.

- Right to restrict processing – this right allows individuals to restrict the processing of their personal data, by blocking or suppressing it. This means that organizations are allowed to store personal data, but not forward it to further processing. When an individual objects the accuracy and the processing of the data, organizations must restrict the processing of this data until the accuracy and processing of this data has been verified.

- Right to data portability – allows individuals to obtain their personal data and reuse it through different services. This way, individuals can move, copy or transfer their personal data from one IT platform to another, easily, safely, securely, without any interference to the data usability. Formally, data portability means a data controller shall give the right of the user to request a transfer of its personal data to another data controller without any hindrance from the data controller which holds the data subjects personal data. Furthermore, the right to portability also applies when the data processing is done through automated means.

- Right to object – individuals have the right to deny access to their private data for direct marketing, profiling, and scientific research and statistics. Organizations have to prove that the data is being used according to its initial purpose and does not infringe the individuals rights, interests or freedom. The right to object has to be informed to the individual through the privacy notice at the first time of the data collection. This must be clearly stated and separated from any other additional information. At any time the individual can object to direct marketing and organizations must accept it free of charge. Upon a request of objection, the data controller can no longer process data which the request relates to, unless the controller can demonstrate legitimate and legal grounds for continue processing such data.

- Rights related to automated decision making including profiling – this right protects the individual rights at scenarios where automated individual decision-making and profiling (automated processing used to learn specific things, behaviour or characteristics, regarding an individual) are present. Automated decision-making or profiling activities can only be performed when this has been consented to by the individual, by law or stated in a form of a contract. In this case, organizations must perform regular checks to verify that the systems are performing as intended.

7. **Security:** This principle refers to appropriate technical or organizational measures used to protect the individual *"against unauthorized or unlawful processing of personal data and against accidental loss or destruction, or damage to, personal data"*. Organizations shall apply a risk-based approach according to their circumstances involving personal data and the level of security they need in order to protect this data. In general, organizations must organize and

design security practices in order to act effectively and promptly in case of any breach of security. Additionally, this principle protects the integrity and confidentiality of the personal data.

8. **International:** The eight and last principle ensures that companies and organizations that want to send data outside the European Economic Area (EEA) must comply with the Data Protection Act. Specifically, personal data shall not be forwarded to anywhere outside the EEA. Additionally, in cases this data is transferred to another country or territory, they must have enough grounds to protect this data.

If organizations want to comply with the Data Protection Act they must follow these eight principles, providing accountability and liability of any and all collected and processed personal data.

3.6 Privacy Solutions

To this point the reader has been given an explanation regarding what privacy means, to whom it applies, to who it applies and common activities it is concerned. In this section an explanation of means one can use in order to mitigate or counter the previously discussed privacy issues. Explicitly, concepts such as patterns and technical implementation strategies are explained. These are alternatives which software engineers and architects can choose to implement during the design phase in order to enforce the the privacy levels in a software system.

3.6.1 Data Classification for Privacy

Enforcing privacy in a software system can quickly become a hassle. Explicitly when one tries to understand what data should be treated as sensitive and which should not. To consider all data in a system as sensitive is unsustainable and often not feasible in terms of cost and performance overhead. Especially when a system grows including more processing and storage of information. An alternative to resolve such problematic can be accomplished through the usage of data classification. By dividing data in different types of groupings (classifiers) one can categorize data to easier understand the uttermost sensitive data in a software system. By understanding the parts where sensitive data being processed and stored, software designers and architects can prioritize and address privacy enforcing strategies to the system in a more sustainable way. A classifier includes a *function* which possess the ability to labeling data into classes by using a set of *attributes*. Such attributes describes the class so the function can determine if a data belongs to that specific class. [33].

3.6.2 Privacy Patterns

To support privacy by design Hoepman [34] proposed a methodology based on the use of software design patterns. Hoepman defines the concept of a design pattern as *"a useful vehicle for making design decisions about the organization of a software* *system"* while Buschmann et al. [35] describes design pattern as a scheme used to refine subsystems or components of a software system, or the relationship between them.

A common description of a design pattern [36] includes its name, purpose, context (when it should be applied), implementation (structure, components and their relationships), and the consequences (the results, effects and trade offs after being applied). "Design patterns may overlap, and may vary in the level of detail they provide; A privacy pattern may sometimes implement several privacy design strategies."

Hafiz [37] has developed a catalogue of various design patterns that focuses at enforcing privacy in software systems. Additionally, he uses Pfitzman and Waidners [38] three different classifications of a potential attackers depending on their motivations and actions made to perform privacy violations. These three classifications are; the active, semi-honest and passive attacker. A passive attacker is one who monitors, for instance through eavesdropping, a data package on a network. An active attacker is one who is not satisfied with the outcome of only monitoring but instead actively performs an action for the purpose of manipulating the data in a network. Finally, the semi-honest attacker is one, who by others, seems as an honest actor in the network but on the contrary has malicious intents. This could be an individual that, instead of only being an eavesdropper, gets access to information about others through being seen as a trusted individual. Instead of performing actions an honest actor would do the semi-honest attacker performs privacy malicious activities. Hence the difference between the active and semi-honest attacker being the effort made to attempt to fool other trusted entities in the network of being such trusted entity or not. Hafiz catalogue has later been extended with eight additional privacy patterns [39]. The remaining of this section will describe these twelve privacy patterns with the Hafiz papers as reference [37, 39].

3.6.2.1 Anonymity set

At its most basic explanation the anonymity set pattern is the concept of mixing data which is considered sensitive with other data for the reason of it acting as an information noise from an attackers point of view. Hence the coveted result being an attacker that attempts to monitor the transaction of data is unable to distinguish the desired data from the added information noise. Hafiz [37] explains this pattern being relevant when there is a desire for designing a system that will maintain send or recipient anonymity, or both, in a messaging scenario. He explains the pattern using a network technology called mix networks where a sender´s data package being, before reaching its receiver, placed into a mixed network where other senders packages being assembled as well. The mixed network is further explained in section 3.6.3.6. When sufficient data packages being assembled in the mixed network they are all being broadcast out to each of the recipients. This meaning each receiver receiving data packages sent from every sender in the mixed network despite their interest in those data packages.

The benefits with the anonymity set pattern are the increased extra work a potential attacker would need to perform in order to distinguish the information of interest and thus an increased privacy of the individual being present. However, the drawback being, in the given scenario with the mixed network, the pattern relying on sufficient data packages reaching the mixed network before the senders data package can be sent. If a data package would be sent with a too small set of data packages this would decrease the anonymity of the sender and receiver. Waiting for sufficient amount of the data packages can develop an increased latency and hence result in a performance bottle-neck. A typical technology being used with this pattern being k-anonymity, which being explained further in section 3.6.3.4.

3.6.2.2 Morphed Representation

The major drawback with the previously explained anonymity set is the in-going and outgoing data packages, to and from the individual, do contain the same data fields. Thus it is possible for an attacker to map the data traffic back to the specific individual, even though it is being encrypted. Instead Hafiz [37] explains, by using obfuscation of the data packages sent in a network the intention of this pattern is the accomplishment of any outgoing data not being a victim of linkage to any incoming data. Again, Hafiz [37] uses the example of the mixed network. He explains this is accomplished by the data sender encrypting the package and sends it to the mixed network. The first node in the network is receiving the data package. Since the encryption key used by the sender is known by the the receiving node, the node can decrypt the data package for the reason of re-encrypt the package with a new key. Just as the new key was shared between the sender and the initial mixed network node the same reasoning exists between the first network node and the next network node. This scenario is repeated until the data package has ended at its intended receiver. The result being, a passive attacker will not be able to identify the package due to that its exterior has been altered at each node it has passed through the route of the network nodes.

Clearly, the benefit here is the increased privacy due to the data package changing form at every node it pass through until it reach its end receiver. Although the drawback being performance overhead, the decryption and re-encryption needed at each mix node the data package is passing through before reaching its end node. Additionally, each mix node where the data package pass through needs its own key to re-encrypt the data package. To conclude this pattern provide a means of unlinkability.

3.6.2.3 Hidden Metadata

The previous pattern does perform an increased use of cryptography throughout the flow in order to accomplish an increased privacy enforcement to prevent a passive attacker to identify a data package and link it to an individual. However, possibilities for an attacker to extract private information of an individual on a network still prevails. This due to the meta data which is included in all traffic sent. In a scenario where an email being sent to a receiver, such meta data could be the email address which commonly has a connection to the individual of the sender of the email. Hence, since the attacker being able to read the sender's email address the attacker has access to some degree of sensitive data. This is also the main concern of the hidden metadata pattern. Another example is given by Hafiz [37] where the data package header, in this case the IP address, discloses private information about the sender's identity and location. Instead to prevent such disclosure Hafiz explains a solution where the metadata being obfuscated with the body of the data package. The solution includes a proxy which performs the actions on the behalf of the sender but strips out sensitive data of the metadata. In contexts where a system aiming at location anonymity, where pseudonyms commonly are used, this pattern can be of interest since the association an individual has to its pseudonym being stripped out.

This pattern does not include the computational overhead of cryptography which the previous patterns do. Hence, the issue of a wide-scale deployment architecture is not present for this pattern. Although, precautions need to be considered so that the system´s availability requirements are not violated due to the risk of a Denial of Service at the proxy if too many actors using the system. Additionally, the proxy needs to remember the incoming data due to it needs to know where to send the content back to. Specifically, the incoming data the proxy receives by performing the actions on the behalf of the originator of the request. [37].

3.6.2.4 Layered Encryption

This pattern make use of the same mixed network which has been used as exemplification scenarios in previous patterns. The mixed network is explained in section 3.6.3.6. The central part of this pattern consists of the user determines a route for which being the flow the data package will travel in order to reach its receiver. This route is determined before the data package is sent by the sender. To accomplish sufficient privacy levels the data is encrypted in several layers. Hafiz [37] relates this to the layers of an onion. The reason being the pattern restricts the amount of content that is accessible to each of the mediating nodes in the mixed network. Nodes which handle the data package on its way to its destination. As well as in the pattern morphed representation, each of the nodes in the mixed network own separate encryption keys for the purpose to be able to decrypt the data package at each node. The privacy advantage compared to morphed representation is that each of the mediating nodes has no more knowledge about the flow, that the data package is traveling, than the next node it being forwarded to. Additionally, the sender of the data package determines the route, which is considered a strong benefit. This pattern is not only strong against passive but as well as semi-honest attackers. Imagine, a data package being received to a semi-honest attacker acting as an ordinary node in a mixed network. The semi-honest then modifies the data and finally send it further to the next node. This next node will detect such malicious modification due to it being distorted and hence instead of forwarding the data package in a direction to the receiver the node sends it in the reversed direction so the sender being aware of such privacy violation being made. To conclude, this pattern provides an improved confidentiality to the system where it is applied. [37].

3.6.2.5 Pseudonymous Identity

As been explained in section 3.3.4 pseudonymity is when someone is assigned a "false name", or alias, for identification purposes instead of its real identity. As the name implies, this is a pattern which has its focus on keeping the connection between a pseudonym and the corresponding individual identity hidden from an potential attacker. After all, if an attacker is able to bypass the protection pseudonyms provided, individuals are left exposed. Hafiz does not explicitly go in-depth how one can enforce the protection of such pseudonyms. He explains this being due to it will vary between domains. Furthermore, he explains in a location based system, such as VANETs [2], one can use location technologies by which provides a less accurate positioning [39]. This making the driver of the vehicle being less exposed.

3.6.2.6 Chaining

An architect designing a system should not rely on only one anonymity solution but instead strive to provide a multi-layered anonymity solution. This is also the fundamental aim of this pattern, to combine multiple patterns with the goal to enforce the system´s privacy protection by ensuring if one privacy solution fails another can have its back. The bottom line is explained by Hafiz as "*adopt multiple instances of an anonymity mechanism in layers.*" as well as "*chain multiple mix nodes instead of one and route traffic through multiple nodes*". If the system using pseudonyms then multiple pseudonyms should be considered in order to protect the users sufficiently. [39].

3.6.2.7 Batched Routing

The batched routing pattern advocates that a data package should not be sent alone but instead being included in collection of other data packages. When the collection has reached a desired amount of data packages the collection is sent as a whole. This will provide an increased privacy protection by means the individual will be less exposed to an attack. Hence, this pattern aims at achieving anonymity. However, unlinkability prevails since an attacker can monitor the traffic and infer correlations of the data over time. [39].

3.6.2.8 Random Wait

The previously explained pattern Pseudonymous Identity can be a suitable privacy pattern in the vehicle domain and VANET systems, later explained in section 3.8.1. This applies to the pattern Random Wait as well. Most of the previously discussed patterns address privacy protection mechanisms and the architectural solution also comes with a performance bottle-neck. Hence, systems where the sending unit suffers from high latency might not be able to use such pattern. Instead this could be an alternative, since the decreased frequency in which data packages will be sent from the sending unit. Hafiz explains "*Add random delays to make a system non-deterministic. Keep a packet in the pool, and forward it immediately after the delay is passed*". [39].

3.6.2.9 Cover Traffic

The Cover Traffic pattern proposes to create dummy entities in order to obtain a better anonymity set. This can be achieved by creating false entities (cover traffic) and mixing them with the entity of interest to generate an anonymity set. *"Cover Traffic creates dummy data and adopts Link Padding to create a continuous flow of dummies between nodes"*. The creation of an independent traffic load produces a more secure cover traffic policy. [39].

3.6.2.10 Link Padding

The Link Padding pattern encourages to keep a constant traffic on each outbound link of a network node. It suggests to *"keep traffic flowing on all outbound links and pad the links with cover traffic, even when actual payload is absent"*. Further, the network node shall hide information regarding payload traffic. Link padding follows the cover traffic pattern and affects the creation of cover traffic policy. [39].

3.6.3 Privacy Enhancing Technologies

Literature explains Privacy Enhancing Technologies (PETs) as Information and Communication Technologies (ICTs) which can be applied to a software system, which using personal data of its users, in order to enforce privacy means. Specifically, Borking and Blarkon et al. [40, 41] explains PETs as *"a system of ICT measures protecting informational privacy by eliminating or minimizing personal data thereby preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system"*.

As has been explained earlier in section 3.6.2, a pattern is a tool which can be used to refine a system, or subsystem, on a design level and additionally implements one or several strategies to fulfill a pattern's desired outcome. Such strategies are in a software architectural context referred to as tactics where their purpose is to promote different quality attributes. This can be confusing for a novice privacy engineer, due to the similar explanation that has been given for the patterns. One can think of an architectural pattern as a description to a recurring problem in a design context. Additionally such pattern provides a well-proven generic scheme for its solution. Tactics, on the other hand, can be understood as the "building blocks" to an architectural design from which a pattern is created. [42]. Thus, a pattern can be seen as a known solution to a problem. Further, the solution is implemented through using one or several tactics.

With such difference between patterns and tactics having been explained the reader can think of a PETs as the similarity to architectural tactics. This since *"PETs are used to implement a certain privacy design pattern with concrete technology"* [34].

3.6.3.1 Crowds Anonymizer

This is a technology which blends groups of users on a network together so the users all represents one and the same group. Such group is referred as a Crowd. Since the individuals all come from different locations the Crowd represents a more diverse and broader location pool, the risk of an individual's origin being exposed is mitigated. Hence, the purpose with Crowds Anonymizer is to provide means of anonymity protection of individuals. Specifically, the target domain of this technology is the network World Wide Web. The intention, of this PET, is to keep the individual protected against three potential adversaries when individuals perform a web request to a web server. These are eavesdroppers, malicious collaborating crowd members and the end server. Eavesdroppers being one who monitoring a transfer of messages. Malicious collaborating crowd members can either assemble information about honest crowd members and potentially deviate from the network with such information or with a malicious end server. [43].

3.6.3.2 Onion Routing

As has been explained in section 3.6.2.3, the meta data between sending and receiving nodes is a concern for systems which use communication through any network. Such metadata can be the origin from where a message is sent, through the IP protocol. Further, the exploits in this context can be countered with an Onion Routing service. Onion Routing is a distributed overlay network in which web applications, and other messaging systems on a network can use in order to accomplish anonymity. Initiating nodes choose their own route in which they desire a message should take before reaching its receiver. More technically, a proxy server (which the initiating sender redirects the message or request to) determines the route for the message. The benefit being no party have the knowledge of any jump to and from any mediating node during such route, except the proxy server itself. Further, the need for identification at the sender's proxy server, any external observer or the receiving parties proxy server is non-present [44]. An example of such Onion Routing services is TOR [45], a service commonly used by individuals which explore the World Wide Web and has an interest in anonymity, plausible deniability and unlinkability.

3.6.3.3 P3P

The Platform for Privacy Preferences (P3P) is a project which focuses at developing privacy protocols which can serve as a standardization for web applications and other user agents (service providers) to lean against in order to accomplish privacy best practices. [46]

The P3P project defines a process which includes an aim for increasing the availability of privacy policies for users. This so a user in an easier way can access and interpret how a web application processes and storing user input data. The protocol defines a standard vocabulary for describing a privacy policy. A difference exists in the degree of detailed content of the digital policy compared to a real and non-digital policy. This due to the privacy policy created under this protocol is derived from a set of multiple-choice questions and are merely intended to be read by machines. Following, the P3P project defines a second protocol for sending and receiving such privacy policies by using the standard networking protocol HyperText Transfer Language (HTTP). The user agent can retrieve a user's privacy policy (which states the conditions the user agrees on its data being processed under) to make actions accordingly. [47].

3.6.3.4 K-anonymity

This is a technique which acts as an enhancement of unlinkability of data subjects in a system where it is applied and hence also a protection against possible inferences to data aggregation. The model focusing on protecting outgoing data from the system. In other words data that can be accessed by external actors. Sweeney propose to introduce a variable k and an identifier in an anonymity set. In her examples the anonymity set is a private table in a database, an identifier being a record in a database including a tuple of columns and k being a literal. The k-anonymity model then compares if such exact tuple occurs at least the amount of times equals to k. If all tuples satisfy this condition in the anonymity set then the anonymity set is considered to satisfy the model and hence is considered anonymous. [48].

3.6.3.5 Differential Privacy

Differential privacy is a privacy preserving technique which has been derived through statistical research. Drowk [49], its creator, explains the main reason behind differential privacy being to prevail privacy of the true information generated when executing an input query to any database. The generated result from performing a query towards a database being referred to as transcripts. As privacy can vary between different contexts differential privacy uses a variable ϵ (epsilon). This is a variable which determines the allowed probability of a record being identifiable when two datasets are compared. The definition precisely says: *"A mechanism is ϵ-indistinguishable if for all pairs $x, x^1 \in D^n$ which differ in only one entry, for all adversaries A, and for all transcripts t"*.

This privacy enhancing technology uses different mechanisms in order to process the data. Such mechanisms being; the laplace, exponential or randomized response mechanism. Each of which process the data in its own way to satisfy different privacy purposes. Here the term mechanism referring to an algorithm which can execute a database query and obfuscate the result so the return data contains a higher degree of noise. An adversary being any actor with a privacy malicious intent. The variable x and x^1 are two different data subsets derived from the dataset D^n. Commonly such data subsets can be two datasets from two different databases, as in the scenario when data aggregation is applied. Differential privacy is the concept of hiding the sensitive data generated from an executed database query by adding noise to this sensitive data and return this to the initiator of the query.

With the benefit of differential privacy prevailing the integrity of sensitive data being explained, another key outcome is it still being possible to make valuable predictions by analyzing such data. This is referred to as data utility. In other words the possibility to make valuable inferences of data which one can use in order to understand a problem. A good balance between privacy and utility is sometimes hard to establish, because differential privacy is enforced by making the data less accurate while increased utility results in the contrary.

Nelson et al. [50] explains the possibility and benefit of applying differential privacy into the automotive domain in the future. The reason being automotive systems having high usage of personally sensitive data, where the biggest and the most sensitive type data is the use of location based data (GPS data) of vehicles. They conclude, due to the complexity of automotive systems (where an extraordinary amount of computational units are used) differential privacy can be rather hard to enforce. Improvements for making it more suitable for automotive systems still prevails, due to the balance between utility and privacy of data. To make use of differential privacy in the context of automotive systems a proper model needs to be established, a model which explains what data needs to be protected.

3.6.3.6 Mixed Network

A mix is a computational unit which possess the power to encrypt data. The mix is used as an intermediary between a sender and a receiver, so a sender's and receiver's anonymity can be established. By encrypting data with a publicly shared key from the sender and sealing the data to be encrypted with a secret key. This secret key is only known by the receiving unit making it being the unit having access to the content of the data. The combination of several mixes creates a so called mixed network. The power of mixed networks can be illustrated by exemplifying the relation between the two mix units A and B interacting with each other. The mix A sending a message to the mix B while A being able to remain anonymous to B. Additionally, B can respond back to A without having A reveled its identity for B. In other words, a sender can send messages to another unit while being anonymous and the receiving unit can still respond to the sender without knowing its identity. However, the possibility still exists for supporting anonymous identification if necessary. The mixed network can be applied in such way that pseudonyms can act as an identification of the individuals using the network. Assuming several actors using the network, the network halts the forwarding of a message as soon as one of the several actors sending a message. Instead, the message being placed to a collection where it is waiting for being forwarded. When a given amount of messages being allocated to the collection of the mixed network they are all broadcasted together as one data package. This meaning the resulting body of the broadcast being multiple messages, each having a pseudonym being attached to it. Finally, the individuals can still being publicly identifiable but still prevailing the anonymity of the sender. [51].

3.6.3.7 PROBE

Privacy-preserving Obfuscation Environment (PROBE) is a framework which targeting at enforcing privacy in location-based services (LSB). The problematic of an individual using a LSB, such as an app in a mobile or a GPS service in a vehicle, lies within the individual being exploited to adversaries reading the data-location traffic sent from the LSB. Damiani et al. [52] explains that even though known obfuscation techniques are used the privacy risk of the individual persist.

PROBE defines regions around the individual using a LBS. Each region is assigned a degree of sensitivity that depends on the type of location (public market, central station or hospital to name a few). Further, the degree of sensitivity depends on the type of location a region representing. This, because the probability of the individual being located inside that type of location varies. The individual is later given the responsibility to determine the possibility of he or she being inside such region. The framework uses this information to obfuscate the data sent from the LBS to include both sensitive and non-sensitive regions. PROBE does not take the individual's specific position explicitly into account. Instead it uses the surrounded regions to accomplish means of obfuscation and hence provides anonymity to the user this way. Although, the possibility still persists that the region the individual is located within still being included in the obfuscation [52].

3.7 Privacy Analysis Methodologies

3.7.1 LINDDUN

LINDDUN [1] [8], [53], is a privacy threat analysis methodology used to identify privacy threats during the early stages of the software development process. The methodology allows requirement engineers and software architects to identify privacy weaknesses and fulfill privacy requirements when developing software systems. The name "LINDDUN" is an acronym that stands for the seven privacy threat types presented under the section 3.4, which are: Linkability, Identifiability, Non-repudiation, Detectability, Disclosure of information, content Unawareness, and policy and consent Non compliance. These threats are presented by LINDDUN as threats to the privacy properties presented under section 3.3.

The LINDDUN methodology is derived from STRIDE [54], a well established framework developed by Microsoft to elicit security threats in software systems. Both methodologies are model-driven and use the data flow diagram (DFD) notation as starting point for the analysis. The DFD describes how data enters, leaves and traverses in a system. This is explained by four different types of components; external entities, data stores and processes and data flows, that are represented as seen in Figure 3.1.

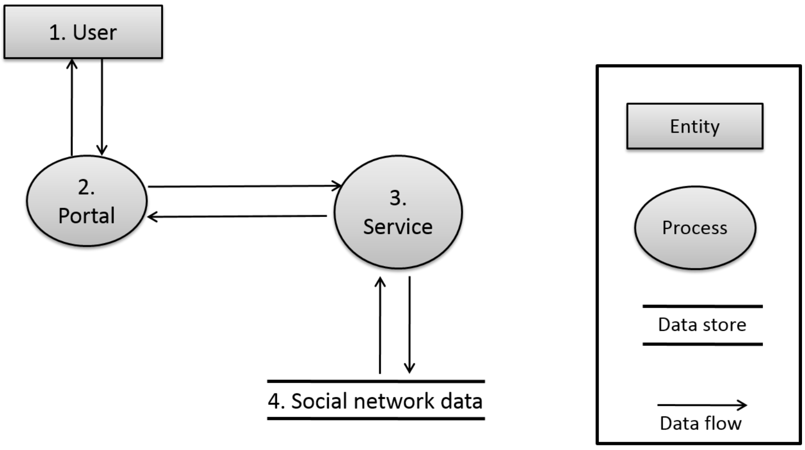


Figure 3.1: The Data Flow Diagram Notation [1].

An external entity is any individual or external actor outside the system and additionally being out of the architects control. A data store is any type of container where data storage is used e.g. databases or files. A process means any part of the system with executable code. Finally, a data flow represents the communication between processes or between processes and datastores, e.g. network communications via HTTP.

LINDDUN consists of six steps, where the first three steps are considered the core of the methodology as they focus on the problem domain and the identification of the privacy threats concerning the system. The other three steps are considered by LINDDUN as solution-oriented, since it provides possible solutions for the presented threats. The steps are presented bellow:

1. **Define DFD:** the analysis starts with the creation of a DFD based on a high-level description of the system.

2. **Map privacy Threats to DFD elements:** the second step consists in mapping the DFD elements with LINDDUN components. For this LINDDUN provides a mapping template as seen in the Table 3.2 below:

Table 3.2: LINDDUN mapping template.

	L	I	N	D	D	U	N
Entity	x	x				x	
Data store	x	x	x	x	x		x
Data flow	x	x	x	x	x		x
Process	x	x	x	x	x		x

3. **Identify Threat Scenarios:** the third step of the methodology consists of a set of sub-steps that include the elicitation and documenting of the threats. All the elements mapped in the LINDDUN template are thoroughly inspected in order to identify if they represent potential threats to the system. This examination is made with the help of privacy threat trees provided by LINDDUN. The threat trees represent the common attack paths according to each specific LINDDUN threat category and DFD element type. An example of a threat tree representing the privacy property Policy and consent non-compliance is shown in Figure 3.2. After the elicitation of the threats is done, the second sub-step is to document the threats according to a threat description template. For this, LINDDUN suggests a template used to document misuse cases as appropriate description template.

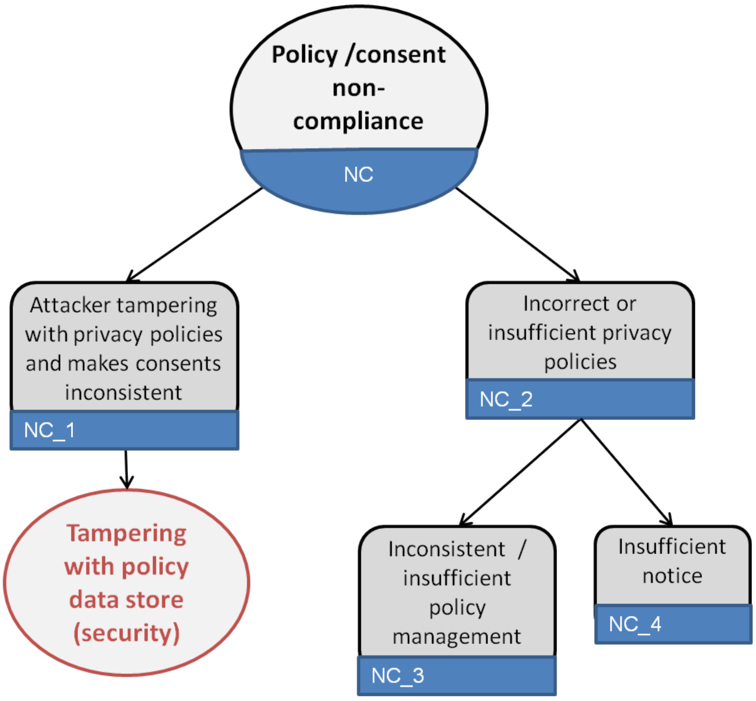

Policy /consent non-compliance
NC
Attacker tampering with privacy policies and makes consents inconsistent
NC_1
Incorrect or insufficient privacy policies
NC_2
Tampering with policy data store (security)
Inconsistent / insufficient policy management
NC_3
Insufficient notice
NC_4

Figure 3.2: Example of a LINDDUN Threat Tree [1].

4. **Prioritize Threats:** the fourth step consists of a risk assessment analysis where the misuse cases are prioritized according to its relevance. LINDDUN does not provide any specific guidelines to risk assessment, as is it up to the analyst to decide which risk assessment strategy to use.

5. **Elicit Mitigation strategies:** the fifth step consists of identifying the suitable mitigation strategies for each privacy threat. The analyst can either choose a proactive or reactive mitigation strategy.

6. **Select corresponding PETs:** the final LINDDUN step consists of selecting appropriate privacy enhancing techniques (PETs) as mitigation strategies for the given privacy threat.

3.7.2 Privacy Impact Analysis (PIA)

Privacy Impact Assessment (PIA) is a process which addresses the need for finding threats towards a software system's privacy needs. By performing a Privacy by Design evaluation PIA considers "*privacy of the person, privacy of personal behaviour and privacy of personal communications, as well as privacy of personal data*" where the emphasis is at the assessment of "*information exchange, organizational learning, and design adaptation*" [55]. Another explanation, is given by the Information Commissioners Office [56], which explains the methodology as a *"process which helps an organization to identify and reduce the privacy risks of a project."*. Additionally, the term project meaning *"any plan or proposal in an organization, and does not need to meet an organization's formal or technical definition of a project"*. Thus, a PIA can involve a rather broad target audience from a managerial point of view to a more technical implementation level. When a project increases in size a PIA can become substantially larger. ICO [56] explains, not all projects need the same depth of PIA. Typically, systems that processes sensitive data (such as user information) requires a more thorough PIA. A PIA can be extended beyond the design phase timeline. It can be considered applicable to execute such PIA in parallel with the development as long as the PIA is well implemented.

A PIA consists of six steps, they are:

1. *Identify the need for a PIA.* By using screening questions and discussions with stakeholders one can determine whether privacy threats are present in the project or not and hence decide if a PIA shall be conducted.
2. *Description of the information flows.* Explanations for how information will be obtained, processed and maintained will be put on the table.
3. *Identification of privacy and related risks.* A common way during this phase is to perform a compliance check to a Data Protection Act (DPA) to reveal potential privacy violations to the project.
4. *Identification and evaluation of privacy solutions.* Discover mitigation strategies for the found privacy threats. Asses the worth of applying such mitigation strategies, investigate cost over beneficial outcome.
5. *Sign off and record the PIA outcome.* Produce a PIA report which states the documentation made in the previous phases.
6. *Consult with internal and external stakeholders throughout the process.* Ensure the previous steps have been implemented according to the PIA.

3.7.3 PriS

PriS [9] is a security requirements engineering method that integrates privacy requirements in the early stages of system development. The method addresses privacy requirements as organizational goals and applies privacy-process patterns to satisfy these privacy requirements. Prior the existence of PriS the author to the methodology claim the absence in the literature for security requirement methodologies which mapped identified requirements with implementation solutions . Thus the reason behind the development of PriS.

As has previously been explained, the methodology aims at eliciting organizational goals and from these goals derive privacy requirements. Kallaniatis et al. advocates the term privacy requirements being rephrased to privacy goals meaning the privacy goals are able to be subdivided into smaller goals.

PriS does consider eight different privacy properties, namely authentication, authorisation, identification, data protection, anonymity, pseudonymity, unlinkability and unobservability. All of these have already been evaluated and discussed in section 3.3 except the two first properties. These two properties will not be explained since they are not related to privacy but instead security and hence out of scope of this work.

3.7.4 PRIPARE

PRIPARE (Preparing Industry to Privacy by Design by supporting its Application in Research) [57] is a systematic methodology for privacy engineering that merges and connects existing best practices for PbD to help engineers to develop privacy-friendly software systems. The methodology combines the two major approaches used to discover and identify privacy requirements during the software life-cycle, which are the risk based and goal-oriented approach. These approaches are usually derived from privacy principles established by legal regulations aiming to support corporate policies.

PRIPARE's methodology tries to reduce privacy uncertainties at the initial phases of the development by eliciting the requirements through community-agreed catalogues, which corresponds to a goal-oriented approach. Second, a risk analysis approach is used to identify system-specific risks and the adequate treatment based on the risk level, among other factors. After applying both approaches, it is possible that some residual risks remain. These risks need to be identified and properly documented.

The methodology proposes the use of PIA (previously described under subsection 3.7.2) to identify potential risks and ensure legal compliance. When performing risk assessments, PRIPARE recommends an analysis based on a dual perspective, assessing the impact of a specific risk for both the organizations and the data subjects. To measure the risks, PRIPARE proposes different types of scale [58], [59]. However, according to the authors of PRIPARE, there are no specific benefits in choosing one scale over another. The analyst should decide which scale is best suitable in accordance with the system domain and the internal and external requirements. Moreover, PRIPARE states the importance of addressing the privacy issues (i.e, threats) identified during the risk management process, and find risk management decisions to mitigate them. These decisions include the avoidance, modification or reduction, sharing or transfer of the affected requirements. If there are any remaining risks, they must be communicated to the all stakeholders involved.

The goal-oriented approach includes a proposed catalogue of requirements that are heuristic, stakeholder-neutral, structured and hierarchized, prioritized and predefined. The idea is to enable the translation process from high-level privacy principles into operational requirements, to be systematic, and easy to follow by less privacy experienced engineers.

3.7.5 Privacy-Friendly System Design

The Privacy-Friendly System Design (PFSD) is a guidance in which from different stakeholder views are considered to understand what approaches are available, needed and how such approaches enforce privacy to software systems. Specifically, PFSD discuss three stakeholder views; the user sphere, recipient sphere and joint sphere. As the name implies, user sphere focuses at the individuals which using the system. The recipient sphere refers to the company or organization by which being responsible for providing the system. This involves backend infrastructure and data sharing networks. The third stakeholder view being parties which host data subjects information. Even though the recipient sphere could be responsible for providing a software system (service) the possibility exists that such recipients do not store the data at their own facilities. [60].

“Privacy-by-Policy” and “Privacy-by-Architecture” are two approaches introduced in PFSD. The former is explained as a strategy where an underlying approach is called notice and choice. The latter uses data minimization, anonymization techniques and evaluates processing and storage of data on client-side. To conceptualize, notice and choice is a set of principles which state the need for organizations to provide the possibility to halt processing and storing of data [61]. Additionally, the user has to have the control over its own data, how it is processed. On the other hand, if a company manages to create a software system, in which it does not need to collect sensitive and personal data in order to provide a service, no such notice and choice implementation needs to be applied to the architecture. Instead the privacy-by-architecture approach can be used to focus on the architecture containing sufficient privacy levels by not storing or processing such information. The authors of the PFSD methodology mainly discuss aspects of designing software with the privacy-by-architecture approach, namely network centricity and identification of data. Here network centricity implies to what parties, in a network infrastructure context, does the system rely on in order to provide the intended service. [60].

3.7.6 STRAP

Structured Analysis for Privacy is a goal-oriented methodology which, as same as Privacy-Friendly System Design, considers the FIP [61] principles as a central part for engineering privacy. By using STRAP, the architect is given four main steps to follow in order to understand the domain (the actors, goals and relevant components), privacy expectations of the stakeholders, vulnerabilities in the system and take appropriate counter measures to such vulnerabilities. The methodology’s four steps are; analysis, refinement, evaluation and iteration. During the analysis phase each goal for the system is asked a set of questions which aims at understanding involved information, actors, and usage of such information. From this the architect understands privacy vulnerabilities. Following, by knowing such vulnerabilities and the degree of risk for the user they are categorized into the FIP privacy categories [62]; Notice/awareness, Choise/Consent, Security/Integrity and Enforcement/Redress. The second phase involves addressing countermeasures to the found vulnerabilities in the analyzis phase. This can either be done through using PETS or by re-evaluating the goals and their sub-goals (as together form a goal-tree) to investigate if a modification of the goal-tree mitigates such vulnerabilities. Further, the authors of STRAP propose to create several alternative mitigation strategies for each vulnerability. This so a comparison can be performed later. Note, the authors of the methodology do not necessary refer mitigation strategies to PETs as explained in section 3.6.3. Instead the methodology advocates to revise the goal-tree, developed in the first phase, in order to eliminate potential threats. When such mitigation strategies have been performed the architect enters the third phase, the evaluation of the addressed mitigations from the design phase. During this phase the comparison of the alternative mitigiation strategies for each vulnerability is made where the strategy which enforces privacy to the highest extent is chosen. The extent of privacy enforcement of a countermeasure is the number of FIP principles (including sub-principles) such countermeasure satisfies. The last phase consists of backtracking the applied solution to the architecture. In other words, the architect do a second iteration of the earlier steps to ensure the applied solution indeed resolve all the found threats and in a correct manner and indeed enforce the privacy of the user as well as ensuring no new threats have arose during the steps. [63].

3.7.7 QTMM

This privacy threat methodology does, same as LINDDUN [8] and STRIDE [54], use a DFD to understand what different types of components exists in the system's architecture. By using such DFD the architect is given a Quantitative Threat Modelling Methodology (QTMM) where security threats in combination with privacy threats are elicited. The methodology consists of five steps; definition of DFD, mapping of security and privacy threats to DFD elements, identification of misuse case scenarios, risk-based quantification and creation of security and privacy requirements. [64].

The methodology considers six security and privacy properties; confidentiality, integrity, availability, unlinkability, transparency and intervenability. The three first properties being related to security and the remaining properties address privacy. Unlinkability has been explained in section 3.3. Transparency means any party involved in a processing of personal data can comprehend the legal, technical and organizational conditions. Finally, intervenability means the possibility by any involved party (including the data subject) to intervene with a processing of personal data. [64].

3.8 Privacy in the Automotive Domain

Vehicles today are not built the same way as they were twenty years ago. Back then the vehicle's main responsibility was narrowed to mechanical-centric matters, such as ensuring the cylinders kept being inflamed and thus make the vehicle moving in a forward direction. While such matters still being truly essential, today the responsibilities of a vehicle have become much widened. Today, they are equipped with an in-vehicle network of computational units that can process data [65]. By having such network within the vehicles the ability of them communicating externally with its surroundings exist. In other words, another external network is needed by which vehicles can communicate to each other. Such external network are usually referred to as Vehicle Ad-hoc Networks (VANET). The benefit with VANETs are several-fold where perhaps one of the biggest benefit being the increased safety insurance for the drivers [66]. By making vehicles being able to communicate to each other about possible accidents in the traffic it is possible to prevent such accidents happening in the first place. According to Hubaux et al. [67] such evolution within the automotive industry has been present since the early 21st century. Additionally, data generated from computational units inside a vehicle, such as event data recorder (EDR), GPS receivers and front-end radars to scan the distances of vehicles can be interchanged to other vehicles by using the communication VANTES provides. The EDR being the same type of 'black box' used in airplanes to record a majority of data of the vehicle. The GPS is used for location-based features and finally the front-end radar being used in collision-avoidance features.

3.8.1 VANET

As explained earlier, VANETs are the type of architectural pattern that are used by vehicles today which enables the possibility for Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I) communications. It enables the possibility for a vehicle to communicate with its surrounded vehicles and automotive relevant external entities. An example of a VANET can be seen in Figure 3.3
The advantages with having technologies using such architectures are many. Cooperative Collision Avoidance (CCA) is one example where a technology specifically using such V2X technologies in order to provide passenger safety [68]. In CCA, vehicles communicating with each other and inform vehicles behind itself about an accident. Imagine the scenario where a driver would not be able to foresee such accident, due too many cars being front. This would lead to a disaster at places where high speed limits are allowed, such as the highway.

Despite all the benefits VANETs provides there prevails important concerns that need to be addressed, as for all software systems dealing with sensitive user information. According to Hubaux et. al, due to the degree of sensitivity of the data which surge through the network there is a need for effective protection of the driver's privacy [67]. Protecting privacy in automotive systems is indeed a hassle. The type of data is different than in ordinary software systems. Due to the dynamic data and heavy use of location based data which is generated by the vehicles during

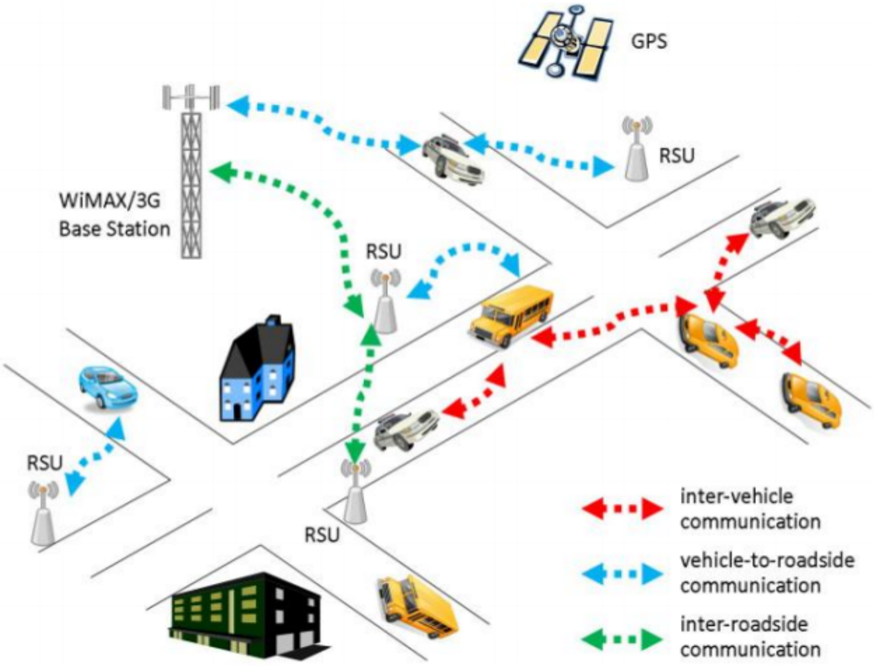

GPS
RSU
WiMAX/3G
Base Station
RSU
RSU
RSU
RSU
inter-vehicle communication
vehicle-to-roadside communication
inter-roadside communication

Figure 3.3: Overview of a VANET taken from [2].

driving, challenges such as data trust, how much and what data should be stored in the vehicle.

Current research already exists which elaborates key aspects on privacy enhancing requirements in automotive systems. Schaub et al. discuss five key privacy requirements that can serve as a guidance and by which needs to be fulfilled in order to reach appropriate privacy levels in automotive systems [69]. The first three requirements have previously already been discussed in section 3.3, they are minimum disclosure, anonymity and unlinkability. Here minimum disclosure refers to the concept of data minimization. The two last requirements being distributed resolution authority, perfect forward privacy.

Distributed resolution authority is explained as the concept of a cooperation of a number of distinct authorities required to link an anonymous credential to an entity (vehicle). They refer to the data management technique resolution identity which helps ensuring the impact to privacy when several authorities split responsibility of the identification of a vehicle, thus no authority can misuse the resolution information. According to Schaub et al. this can be implemented through policies and regulations but should be implemented technically as well.

Perfect forward privacy is a requirement that as well as the previous requirement relates to identity resolution. Schaub et al. mean that it is a requirement where one anonymous credential only enables linking of messages sent under this credential, but no information is provided about other credentials held by the same user [69].

4
Iteration 1 - LINDDUN versus PIA

As previously explained, due to limited resources the focus of this study was on the two methodologies: LINDDUN and PIA, which are described in section 3.7.2. Again, the reason for choosing the two methodologies, instead of other privacy methods presented in Section 3.7, is out of convenience, and also because they are well regarded in the privacy community. The analysis of LINDDUN and PIA is divided in two parts: 1) LINDDUN and PIA comparison and 2) an empirical evaluation of the two methodologies. The first part was used to map and identify which provisions from GDPR were covered by each methodology, while in the second part, a threat analysis session was performed for both methodologies on a case within the automotive domain provided by Volvo Group Trucks Technology.

4.1 LINDDUN and PIA Comparison

During the literature review different methodologies used to analyze privacy threats were identified, as presented in section 3.7. However, the GDPR suggests that when handling PII organizations should perform a PIA/DPIA to ensure compliance to the regulations [17]. This is also another reason why PIA was chosen, instead of other methodologies, in the comparison with LINDDUN. The aim of the comparison was to investigate the coverage provided by both methodologies in terms of compliance with the GDPR principles and user rights. Thus, a mapping table relating LINDDUN and PIA to the GDPR was created to establish a base line for further research. The mapping table is a conceptual effort, created after a close analysis of the documentation available for both methodologies and the regulation, to map LINDDUN and PIA to the provisions presented by GDPR.

4.1.1 LINDDUN and PIA Mapping Table

In order to identify possible LINDDUN and PIA limitations in terms of compliance to GDPR, a systematic comparison of the two methodologies was performed. Initially, all GDPR main principles and user rights presented under section 3.5 were extracted. The extracted principles and user rights are listed in Table 4.1. After extracting the GDPR principles and user rights, each methodology was analyzed individually in order to identify to what extend they cover the provisions from the GDPR. For LINDDUN, the Unawareness (U) and Policy/Non-compliance (NC) threat trees were included. These threat trees were considered to be the only threat trees of the methodology aiming to address legislation policies and to provide user awareness [8]. Thus, these are the only threat trees from LINDDUN used in the analysis. For the PIA, the material from the PIA guidelines defined by Oetzel et al. [10] was used to identify the principles and rights addressed by the methodology. The PIA guidelines consists of a set of predefined privacy targets and threats that aims at providing the practitioner with a robust material for finding relevant threats in the software and hardware domain. One can see these guidelines as a set of screening questions that one can follow in order to feel comfortable in knowing that all relevant privacy threats are found. These defined privacy targets and threats, defined by Oetzel et al., was thus manually analyzed to understand if they had any correlation to a certain principle or user right presented in Table 4.1.

Table 4.1 presents an overview of the GDPR principles and user rights used for the mapping and comparison of the two methodologies. In total, 14 principles and rights were included. Some of these principles and user rights were expanded into sub-principles to accomplish a more explicit mapping. This, since some of the principles are broader than others and hence harder to relate a certain threat to. In other words, these sub-principles were used to provide a more clear definition of what is actually required in order to become compliant with the regulation. If one of these sub-principles were not identified in the methodologies, the main principle or user right was considered as not covered. Overall, LINDDUN was mapped to 2 principles, with a percentage rate of 14 percent coverage, while PIA was related to 4 principles and 6 rights, with 71 percent coverage.

After mapping the principles and user rights with the two methodologies, it was possible to identify which of the principles and user rights from the GDPR were not covered by LINDDUN. A more detailed description of the gaps found in the methodology are presented below:

- **Lawfulness, fairness and transparency:** this principle was divided into four sub-principles: 1.1 User consent for processing personal data, 1.2 User consent for sharing data to third parties, 1.3 User is able to view/update/withdraw consent, and 1.4 Lawful processing due child's consent. The principle is presented by LINDDUN as a high level goal under the content unawareness and policy and consent non-compliance threat trees. However, it does not make any reference to user consent for sharing data to third parties, neither to user being able to view/update/withdraw consent, or to child's consent.

  PIA did define threats for consent for processing as well as sharing personal data to third parties. Regarding the sub-principle 1.3, PIA guidelines do cover threats related to a user is unable to view consent. However, the provision to provide a user the possibility to update and withdraw a consent was not addressed. For this reason the sub-principle was considered not covered. Regarding consents for children being the age of 15 or below, the sub-principle 1.4, the PIA did not address this at all. This is not surprising, since this is

Table 4.1: LINDDUN versus PIA (GDPR mapping table).

GDPR Principles and User Rights	LINDDUN	PIA
1. Lawfulness, fairness and transparency	No	No
1.1 User consent for processing personal data	x	x
1.2 User consent for sharing data to third parties	-	x
1.3 User is able to view/update/withdraw consent	-	-
1.4 Lawful processing due child's consent	-	-
2. Purpose	No	Yes
2.1 Data must be collected for specified, explicit and legitimate purposes	-	x
2.2 Collected data for one purpose is not used for another purpose	-	x
3. Adequacy and data minimization	No	Yes
3.1 Data is limited to what is necessary to the purpose	-	x
4. Accurate and up-to-date processing	Yes	Yes
5. Storage limitation and retention time	No	Yes
6. Accountability	No	No
6.1 Notify data subject about data breach	-	-
6.2 Notify supervisory authority about data breach	-	x
6.3 Perform logs	-	x
7. Right to be informed	Yes	Yes
7.1 Inform the user how data is processed/stored	x	x
7.2 Inform the user how/where data is disseminated	x	x
8. Right of access	No	Yes
9. Right to rectification	No	Yes
10. Right to erasure	No	Yes
11. Right to restrict processing	No	No
12. Right to data portability	No	No
13. Right to object	No	Yes
14. Rights related to automated decisions	No	Yes

one of the newly introduced provisions of GDPR compared to the previous EU data protection directive. Since, two of the sub-principles were not covered by PIA the lawfulness, fairness and transparency principle is considered not covered for the PIA methodology.

- **Purpose:** includes two sub-principles 2.1 Data must be collected for specific, explicit and legitimate purposes, and 2.2 Collected data for one purpose is not used for another purpose (purpose limitation). In LINDDUN there is no reference regarding purpose limitation, neither under the threat trees nor the provided PETs. The PIA guideline does define privacy targets, threats and solutions to counter both sub-principles. Hence, PIA is considered to cover the purpose principle.

- **Adequacy and data minimization:** this principle includes the sub-principle 3.1 Data is limited to what is necessary to the purpose. LINDDUN makes a reference for data minimization under the linkability and identifiability threat tree. However, LINDDUN does not provide any PET to address data minimization directly. Privacy targets, threats and countermeasures which address the concern for adequacy and data minimization was defined in the PIA. This was considered enough to classify PIA as covering the third principle of GDPR.

- **Accurate and up-to-date processing:** LINDDUN makes a reference to data accuracy under the unawareness threat tree. It also provides a mitigation strategy to increase accuracy. The strategy includes the possibility to review, update, and/or delete data. The PIA refer threats related to Accuracy and up-to-date processing as data quality. The guidelines do define targets and threats which relates to violation of quality assurance of data.

- **Storage limitation and retention time:** LINDDUN provides a reference for storage limitation and retention time under the Linkability of a data store threat tree. It shows possible threats that can occur when storing data for too long and when storing too much data (data minimization). As previously explained, this threat tree is not considered to be included in this work, as the aim of the tree is not to comply with regulations. Therefore, it is considered that LINDDUN does not cover this principle. PIA does include two threats which defines the threat of inability for deleting data either due to missing erasure policies or retention rules.

- **Accountability:** is divided into the following sub-principles 6.1 Notify data subject about data breach, 6.2 Notify supervisory authority about data breach, and 6.3 Perform logs. LINDDUN does not provide any direct reference to accountability. Instead, it addresses accountability through the transparency principle, which is included under the Content Unawareness threat tree. However, this is considered to be insufficient, since it does not address any threat related to the new regulation.

PIA does bring up the concern for notifying supervisory authorities upon data breach and other related matters. However, nothing is addressed for notifying the affected data subject. This, most probably due to the recent change in the regulation valid since May 2018. In PIA, there are defined threats for not performing sufficient amount of logging. However, those threats are not directly tied to an accountability target (privacy principle). Since not all sub-principles were addressed the accountability is considered not sufficiently covered in PIA.

- **Right to be informed:** includes the sub-principles 7.1 Inform the user how data is processed or stored and 7.2 Inform the user how or where data is disseminated. LINDDUN addresses this right under the policy and non-compliance threat tree, where it refers to treats related to insufficient policies, which in turn can be related to the right to be informed. One of the most defined threats in PIA is related to inform the user regarding how data is processed, stored and disseminated to other third parties. Hence, both sub-principles are addressed and consequently the right to be informed as well.

- **Right to to access:** LINDDUN makes a reference to this right under the Unawareness threat tree, including a leaf node that refers to "Unable to review personal information". However, this does not explicitly mention whether it refers to only reviewing the data or if this includes physical access to the data. Under the Non-repudiation of a data store threat tree, there is a reference to a person not being able to remove/alter all data. However, as has been explained earlier in this section, only the content unawareness and policy and consent non-compliance threat tree is considered in this work. Therefore, this will not be included in the mapping of the methodology. The threats relating to this principle is referred to as inability to provide information about processed data and purpose in the PIA guidelines. The threats involves providing an interface, or through other means, for the individuals where they can identify what data about him or her is processed.

- **Right to rectification:** LINDDUN does not provide any reference under the threat trees, tree node, or leaf node which can be related to the GDPR right to rectification. Thus, the right to rectification was considered not to be addressed by the methodology. The privacy target in the PIA which relates threats for this principle is called "Inability to rectify, erase or block individual data" and does have supplemented defined threats, explaining the concern for user rectification, automatically rectification of error and similar correlated to it. Thus, this right is considered covered by PIA.

- **Right to erasure:** there is no direct mention regarding the right to erasure under the threat trees. However, under the Non-repudiation of a data store threat tree there is a reference to a threat when the user is unable to remove its own data. This reference was considered to be insufficient, thus, it was considered that the methodology does not address the right to erasure. As for PIA, the same privacy target as mentioned for the previous user right, contains threats that relate to this user right. These threats are defined as the non-existent, technical or process, and means for erasing data about a data subject in the system.

- **Right to restrict processing:** no mention was found in the LINDDUN methodology regarding the right to restrict processing, neither at the threat trees nor under the PETs. No defined privacy target, threat or countermeasure with an explicit correlation to this right is existing in the PIA. Hence, this user right is considered not covered by the PIA.

- **Right to data portability:** there is no reference to the GDPR right to data portability under the LINDDUN methodology, neither in the threat trees nor under the PETs. The PIA does not address any threats or other means as an indication for addressing this user right. Just as for some of the previously discussed principles this principle is newly introduced with the GDPR. Hence, this was expected.

- **Right to object:** no mention was found regarding the right to object in LINDDUN. Not under the threat trees or under the PETs. The PIA does include defined threats to the user right to object on processing. In PIA threats which relate to this user right includes the concrete implementation, technically or process, of accomplish a user request for objecting on his or her data being processed. Hence, this user right is considered to be addressed by PIA.

- **Rights related to automated decision making and profiling:** no mention was found regarding the rights related to automated decision making and profiling in LINDDUN. Neither under the threat trees nor under the provided PETs. PIA does not include many threats relating to the user's right for automated decision making and profiling. Although, a threat is defined for the inability of a user to object on such actions taking place in a system. Even though one defined threat can be considered limited, it does address the the user right and hence makes the PIA cover this user right.

It is important to mention that some of these principles and rights can also be found under the other LINDDUN threat trees, as described above. However, in order to provide awareness of the user and compliance with regulations and legislations, it is believed that these principles and rights should be addressed under the Unawareness (U) and Policy/Non-compliance (NC) threat trees.

4.2 Empirical evaluation of LINDDUN and PIA

4.2.1 Case Study

The case where LINDDUN and the PIA analysis were performed was provided by Volvo Group Trucks Technology. It represents a truck platooning system, which utilizes V2V communication to share private data and communicate among other vehicles equipped with the same system, a similar system can seen be in [70]. The platooning system is an advanced version of an Adaptive Cruise Control (ACC) and a Cooperative Adaptive Cruise Control (CACC) system [71]. One of the purposes of the system is to reduce the fuel consumption of the vehicles, just as described in [72]. A simplistic explanation of the general functionality of the platooning system is shown in Figure 4.1.

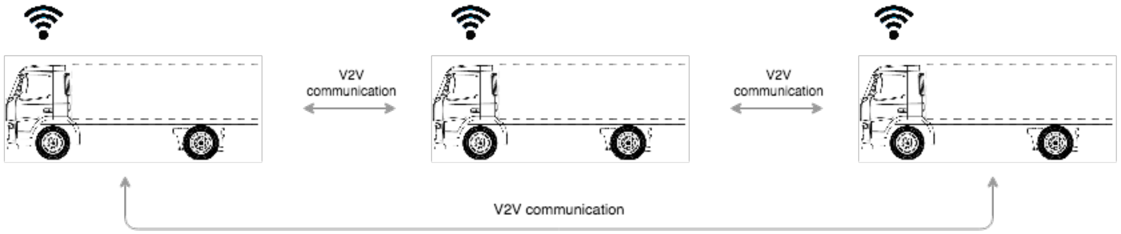


Figure 4.1: Overview of the platooning system's functionality

A system description of the main functionality of the system was provided to the authors, together with some use cases and information regarding the components of the system. From the provided document four use cases were extracted and a component diagram was created. This component diagram was used as starting point for both of the analyses. However, the system is still in its early stages, therefore due to confidentiality polices it was not possible to include additional information regarding the case in the study, for instance certain use cases used to model the system or the architectural views.

4.2.2 LINDDUN and PIA analysis

The LINDDUN and PIA analysis were performed by the authors of this study in two different, independent and isolated sessions. Each author was assigned to one method and performed the analysis without communicating or interacting with the other author. As previously mentioned, the analyses were performed to identify how much coverage both methodologies provide in accordance to the GDPR principles and user rights. Each of the authors were responsible for performing one of the methodologies, and the sessions were performed in an isolated environment without external interference. As stated earlier in subsection 4.2.1, for the analyses a system description of the system, together with the four use cases, and a modeled component diagram were provided. The sessions are fully described below.

During the LINDDUN session a DFD was created from the provided component diagram and use cases description. Only the three first steps of the methodology were addressed, together with the two last threat trees: Content Unawareness and Policy and Non-compliance. As explained in section 4.1.1 , the reason why to only focus on these threat trees is due to the nature of the trees. It is believed that these trees are used to provide awareness for the users and help to provide compliance with legislation policies. Hence, the provisions of the GDPR should be included in these trees. In the first step a DFD was created based on the documentation provided; in the second step the elements of the DFD were mapped with the provided LINDDUN mapping table; and in the last step, the threats identified with the help of the threat tree catalogs were documened using a misuse template provided by LINDDUN [1]. Moreover, the assumptions were also documented.

For the PIA session, the same documentation provided for the LINDDUN analysis was used. The analysis followed the PIA Framework presented by Oetzel et al. [10]. Their work includes a template used to access a Radio-frequency identification (RFID) application in an automotive context. This template was extracted from the paper and was used as guideline for the analysis of the platooning system. The template contained six steps that aimed to discuss and identify privacy targets and its corresponding threats.

All the results obtained during the LINDDUN and PIA analyses are presented in the following section.

4.3 Results

In total, 7 threats were identified with the LINDDUN methodology, while 44 were identified with the PIA analysis. In order to calculate the coverage provided by the two models in relation to the GDPR principles and user rights, the threats were mapped in a systematic fashion with the principles and user rights presented in Table 3.1.

The Figure 4.2 demonstrates the amount of threats considered to have any relation to the GDPR. As presented in the figure, from the 7 threats identified with LINDDUN only two were considered to be related to the regulation. As for the PIA methodology, 44 threats were identified. These 44 threats were mapped to 11 principles from the GDPR.

The approach used to consider if a principle or user right was fully addressed by LINDDUN and PIA through the threats found, was the same approach used during the mapping table. A principle or user right containing sub-principles was only considered to be addressed if all of its sub-principles were also addressed. For instance, during the PIA analysis the principle number 6, Accountability, as seen in Figure 4.2, was associated to 5 threats. However, the threats found were related to the sub-principles 6.2 and 6.3, thus the accountability principle was considered not fully covered by PIA.

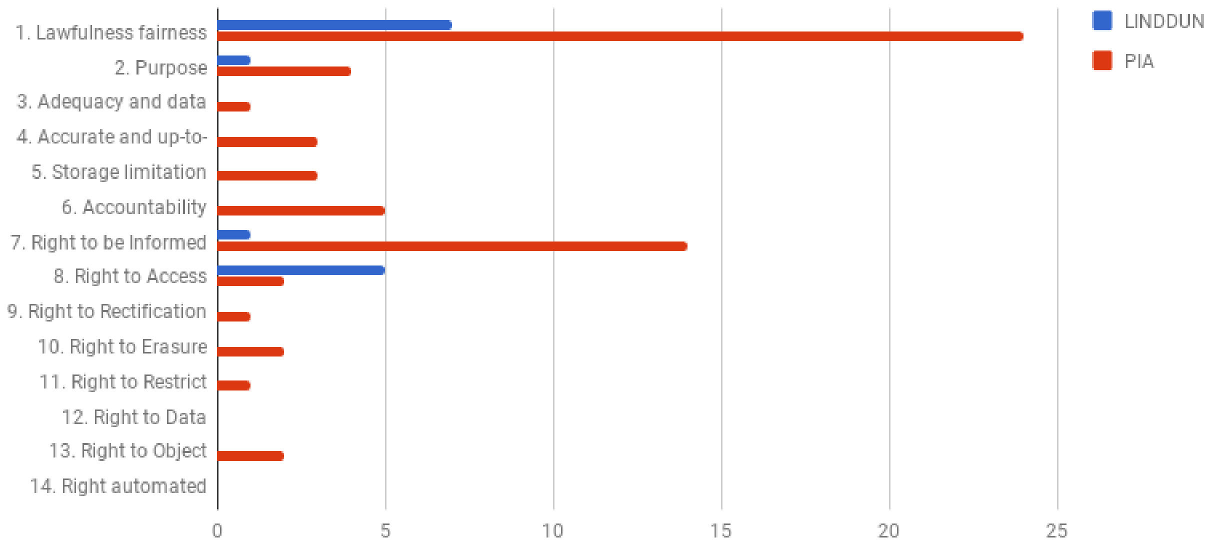


Figure 4.2: Number of threats found for each GDPR principle or user right

The result gathered from the LINDDUN and PIA analyses were very much alike the result obtained during the mapping table. This confirmed the need for improvements of the LINDDUN methodology and further analyse explicitly which these areas are to be improved. An overview of the results can be seen in Table 6.3.

Table 4.2: Overview of the Results from the Mapping and empirical evaluation of the methodologies.

	Conceptual Mapping		**Empirical Analysis**	
	LINDDUN	**PIA**	**LINDDUN**	**PIA**
Covered principles	14%	71%	14%	71%
No of threats found in the platooning case	-	-	7	44

Overall the results show that PIA can provide a better coverage of the principles in comparison with LINDDUN. Therefore, the motivation to propose changes to the LINDDUN framework. In Chapter 4, the gaps found during this first iteration were addressed in an attempt to develop an improved version of the LINDDUN methodology, with a better coverage of the regulation.

5
Iteration 2 - LINDDUN+

LINDDUN and PIA are not the only methodologies used to elicit privacy threats for software systems, as seen in section 3. Earlier, it was explained that due to convenience, and due to limited resources, only LINDDUN and PIA were addressed in this study. Furthermore, PIA is cited by GDPR as a way to help organizations address PbD and demonstrate compliance when required. The results presented in section 4.1 and 4.2 showed that PIA already provides a better coverage in identifying privacy issues related to GDPR compliance in comparison with LINDDUN. Also, LINDDUN contains a different approach compared to PIA, as it is a robust step-by-step strategy that allows practitioners to perform privacy analysis from a data flow perspective. By understanding how data flows in a system one can easier, and to a much higher extent, find privacy threats at specific parts of the software system. Moreover, the identification of privacy threats can be accomplished with the help of the catalog of threat trees provided by LINDDUN. Therefore, the motivation for the proposal to extend LINDDUN. It would be of great interest to investigate the possibility to address the identified limitations of the methodology and enhance its effectiveness in identifying design issues related to GDPR, and thus increase its coverage rate in comparison with PIA.

5.1 Perceived limits of LINDDUN

By looking at the results in section 4.1 and 4.2, from the comparisons made during the first iteration, it is clear that LINDDUN suffers various improvement points regarding compliance to the GDPR principles and user rights. Specifically, looking at the Table 4.1 the reader can understand that the methodology fails to support seven out of the fourteen GDPR principles and user rights. Nevertheless, in order to establish compliance, organizations are required to reach these principles and rights. Consequently, the principles and user rights must be treated just as any other functional or quality requirement that a system has to fulfill.

From the literature review, it was understood that the principles and user rights are all threats that are related to LINDDUN's policy and consent non-compliance threat trees. As it was also understood that the possibility for a practitioner of LINDDUN to address and assess the effects of privacy threats in a design solution does not solely depend on the threat trees. Therefore, after completing the literature review and the comparison between LINDDUN and PIA, it was perceived that the business-oriented DFD did not provide enough details and information that could be used to help practitioners assess the likelihood of a possible threat to occur.

Another perception, experienced by the authors after completing the analysis session of LINDDUN in the first iteration, was that the methodology is time consuming. This due to all redundant time of performing administrative tasks, such as analyzing threats through the threat trees, documenting assumptions as well as documenting applicable threats. Moreover, the LINDDUN methodology can be hard to learn for an inexperienced privacy engineer, e.g. realizing relevant threats or assumptions. This could lead to, even though the methodology does cover a certain threat that the practitioner is unable to successfully find the threat.

5.2 LINDDUN+ at a glance

LINDDUN+ extends LINDDUN in three ways: 1) with a PA-DFD, 2) the Threat Trees Extensions, and 3) the Threat Trees Rules. The PA-DFD provides the possibility to create a business-oriented DFD with a more privacy aware perspective and is fully explained in section 5.3. The threat tree extensions extend the Content Unawareness and Policy and Consent Non-Compliance threat trees and are described in section 5.4.1 and 5.4.2. Finally, the threat tree rules represent an attempt to demonstrate the possibility to automate the methodology and are explained in section 5.5.

The desired outcome expected to be achieved with LINDDUN+ was:

- Allow the practitioner to better understand where in the architecture the most privacy sensitive information is flowing.
- Provide a more complete coverage of the design issues related to GDPR compliance.
- Take an initial step towards automation of the LINDDUN methodology

To cover the desired outcome the following areas have been in focus:

1. **The PA-DFD:** the PA-DFD, an extended version of the business-oriented DFD, in LINDDUN+ is used to fulfill two purposes. The first purpose is to provide the practitioner with a more detailed and privacy-aware DFD. Although the business-oriented DFD provides a complete view of how data flows in the system, the practitioner can be unaware of the most sensitive parts of the system, i.e. the flows and processes which handles PII. With a PA-DFD the practitioner can identify which parts in the architecture require a more close analysis. The second purpose of the PA-DFD is related to the development of the threat trees rules used to demonstrate the possibility of automation of the methodology. For this, the extensions proposed by Antignac et al. [12], [13] were reused, resulting in the creation of a meta-model that is further explained in Section 5.2.2.

2. **LINDDUN+ Threat Trees:** threat trees are not meant to showcase all possible attack paths to a software system, but instead demonstrate common state-of-art attack patterns. Although, it was clear to the authors of this study that it was not possible for the inexperienced architect to understand when a certain threat from the threat tree is present in the architecture or not. Further, the creators of the LINDDUN methodology expressed the need to further develop the catalog to include more threats in the future [8]. The proposal for extending the threat trees consists of revising the already included threat tree nodes and including new threats to increase completeness with regards the GDPR. The extended threat trees were the Content Unawareness and Policy and Non Compliance. They were extended for the same reason they were included in the analysis of LINDDUN versus PIA, as they are used to address user awareness and compliance with regulations. The extended threat trees are further explained in section 5.4.1 and 5.4.2.

3. **LINDDUN+ Threat Tree Rules:** from the literature review and after the first iteration it was observed that LINDDUN methodology is time consuming and might be considered hard to apply by an inexperienced practitioner. A belief is, an alternative to mitigate the struggle for a privacy engineer of applying LINDDUN is to, in the future, develop an automated tool. The automated tool could help the inexperienced privacy engineer by suggesting where in the architecture a threat can occur, and speed up the time spend to perform a full session of LINDDUN. Although the scope of this work is not to develop such automated tool. Instead it can be seen as initial work to demonstrate the possibility of developing an automated tool. Therefore, a proposal for creating rules which can explain, where in an architectural design, a given threat is applicable. Moreover, the rules can serve a second purpose. They can help practitioners to manually, by reading the rules, understand where such threats are present in an architecture. Hence, the believed benefit for the methodology of an increased coverage of the GDPR provisions. The rules are explained more in detail in section 5.5.

5.3 The PA-DFD

As previously explained, the first goal of the PA-DFD is to help practitioners become aware of the privacy sensitive areas of a DFD. Also, the goal is to let the practitioners to reflect about which countermeasures are assumed to be implemented in the system, by default, in order to mitigate possible threats before going further into the analysis. The assessment of the countermeasures can also be done with the help of stakeholders of the system, such as domain experts. The creation of a PA-DFD include the following steps:

1. Create a business-oriented DFD.
2. Specify, using annotation nodes, what data is passed, stored or processed at each DFD element.
3. Include in the nodes the countermeasures, or the defense mechanisms, used

in the system. This step might require an interaction between practitioners and stakeholders to identify what will be addressed in the system before the LINDDUN+ analysis.

4. Use a data classification specification to compare which processes, data flows or data stores contain personal sensitive information.
5. Mark which nodes contain personal sensitive information.

A simple business-oriented DFD of a V2I application, that sends data from a vehicle to its back-office is shown in 5.1. The DFD presented is a very high-level description of an actual system and is provided for the purpose of helping the reader understand the implied work of creating a PA-DFD. A vehicle is represented as an external entity (1.Vehicle) that sends data to a data handler (2.Data Handler), represented as a process. The database handler (3.DB Handler) is also a process that sends private data to be stored in the data store (4.DB).

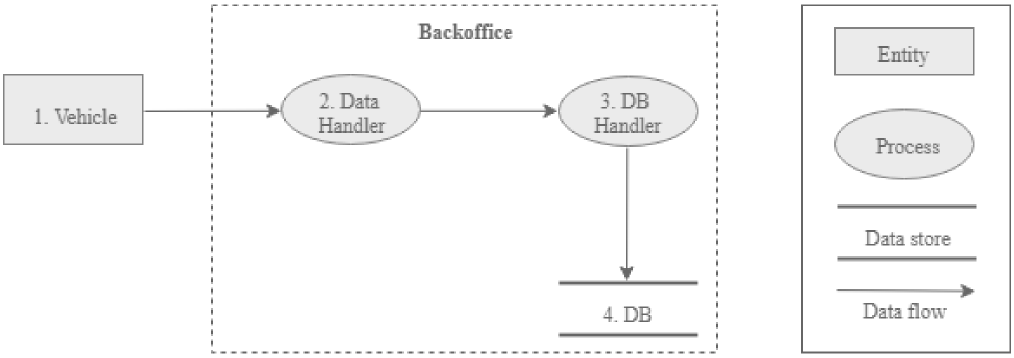


Figure 5.1: Overview business-oriented DFD

Prior to the work for creating a PA-DFD it is assumed the practitioner has a business-oriented DFD as well as an understanding of what are the sensitive data attributes for the specific domain. The latter can be achieved by having a data classification specification at hand. The architect will identify which flows contain PII data by marking notes at each DFD element. The notes explain what data attributes are essential to be contained in the specific DFD element, as well as assumed implemented privacy aware techniques which mitigate potential privacy threats. Here, the introduced privacy-aware elements, explained in section 5.3.1, can help the engineer to reveal such privacy aware techniques. The Figure 5.2 shows a potential business-oriented DFD with attached notes and assumptions. As the figure shows, each note contains two parts; the data attributes used by the DFD element (Data) as well as the assumed privacy mitigation technique (Countermeasure) implemented. Looking at the data base handler it is the DFD type of a *Log* process, due to logs are performed for the data that is processed. Further, the data store is considered to be of a type *DataStoreLog* and an ordinary *DataStore*. The remaining DFD elements are ordinary element types such as External Entity, Process, Data Store or Data Flow.

After all DFD elements have been marked, the engineer has to understand which privacy sensitive areas are present in the DFD. This can be achieved by using the data classification specification to evaluate if a certain DFD element matches the condition for PII. Explicitly, the information specified in the note for each element is compared to the PII definition in the data classification specification. If the data is considered PII the practitioner understands an increased privacy-aware mindset for that specific part of the system is needed. The architect then marks the note to understand if this needs to be further analyzed to enforce privacy. It will help the practitioner to later reveal privacy threats that the practitioner would not be able to find otherwise. When all notes have been compared to the definition of PII the DFD is considered to be privacy aware.

It is known that in all architectural system designs the architect has to prioritize the most important requirements. This due to various reasons, such as economical cost, performance overhead and customer needs. The same exists for an architect practicing LINDDUN. Consequently, it will in most cases not be possible to develop a PA-DFD which at all parts are completely privacy aware. Instead the architect needs to be aware of the most important parts (where the most sensitive data is flowing, processed and stored). Therefore, a proposal for changing the initial steps of the methodology and guide a practitioner of the LINDDUN methodology in the creation of a PA-DFD.

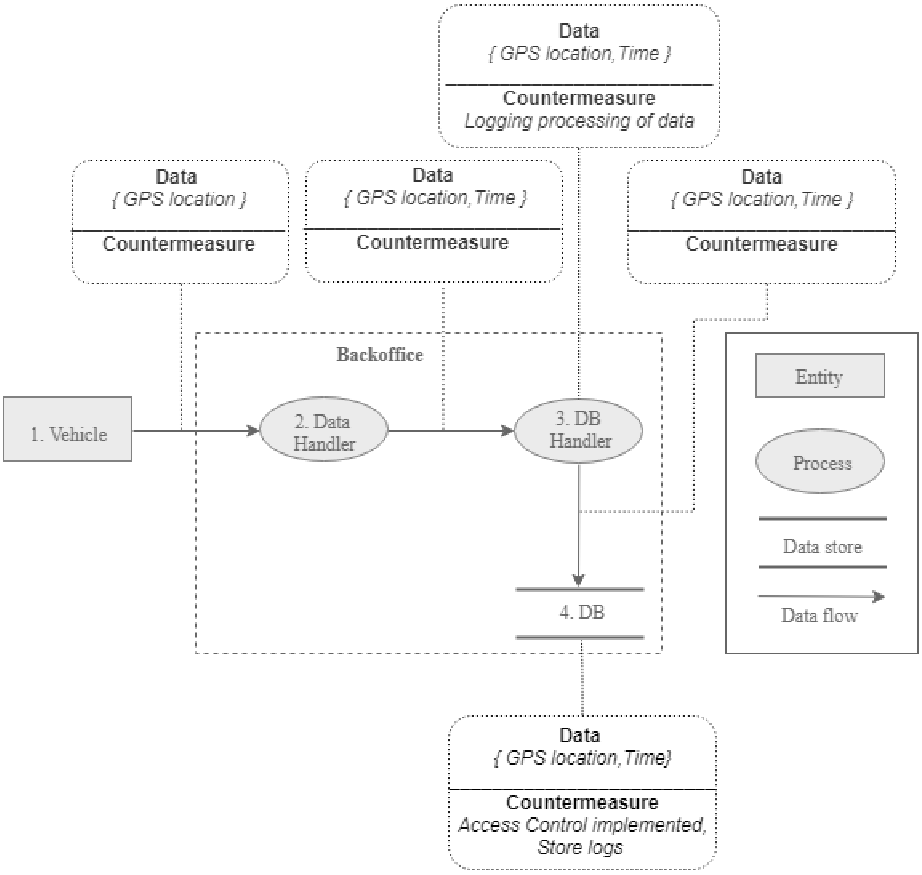


Figure 5.2: The PA-DFD with notes of data flowing in the system

Currently, the DFD used by LINDDUN considers four element types: *External Entity*, *Data Flow*, *Data Store* and *Process*. In order to address the privacy-aware elements included by Antignac et al. [13], in their work for developing a PA-DFD, a new meta model was created for the DFD. The development of the meta model is fully explained in the following section. The meta model is used to model all of the privacy-aware components and their relations in a PA-DFD. In total, the PA-DFD can use eleven element types.

5.3.1 The PA-DFD Meta Model

As explained earlier in section 5.2, one of the proposals for extending LINDDUN is to elicit threats from a PA-DFD instead of a business-oriented DFD. In order to define which elements would represent a PA-DFD, a meta model was created. Antignac et al. [12], [13] developed a model-to-model transformation which takes a business-oriented DFD and transforms it to a PA-DFD. Their work focuses at finding hotspots in a given business-oriented DFD and transforming the hotspots into a more privacy aware architectural solution. The transformations include architectural elements that are used to promote compliance with the legislation. Inspiration has been taken from the work from Antignac et al. Specifically, the new DFD element types introduced in their work. Thus, the Meta Model considers more element types than the four element types of the ordinary business-oriented DFD. The DFD elements that have been included in the Meta Model are; *Limit* process, *Request* process, *Reason* process, *Log* process, *DataStoreLog*, *DataStorePolicy*, ordinary *DataStore* and *DataFlowErase* as well as the four ordinary elements from the business-oriented DFD. The purpose for including more elements is to help practitioners become aware of other privacy related scenarios in the software design. In other words, to evaluate if including the privacy-aware elements would significantly benefit the LINDDUN methodology to become more compliant with the GDPR. The second purpose for the creation of the Meta Model was to demonstrate that by including these privacy aware elements, it would be possible to create rules for the nodes in the extended threat trees. As explained earlier, the rules are used to showcase the possibility to create a tool that in an automated way can identify possible threats in a software architecture. These rules are explained further in section 5.5.

Figure 5.3 presents the meta model created over the privacy-aware DFD. The root node of the meta model is a DFDDiagram. This node can contain one of more DFDNodes. A DFDNode consists of two attributes, isPII and a Name. The former is a boolean attribute and aims to specify if the node does in any form handle data classified as personally identifiable. The latter attribute is a string. It is used to specify the name of the node, for the purpose of increasing the comprehension of a modelled system. A DFDNode can be any type of the following nodes; DataFlow-Component or DestComponent.

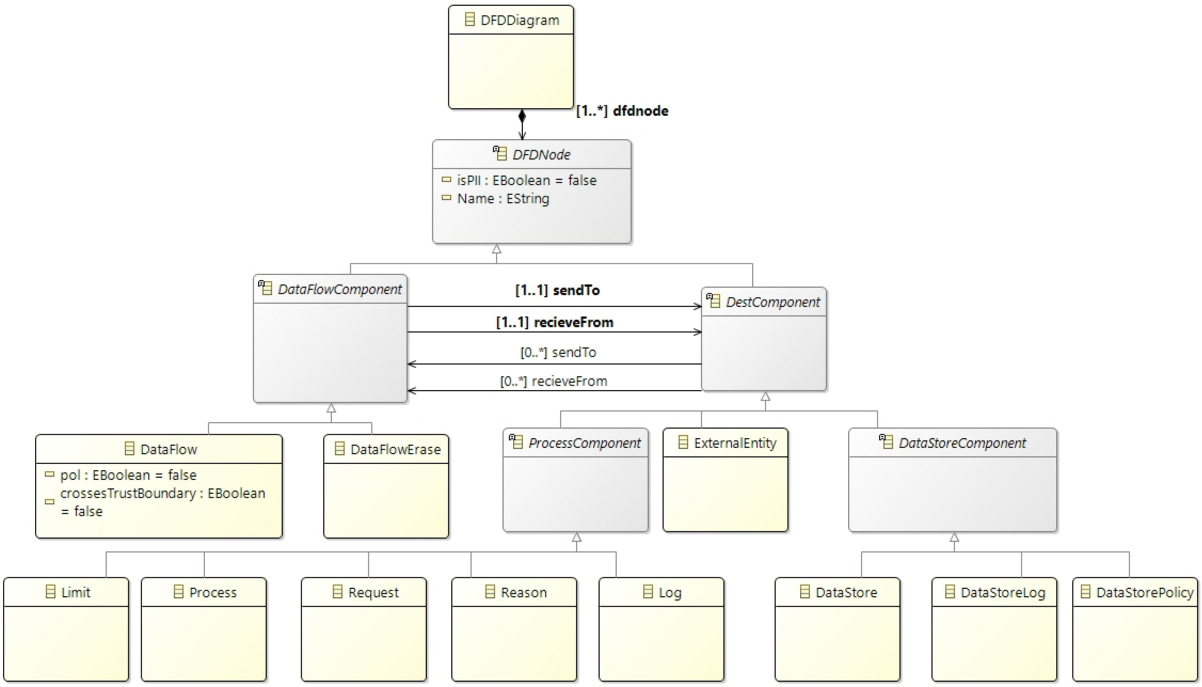


Figure 5.3: Meta Model over the privacy-aware DFD.

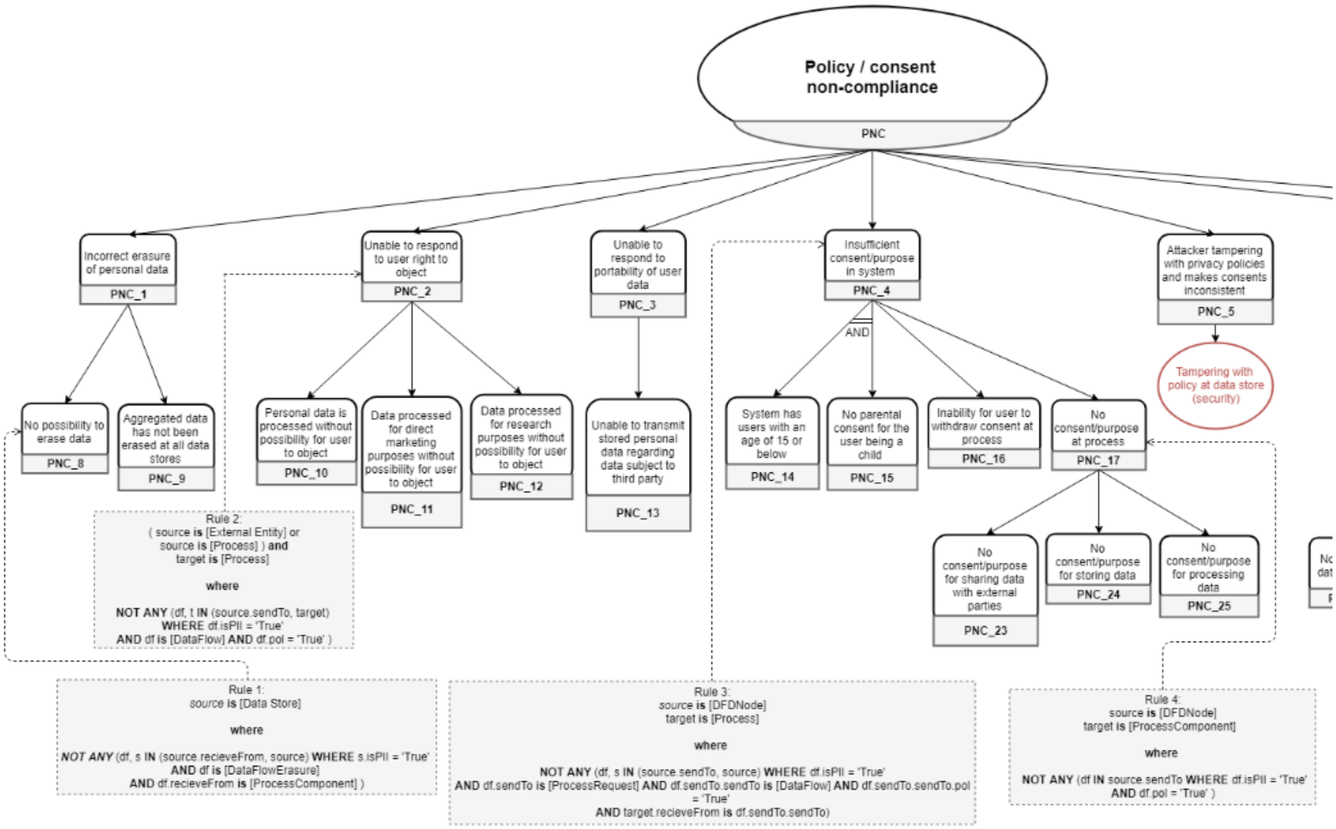
A DestComponent can be any of the following type of element:; ProcessComponent *ExternalEntity* DataStoreComponent. Here a ProcessComponent can be any type of the following five types; *Limit*, *Process*, *Request*, *Reason*, *Log*. A *Limit* process is any process in a DFD which performs any data minimization technique. Additionally, it receives a consent which is evaluated in order to understand if the data subject does agree to the processing or not. A *Request* process is a process which solely does receive a consent and forwards it to another element. A *Reason* process does, just as the *Request* process, receive a consent. However, the purpose of the *Reason* process is to update the value of a certain consent which is stored in the system. Hence, the *Reason* process forwards the consent to be updated to a DataStorePolicy. A *Log* process performs the work of logging what data has been processed at a certain ProcessComponent. The practitioner of LINDDUN should consider the *Log* process to log not only what data is processed but also what consent has been considered at the processing. A *Process* processes data that does not fit into the purposes or contexts of the previously explained ProcessComponents. In other words, it is an element which process data without considering a consent, not logging what data is under processing or does not forward or update a consent.

A DataStoreComponent can be any of the following types: *DataStoreLog*, *DataStorePolicy* or *DataStore*. The first element type serves the purpose of storing the information provided by a *Log* process. The second element serves the purpose of storing consents about the system users as well as the policies which are used by the system in order to provide compliance to the regulation. The last element, the *DataStore* element, is any data store which does not share the same purpose as the two previously explained DataStoreComponents. Hence, the reader can relate this to an ordinary data store, the same type which is defined in the business-oriented DFD.

A DataFlowComponent is either an ordinary *DataFlow* or a *DataFlowErase*. The DataFlow contains two attributes: pol and crossessTrustBoundary. Both attributes are of type boolean. The former aims to explain if the *DataFlow* does contain a policy or a consent. The latter, as the name implies, explains if a DataFlow crosses a trust boundary or not. A *DataFlowErase* provides the architect the possibility to design software solutions that can comply with the right to erasure. Hence, it aims to model the scenario of erasing data at a given reference in a DataStoreComponent.

5.4 LINDDUN+ Threat Trees

5.4.1 Policy and Consent Non-Compliance Tree



Policy / consent non-compliance
PNC
Incorrect erasure of personal data
PNC_1
Unable to respond to user right to object
PNC_2
Unable to respond to portability of user data
PNC_3
Insufficient consent/purpose in system
PNC_4
Attacker tampering with privacy policies and makes consents inconsistent
PNC_5
Tampering with policy at data store (security)
No possibility to erase data
PNC_8
Aggregated data has not been erased at all data stores
PNC_9
Personal data is processed without possibility for user to object
PNC_10
Data processed for direct marketing purposes without possibility for user to object
PNC_11
Data processed for research purposes without possibility for user to object
PNC_12
Unable to transmit stored personal data regarding data subject to third party
PNC_13
AND
System has users with an age of 15 or below
PNC_14
No parental consent for the user being a child
PNC_15
Inability for user to withdraw consent at process
PNC_16
No consent/purpose at process
PNC_17
No consent/purpose for sharing data with external parties
PNC_23
No consent/purpose for storing data
PNC_24
No consent/purpose for processing data
PNC_25
Rule 2:
(source is [External Entity] or
source is [Process]) and
target is [Process]
where
NOT ANY (df, t IN (source.sendTo, target)
WHERE df.isPII = 'True'
AND df is [DataFlow] AND df.pol = 'True')
Rule 1:
source is [Data Store]
where
NOT ANY (df, s IN (source.recieveFrom, source) WHERE s.isPII = 'True'
AND df is [DataFlowErasure]
AND df.recieveFrom is [ProcessComponent])
Rule 3:
source is [DFDNode]
target is [Process]
where
NOT ANY (df, s IN (source.sendTo, source) WHERE df.isPII = 'True'
AND df.sendTo is [ProcessRequest] AND df.sendTo.sendTo is [DataFlow] AND df.sendTo.sendTo.pol
= 'True'
AND target.recieveFrom is df.sendTo.sendTo)
Rule 4:
source is [DFDNode]
target is [ProcessComponent]
where
NOT ANY (df IN source.sendTo WHERE df.isPII = 'True'
AND df.pol = 'True')

Figure 5.4: Extended Threat Tree Policy and Consent Non-Compliance.

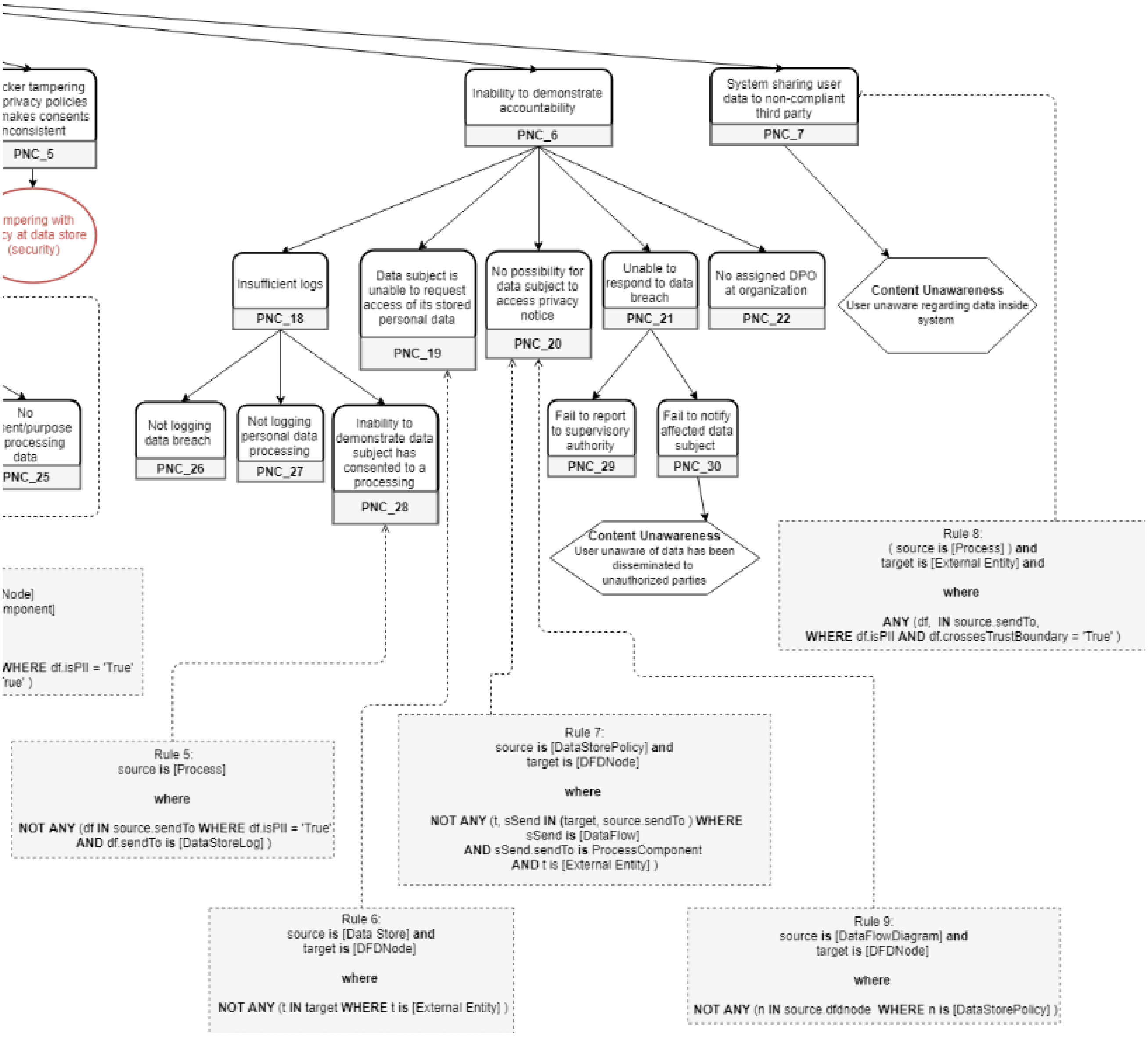

cker tampering
privacy policies
nakes consents
nconsistent
PNC_5
mpering with
cy at data store
(security)
Inability to demonstrate accountability
PNC_6
System sharing user data to non-compliant third party
PNC_7
Insufficient logs
PNC_18
Data subject is unable to request access of its stored personal data
PNC_19
No possibility for data subject to access privacy notice
PNC_20
Unable to respond to data breach
PNC_21
No assigned DPO at organization
PNC_22
Content Unawareness
User unaware regarding data inside system
No
ent/purpose
processing
data
PNC_25
Not logging data breach
PNC_26
Not logging personal data processing
PNC_27
Inability to demonstrate data subject has consented to a processing
PNC_28
Fail to report to supervisory authority
PNC_29
Fail to notify affected data subject
PNC_30
Content Unawareness
User unaware of data has been disseminated to unauthorized parties
Rule 8:
(source is [Process]) and
target is [External Entity] and
where
ANY (df, IN source.sendTo,
WHERE df.isPII AND df.crossesTrustBoundary = 'True')
Node]
mponent]
WHERE df.isPII = 'True'
'rue')
Rule 5:
source is [Process]
where
NOT ANY (df IN source.sendTo WHERE df.isPII = 'True'
AND df.sendTo is [DataStoreLog])
Rule 7:
source is [DataStorePolicy] and
target is [DFDNode]
where
NOT ANY (t, sSend IN (target, source.sendTo) WHERE
sSend is [DataFlow]
AND sSend.sendTo is ProcessComponent
AND t is [External Entity])
Rule 6:
source is [Data Store] and
target is [DFDNode]
where
NOT ANY (t IN target WHERE t is [External Entity])
Rule 9:
source is [DataFlowDiagram] and
target is [DFDNode]
where
NOT ANY (n IN source.dfdnode WHERE n is [DataStorePolicy])

Figure 5.5: Extended Threat Tree Policy and Consent Non-Compliance.

The Policy and Consent Non-Compliance Threat tree addresses threats that are directly related to the compliance of software systems and organizations with GDPR principles and user rights. The tree can be seen in Figure 5.4 and Figure 5.5, and its purpose is to make practitioners aware of the implications and issues involving compliance with legislations. This includes threats related to accountability and user consents for handling private data. The tree contains thirty nodes, including parent nodes, and child nodes, together with nine rules used to identify the presence of a specific threat in the software architecture. An overview of the whole tree can also be seen in Appendix A.1.

Out of these thirty nodes, only two have been kept from the original LINDDUN, which means that a vast rearrangement of the original tree has been made. Some of the new threats have a direct relation to the Meta Model earlier described in section 5.3.1. However, possible solutions to counter these threats are not provided in this work. All the threats are explained below.

5.4.1.1 Incorrect erasure of personal data

The threat of *incorrect erasure of personal data* (PNC_1) means to demonstrate the threat in a system where any personal data regarding a data subject that is being stored does not have the possibility to be erased. This threat, also relates to the right to erasure defined in the GDPR. Hence, the DFD elements which relates to such threat are DataFlowComponents that interact with DataStoreComponents. The incorrect erasure threat does contain two child nodes, *no possibility to erase data* (PNC_8) and *aggregated data has not been erased at all data stores* (PNC_9).

The desired outcome of the first child threat (PNC_8) is to make the practitioner of LINDDUN identify which elements in the business-oriented DFD are used to represent a Data Store. If the Data Store contains sensitive data regarding individuals, an architectural design solution to provide the possibility to erase such data needs to be included.

The second child threat (PNC_9) relates to data aggregation. It addresses aggregated data that remains in the system after an erasure operation has been performed. This happens when data regarding a data subject is processed or stored in some architectural element is then forwarded to another architectural element and then combined with another data set. The problem arises when a data subject requests to erase certain data regarding her or him, since the data has to be removed from all points inside the system. Thus, identifying where all data attributes are located in a system can become a hassle. Figure 5.6 illustrates the problem of aggregated data in a system.

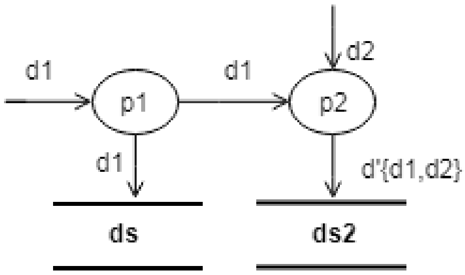


Figure 5.6: An example scenario of aggregated data in the system

5.4.1.2 Unable to respond to user right to object

This threat relates to the right of the user to object on a certain processing of her or his personal data. Hence, it relates to the right to object defined in the GDPR. Here, the system is a victim of being *unable to respond to the user right to object* (PNC_2) as soon as the system contains a process which processing personal data without the possibility to make a request for objecting to such processing (PNC_10). Another more special use cases of data processing are when data is processed for direct marketing purposes without possibility for user to object (PNC_11) or data is processed for research purposes without possibility for user to object (PNC_12) which are also specified in the regulation. When the system does not provide a possibility for the user to object to direct marketing, the system becomes non-compliant with the legislation. Consequently, the system can become non-compliant if it is not capable to provide the user the possibility to object to data used for research purposes.

5.4.1.3 Unable to respond to portability of user data

The right to portability is one of the new and introduced provisions in the GDPR. Organizations shall provide users with the possibility to transfer and access their personal data through different platforms. If the system is unable to respond to a user request regarding data portability and send this data to a specified third party (PNC_13) the system can become non-compliant (PNC_3). The GDPR does not specify any specific way of how to comply with this provision. Although, it is stated for data controllers (organizations), when technically feasible, the user data shall be transmitted directly to the third party.

5.4.1.4 Insufficient consent/purpose in system

This threat (PNC_4) relates to the GDPR principle Lawfulness, fairness and transparency, as well as the user consents. In order to collect and process personal information, organizations must define and specify the reason and purpose for the collection and processing, and obtain consent from the user. If this information is not included and the organization fails to reach sufficient privacy standards according to the regulation, the system becomes non-compliant, and thus represent a threat towards the privacy of its users. The threat does contain four child nodes explained below.

The two first child nodes of this relates to a new provision of the GDPR, conditions for consent when a system contains users who is the age of 15 or below (PNC_14). If this is true, then the system must provide its users the possibility for the parental responsible person to consent on behalf of the child, who is 15 years of age or below (PNC_15). Further, description on how to provide the users with this possibility is not explained in the regulation. Therefore, it is up to the architect to find a technical design solution that satisfies this provision.

Insufficient consent (PNC_4) can also occur when a data subject does not possess the ability to change its reasoning for the processing of its personal data. In other words, the system does not provide the user the ability to withdraw consent(s) at a process (PNC_16).

The fourth child threat that leads to insufficient consent or purpose in a system is when no consent or purpose exists at a process (PNC_17). Such non-existent consent can be at any of the following processing scenarios: when a ProcessComponent sharing data with external parties (PNC_23), when data is stored at a DataStoreComponent (PNC_24) or overall when data is processed at a ProcessComponent (PNC_25).

5.4.1.5 Attacker tampering with privacy policies and makes consents inconsistent

The (PNC_5) threat remains the same as in the original LINDDUN version. Since it is believed to represent a valid and accurate threat regarding consents and privacy policies.

5.4.1.6 Inability to demonstrate accountability

To process or store personal data in a system, the data subject must approve it. If these actions are performed with the absence of an approval from the data subject, it violates the principle of Lawfulness, fairness and transparency. In most cases the approval is accomplished by the data subject sending a consent which implies he or she either agrees or disagrees with the action. If the system supports consents but does not have the ability to demonstrate that consents are present in the system, it can become non-compliant. Therefore, the PNC_6 threat represents the inability of the organization to demonstrate compliance with the regulation. The cause to the inability to demonstrate accountability threat can vary depending on its context. Thus, this threat node is divided into smaller threat nodes. They relate to a software system insufficiently performing logs (PNC_18); a data subject that cannot access a copy of stored data to which the data relates (PNC_19); no possibility for data subject to access privacy notice (PNC_20); a system being unable to respond to a data breach (PNC_21); and no assigned DPO at organization (PNC_22).

The implications of a system having *insufficient logs* (PNC_19) can vary depending on the context. One of the contexts can refer to when a software system has been a subject of a data breach. If no logs of this data breach has been made (PNC_26) the organization most likely will be unable to demonstrate accountability. Another context would refer to the absence of logs for when a system is processing personally sensitive data (PNC_27). Finally, when a system is insufficiently performing logs, this can lead to the inability to demonstrate that a data subject has consented to a processing (PNC_28). More specifically to where and when in an architectural design a data subject has made a consent and what such consent covers.

The threat is regarding when a data subject is unable to access a copy of his or her stored data (PNC_19). This threat is directly related to the GDPR right to access.

In essence, the architect has to ensure that there is a way for the data subject to receive a copy of his or her personal data upon request. Consequently, from where in the system the data is stored there has to exist one or several data flows by which, alone or together, serves the purpose of making a transaction of this data to the data subject. There is no “true solution” to this due to the vast difference in how systems vary in size, complexity, computational performance etc. The same reasoning, of accessing of stored data as explained above, exists for the inability to access the organization’s privacy notice (PNC_20). The difference is that this threat focuses at the privacy notice of the organization instead of the actual data.

Another of the new provisions of the GDPR is the obligation of the data controllers to perform notifications upon a data breach. The provision states that two parties shall be notified is a data breach has occured; the supervisory authority and the affected data subjects as well. Therefore, the threat unable to respond to data breach. This is also the reasoning for including the fourth child node, *unable to respond to data breach* (PNC_21). This threat node is parent to two child nodes, *fail to report to supervisory authority* (PNC_29) and *fail to notify affected data subject* (PNC_30). The two child nodes explains the two parties that shall be notified.

The last child node, no assigned DPO at organization (PNC_22), does not have a direct connection to a threat modelling context but instead aims at at expressing the GDPR provision of assigning a responsible person at the organization for ensuring and maintaining compliance towards the regulation. he purpose of including this threat in the tree is to let the practitioner of the methodology aware of the possible need for assigning a responsible position at the organization. The DPO can ensure privacy enforcement for tasks that might not be feasible to implement into the software system. This could for instance be the request for accessing a privacy notice. The feature where such request is made electronically by a user through the software system might perhaps not be feasible to implement due to economical costs. Instead this responsibility perhaps could be assigned to the DPO if applicable. Note, this is an exemplification of a scenario which might be applicable for a one organization, but for another it might not be applicable. Further, the DPO is not limited to the responsible task described in the scenario above only. It is up to the organization to decide what is applicable in their context.

5.4.1.7 System sharing user data to non-compliant third party

When a system that processes personal data shares this data with a third party which is not compliant to the GDPR, the user’s power over his or her own data gets severely violated. Firstly, if data is shared to any third party (even GDPR compliant third parties) and the user requests for erasure of his or her data, then the system is compelled to reach out to all third parties to which the data subject’s data has been disseminated to. Secondly, if such request is made and the third party is non-compliant (PNC_7), this makes the system itself (which has shared the data) violating the user rights. This since sharing of data has been made to an organization which potentially will not take actions regarding the data on behalf of whom the data relates to. Hence, this threat relates on the phenomenon of the transaction of outgoing data from the system to external entities.

5.4.2 Content Unawareness Tree (CU)

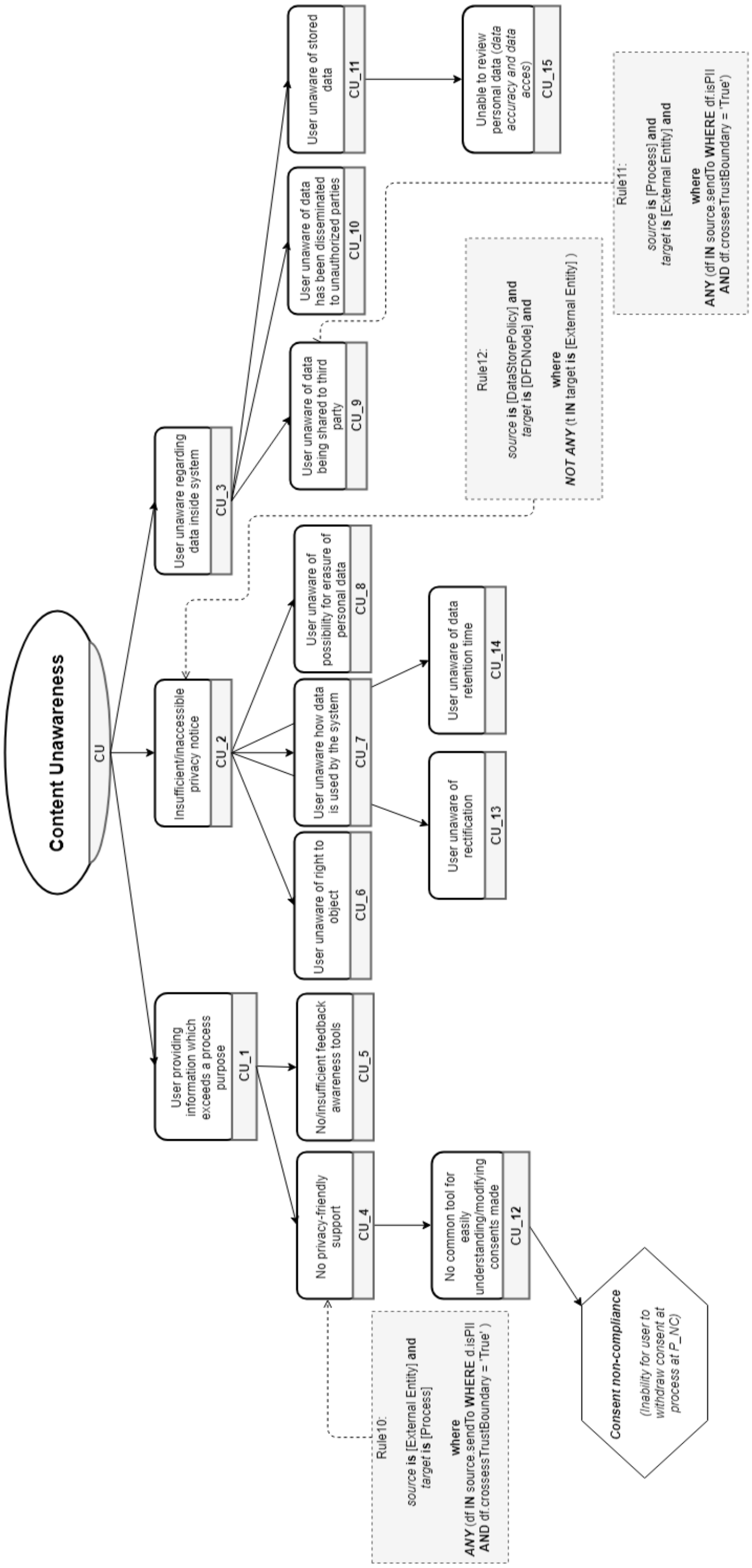

Content Unawareness
CU
User providing information which exceeds a process purpose
CU_1
Insufficient/inaccessible privacy notice
CU_2
User unaware regarding data inside system
CU_3
No privacy-friendly support
CU_4
No/insufficient feedback awareness tools
CU_5
User unaware of right to object
CU_6
User unaware how data is used by the system
CU_7
User unaware of possibility for erasure of personal data
CU_8
User unaware of data being shared to third party
CU_9
User unaware of data has been disseminated to unauthorized parties
CU_10
User unaware of stored data
CU_11
No common tool for easily understanding/modifying consents made
CU_12
User unaware of rectification
CU_13
User unaware of data retention time
CU_14
Unable to review personal data (data accuracy and data acces)
CU_15
Consent non-compliance
(Inability for user to withdraw consent at process at P_NC)
Rule10:
source is [External Entity] and target is [Process]
where
ANY (df IN source.sendTo WHERE d.isPII AND df.crossessTrustBoundary = 'True')
Rule11:
source is [Process] and target is [External Entity] and
where
ANY (df IN source.sendTo WHERE df.isPII AND df.crossesTrustBoundary = 'True')
Rule12:
source is [DataStorePolicy] and target is [DFDNode] and
where
NOT ANY (t IN target is [External Entity])

Figure 5.7: Extended Threat Tree Content Unawareness

The *Content Unawareness* tree in LINDDUN takes the perspective as an external entity. The aim is to provide the external entity awareness of the implications of sharing data to a software system. The tree also advocates that the user is given enough information from the software system in order to be able to take his or her own decision whether to submit or not submit his or her data for further processing. The tree also aims to make the practitioner aware of the threats related to the user not being aware of the principles and user rights from GDPR. This is just as important as the actual implementation of such principles and rights. If a user is not aware of the rights regarding his or her own data, the implementation of the principles and user rights is nonsensical.

The tree addresses three major contexts: the unawareness of the user regarding the privacy implications before submitting his or her data (CU_1), the unawareness of the user regarding his or her data after it has entered the software system (CU_3), when the user has no access or given an insufficient privacy notice (CU_2). The CU threat tree is further explained below.

5.4.2.1 User providing information which exceeds a process purpose

This threat is the same as the threat “User providing too much information” (U_1) from the original Content Unawareness threat tree. The original threat tree node (U_1) is interpreted as ambiguous since one cannot, by reading the name of the threat node, understand the definition of “too much” [1]. In order to provide an easier and concrete definition of what is considered to be “too much”, the name of the threat tree has been changed. The threat contains three child nodes, which all specify the root cause to the threat discussed above. The threats are: *No privacy-friendly support* (CU_4) and *No/insufficient feedback awareness tools* (CU_5).

The first child (CU_4) node resides unchanged from the original version of LINDDUN. The threat refers to the non-existence of a way for a user to modify and set a personal privacy preference profile. A privacy profile by which the system uses and takes in consideration when processing personal information about the data subject. This is considered indeed essential and hence the reason to why it has been left unchanged. Earlier it was explained, in section 5.4.1.4 regarding the proposed policy and consent non-compliance threat tree, the threat of the inability for a user to withdraw a consent (PNC_16). This specific threat also relates to the unawareness of a user. When systems escalate in size and complexity the GDPR principle of consent compliance may easily get a hassle to manage. It can become really hard for an organization to provide the possibility for the user to understand what has been consented to and what has not in an easier way.

The GDPR suggests that organizations shall allow users to modify the consents made when desired [73]. This is one way to empower users to decide what to consent to and have control over the consents made. According to GDPR, this can be achieved through the use of dashboards. Dashboards are a typical tool that can be used to gather and organize user consents, making all the consents accessible in one place. It is a tool which helps an organization to provide privacy-friendly support

(CU_4). Thus, the user can access the dashboard to review and/or modify his or her consents. If the organization does not provide the possibility for the user to modify (withdraw) his or her consents, it faces the risk of being non-compliant with the GDPR. Hence, the threat (CU_12) in the Content awareness (CU) tree. However, the GDPR literature does not put an obligation for a presence of such dashboards for consents in order make the organizations to be compliant to the GDPR. Nevertheless, the decision of including such threat in the tree was due to the believed benefit for the inexperienced practitioner of the LINDDUN methodology be aware of such potential threat. In this case, the threat of a user being unaware of what has been consented to.

The second child node, *no/insufficient feedback awareness tools* (CU_5), has been kept the same as the original version of LINDDUN. This threat refers to the unawareness of the user regarding the information he or she is sharing. If no feedback or awareness tools are used, the user is most likely to be unaware of the impact of sharing his/her personal data.

5.4.2.2 Insufficient/inaccessible privacy notice

As has been explained in section 3.5 the GDPR advocates the data subject's (user's) right to be informed. The right to be informed states that the user has the right to know how his or her personal data is processed and for what purposes. The second child node, *insufficient/inaccessible privacy notice* (CU_2), refers to the same context. it will be considered non-compliant with the regulation and face possible fines defined by the regulation. In a software design context, the risk of these fines can be invoked by either not including sufficient information in the privacy notice or not providing the user access to such privacy notice. This threat specifies what shall be included in the privacy notice, the right to object (CU_6), how data is used by the system (CU_7), the right to erasure personal data stored in the system (CU_8), right to a data subject to update inaccurate private data regarding himself (CU_13), and for how long such data is stored (CU_14). Although, the scenario of an insufficient privacy notice might not be invoked by a design flaw of the system's architecture it still provides value to the practitioner of the LINDDUN methodology by proving awareness.

5.4.2.3 User unaware regarding data inside system

The third child node, *User unaware regarding data inside system* (CU_3), can occur when a user has not received sufficient information regarding how his or her personal data is stored and processed in the system. This can be due to either *User unaware of data being shared to third party* (CU_9), *User unaware of data has been disseminated to unauthorized parties* (CU_10) or *User unaware of stored data* (CU_11). The former node is applicable to systems where intentions of disseminating personally sensitive data about its users exists. The same reasoning applies for the second child node but with a restriction to the third party is not authorized to access such information. The third child node has been kept from the original version of

LINDDUN. It explains the possibility for the user to be able to retrieve the data stored by the system.

5.5 LINDDUN+ Threat Tree Rules

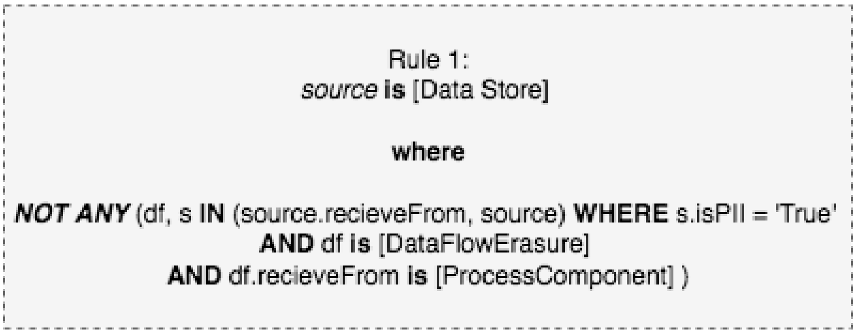
LINDDUN+ threat tree rules aim to take an initial step towards a future automated tool and have been developed with the help of a pseudo code. Again, as mentioned earlier they are an attempt towards automation. The rules address the extended threat threes, Content Unawareness and Policy and Consent Non-compliance. Also, not all threats inside the trees were included. This due to time constraints and because of the subjective nature of the threats. Each rule consists of two parts: a definition of the elements to which the rule applies, and the condition when the threat is applicable. The structure of the first part has close similarities to the language used by the Microsoft Threat Modeling Tool. Additionally, the rules can be seen as an exploratory attempt to see feasible techniques to identify the presence of LINDDUN+ privacy threats inside the architecture.

As was described in section 5.3.1 the work by [13], [12] has served as inspiration for developing the Meta Model. Their work can be seen as a significant step forward for the privacy modelling literature towards automation, i.e. the development of an automated tool similar to the Microsoft's Threat Modeling Tool. Since the same interest of accomplishing automated means shared for this study, their work has served as inspiration for the rules of LINDDUN+ as well. Each rule considers one or several PA-DFD elements from the Meta Model explained in section 5.3.1. The goal of the rules is to help the practitioner become more aware of the privacy violations.

5.5.1 Rule 1: No possibility to erase data

This threat node (PNC_8) can be found in the policy and consent non-compliance tree. The fist part of rule 1, as presented in figure 5.8, shows that in order for the threat to be present in the architecture, the source (or the starting point) has to be a data store. In this case, the (PNC_8) only relates to data stores.

After meeting the first condition, the second part of rule 1 tells the practitioner when the treat applies. In this case the threat reveals when the data store contains sensitive personal information about the users. Additionally, there are no data flows (df) of type DataFlowErase connected to the data store under evaluation which by itself has a sending data flow which being a process.



Rule 1:
source **is** [Data Store]

where

NOT ANY (df, s **IN** (source.recieveFrom, source) **WHERE** s.isPII = 'True'
AND df **is** [DataFlowErasure]
AND df.recieveFrom **is** [ProcessComponent])

Figure 5.8: Rule 1 in Policy and Consent Non Compliant Tree.

Figure 5.9 shows an example of how Rule 1 can be identified in a PA-DFD. As presented in the picture, the source is a data store which receives incoming data from a process. The red rectangle in the PA-DFD demonstrates the presence of the threat. It indicates that there is no possibility to erase data at the specific database, since the data flow component is an ordinary data flow. The red rectangle inside the DataFlowComponent shows the element used to represent the possibility to erase data in a PA-DFD, and thus mitigate the presence of the PNC_8 threat.

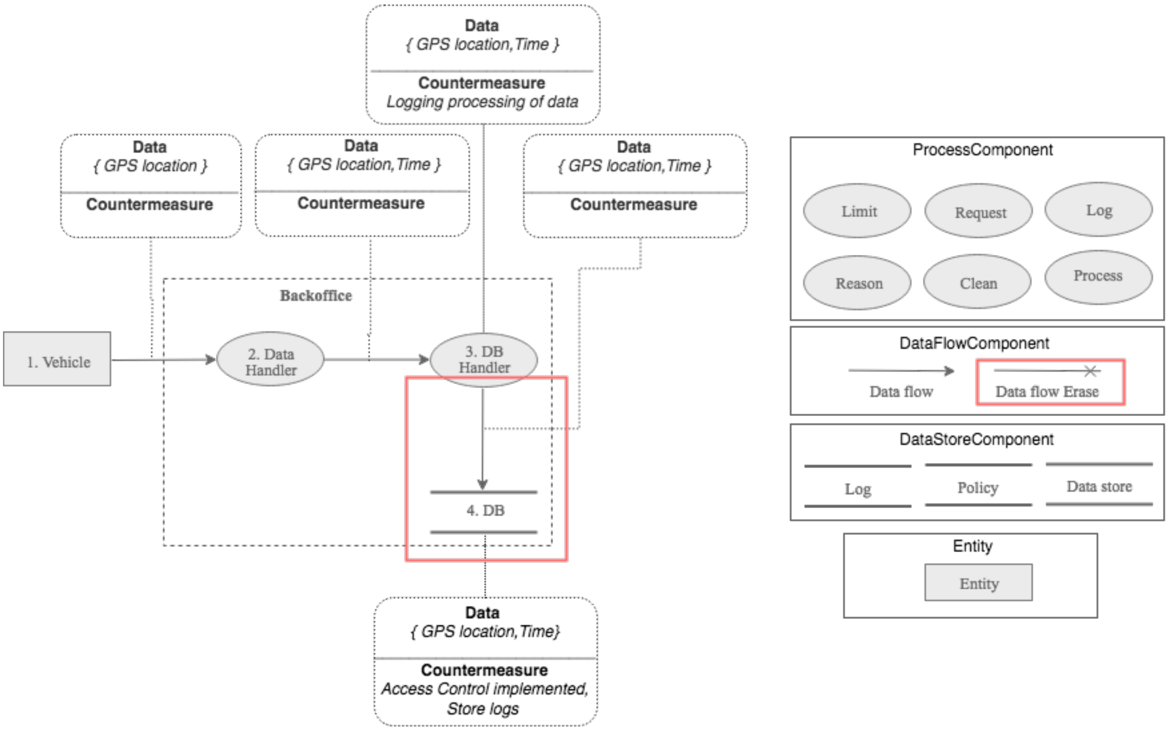
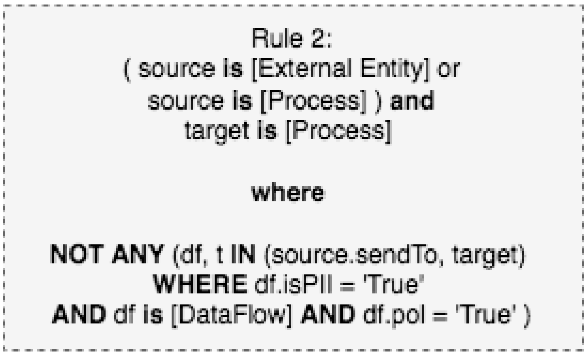


Figure 5.9: Rule 1 example.

5.5.2 Rule 2: Unable to respond to right to object

The rule for triggering the inability to respond to right to object can be seen in figure 5.10 and refers to the node (PNC_2) in the policy and consent non-compliance tree. The sources can be either external entities (users) or processes in a software system. While the target (or ending point) can only be processes.

The condition for when the threat applies is when a process does not possess any incoming data flow that passes a consent. Also, the data received by the process is considered to be personally identifiable. The consent represents the decision of the data subject in accepting or not the processing of the data. Additionally, the consent can either come directly from the external entity or be forwarded from a preceding process. Although the rule focuses at a certain process, the practitioner of LINDDUN can use the rule to trace a chain of processes which sends consents between each other in the architecture.



Rule 2:
(source **is** [External Entity] or
source **is** [Process]) **and**
target **is** [Process]

where

NOT ANY (df, t **IN** (source.sendTo, target)
WHERE df.isPII = 'True'
AND df **is** [DataFlow] **AND** df.pol = 'True')

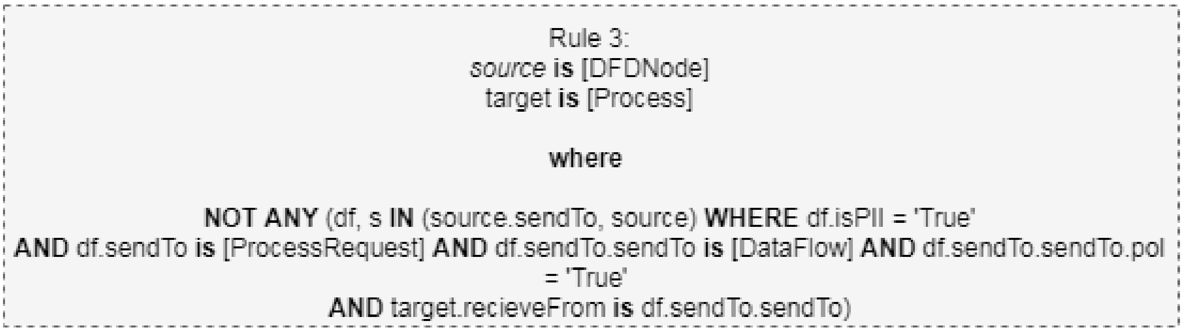
Figure 5.10: Rule 2: unable to respond to user right to object

5.5.3 Rule 3: Insufficient consent/purpose in system

This rule represents the (PNC_4) in the threat tree. This threat can occur when the source which can be any type of DFD element(DFDNode) sends data to a process (target).

In order for the threat to be present, there is no incoming data flow containing a consent to the processing at the process. The consent should come from a DFDNode Request process that can receive it from the source node or from a relevant previous node. A consent can be received at a process either from a external entity or a preceding process. The preceding process can obtain the consent from another process or external entity.

Figure 5.12 tries to illustrate when this threat can apply. The figure is a combination of two design implementations and a logical expression. The left side of the figure explains the scenario of an interaction between two processes. The right side of the figure explains the reason that triggers the threat. Furthermore, the data flows, to the right of the conditional operator which are marked as bold, is the non-present part of the left hand side of the figure, and hence the reason to the threat.



```
Rule 3:
source is [DFDNode]
target is [Process]

where

NOT ANY (df, s IN (source.sendTo, source) WHERE df.isPII = 'True'
AND df.sendTo is [ProcessRequest] AND df.sendTo.sendTo is [DataFlow] AND df.sendTo.sendTo.pol
= 'True'
AND target.recieveFrom is df.sendTo.sendTo)
```

Figure 5.11: Rule 3: insufficient consent/purpose in system

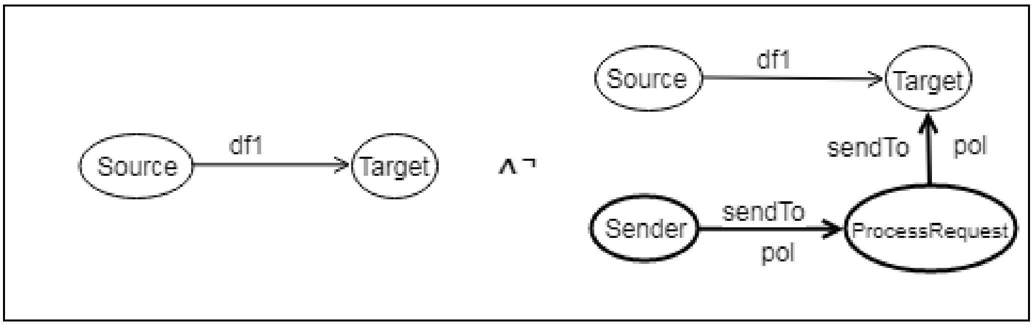
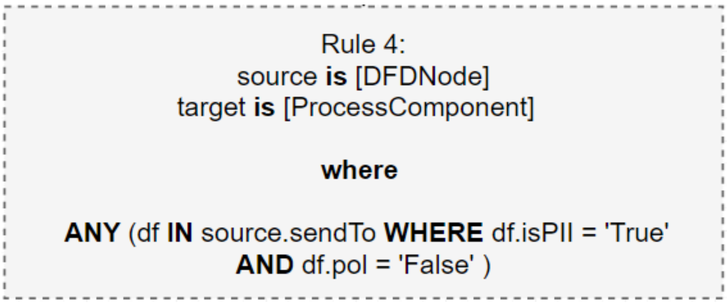


Figure 5.12: Logical expression of design flaw when rule3 being present in a system

5.5.4 Rule 4: No consent/purpose at process

One of the biggest concerns of the GDPR is that personal data about a data subject shall be processed only upon approval from the data subject himself/herself. This is what the fourth rule addresses as well, the threat of a process processing data without taking consideration from the data subjects approval of, or not having a legitimate purpose behind, such processing.

The threat rule can be seen in Figure 5.13 and tries to illustrate the design flaw of *No consent/purpose at process* (PNC_17). The rule specifies that the rule applies when there is any type of DFD element (DFDNode) that has a connection to a ProcessComponent. Further, there is a Data Flow from the source node (DFDNode) which transmitting data classified as PII to the ProcessComponent but there is not any consent provided for making the processing legitimate.



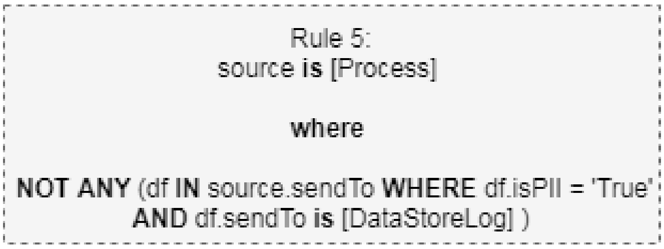
```
Rule 4:
source is [DFDNode]
target is [ProcessComponent]

where

ANY (df IN source.sendTo WHERE df.isPII = 'True'
AND df.pol = 'False' )
```

Figure 5.13: Rule 4: No consent/purpose at process

5.5.5 Rule 5: Inability to demonstrate accountability

As explained earlier, this threat is found in the policy consent non-compliance threat tree and it is a parent node of six child nodes. Rules have been created for three of these child nodes. Since the intention of the policy non-compliance tree is to focus on a system overview not all threats nodes are possible to create query rules for.

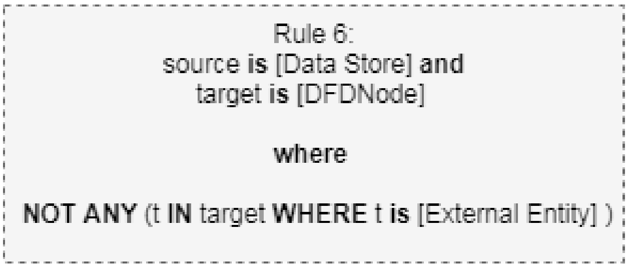
The threat *Inability to demonstrate data subject has consented to a processing* (PNC_28) is represented in figure 5.14. Here, a process is unable to demonstrate that a data subject has consented to the processing of his or her personal data, since the process does not perform any log for the consent and for what data has been processed. Hence, the threat is present when a connection between the process and a DataBaseLog is non-existent.



```
Rule 5:
source is [Process]

where

NOT ANY (df IN source.sendTo WHERE df.isPII = 'True'
AND df.sendTo is [DataStoreLog] )
```

Figure 5.14: Rule 5: inability to demonstrate data subject has consented to a processing

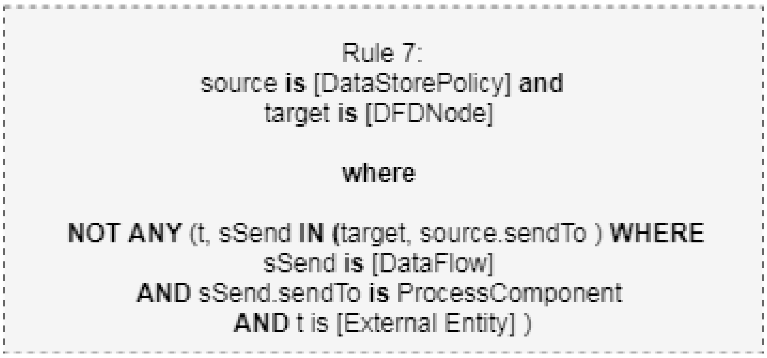
The rule for the threat *Data subject is unable to request access of its stored personal data* (PNC_19) is illustrated in figure 5.15. The rule focuses at a data store where a privacy notice is stored. Additionally, the rule specifies the nonexistence of a flow where the destination (target) is an external entity. Important to add is that the flow here is not limited to only the interaction between two DFD elements but can also consists of the interaction of several DFD elements.
The third child node to the accountability threat node is the *No possibility for data subject to access privacy notice* threat (PNC_20), and it has two defined rules. The rules are shown in the figures 5.16 and 5.17. Note, assumptions have been made, such as that the privacy notice is stored in the PA-DFD element DataStorePolicy. If this would not be the case, the DataStoreComponent would be the same element



```
Rule 6:
source is [Data Store] and
target is [DFDNode]

where

NOT ANY (t IN target WHERE t is [External Entity] )
```

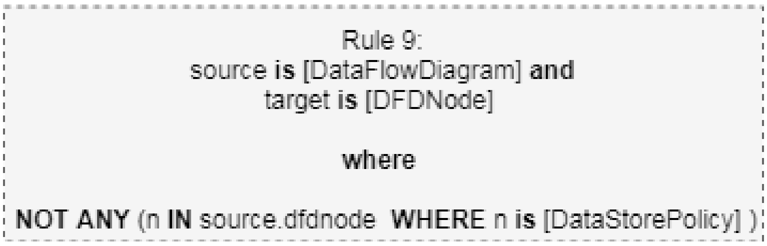
Figure 5.15: Rule 6: data subject is unable to request access of its stored personal data

type as the element type where the privacy notice would reside. This threat can arise when a policy database exists in the system and there is no flow or several ProcessComponents where the target element is an external entity (system user). The rule is present when no such data base element is present in the system in the first place.



```
Rule 7:
source is [DataStorePolicy] and
target is [DFDNode]

where

NOT ANY (t, sSend IN (target, source.sendTo ) WHERE
sSend is [DataFlow]
AND sSend.sendTo is ProcessComponent
AND t is [External Entity] )
```

Figure 5.16: Rule 7: no possibility for data subject to access privacy notice

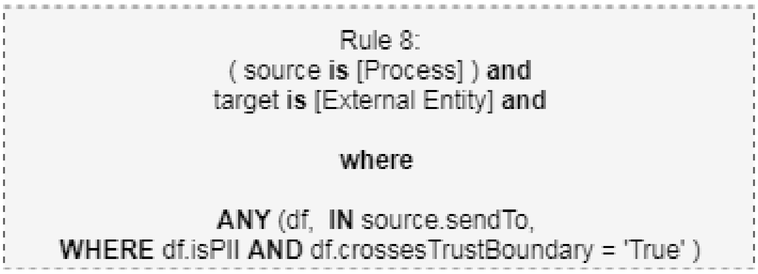


```
Rule 9:
source is [DataFlowDiagram] and
target is [DFDNode]

where

NOT ANY (n IN source.dfdnode  WHERE n is [DataStorePolicy] )
```

Figure 5.17: Rule 9: no possibility for data subject to access privacy notice

5.5.6 Rule 8: System sharing user data to non-compliant third party

This rule represents the presence of the (PNC_7) threat. It can occur when the source is a process that has outgoing data transactions which leave the system's trust boundary. In this case, the outgoing data contains personal information regarding systems' users. This threat is shown in figure 5.18. It is also important to note that, the interaction between a single process sending data through a data flow which crosses a trust boundary to a single external entity, does not automatically lead to the violation of *system sharing user data to non-compliant third party.* Instead, it is up to the practitioner to look at the whole picture and judge the likelihood of such threat to happen. The intention of the rule is to present, to the practitioner, the most common scenario when the threat applies.

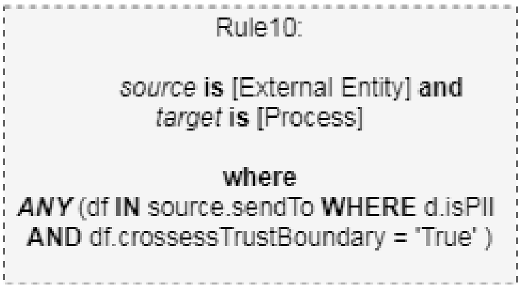


```
Rule 8:
( source is [Process] ) and
target is [External Entity] and

where

ANY (df,  IN source.sendTo,
WHERE df.isPII AND df.crossesTrustBoundary = 'True' )
```

Figure 5.18: Rule 8: System sharing user data to non-compliant third party

5.5.7 Rule 10: No privacy-friendly support

This threat node resides in the content unawareness tree, and as in the previous threat, this threat (CU_3) also relates to the interaction between an external entity and a process. However, in this case, the external entity is the source (initial point) while the process is the target. The threat is applicable for flows where an external entity sends personal sensitive data over a trust boundary to a process. Of course this is a rather generic interaction exists in most software systems. As has been explained in the previous rules, it is up to the practitioner of LINDDUN to consider if this threat is applicable or not. This can of course vary between the context a software system is working in.

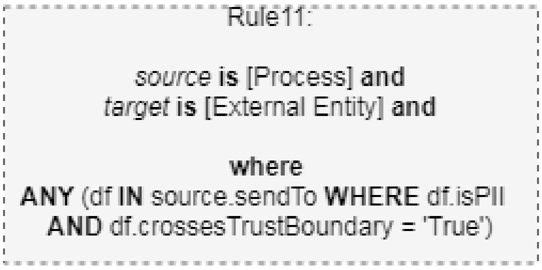


```
Rule10:

source is [External Entity] and
target is [Process]

where
ANY (df IN source.sendTo WHERE d.isPII
AND df.crossessTrustBoundary = 'True' )
```

Figure 5.19: Rule 10: no privacy-friendly support

5.5.8 Rule 11: User unaware of data being shared to third party

The sharing of personal sensitive data to third parties can only be done by a system through a ProcessComponent. The source is a process that sends data to an external entity (target). The rule for this threat tells the practitioner to look at each process in the design that is processing personal data. The purpose is to investigate if there exists a process which has a single data flow or chain of flows that crosses a trust boundary. If the interaction exists the condition for the threat is fulfilled, and hence the risk of the threat to be present in the system.



Rule11:

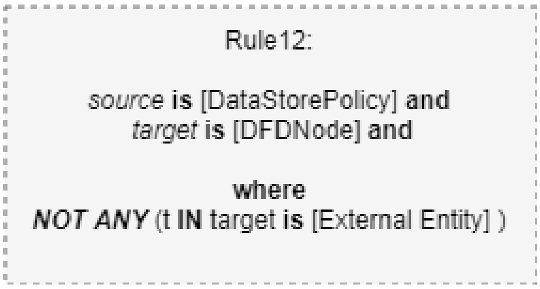
source **is** [Process] **and**
target **is** [External Entity] **and**

where
ANY (df **IN** source.sendTo **WHERE** df.isPII
AND df.crossesTrustBoundary = 'True')

Figure 5.20: Rule 11: System sharing user data to non-compliant third party

5.5.9 Rule 12: Insufficient/inaccessible privacy notice

The name of the threat implies that a privacy notice is either insufficient or inaccessible. However, the rule for this threat focuses on the latter part, since it is not possible to model an insufficient privacy notice. The threat of insufficient or inaccessible privacy notice (CU_5) can arise when a policy data store exists in the system but the target does not represent an external entity. In this case, the system does store the organization's privacy notice in a policy data store but no single data flow or chain of flows exist, which ends with an external entity.



Rule12:

source **is** [DataStorePolicy] **and**
target **is** [DFDNode] **and**

where
NOT ANY (t **IN** target **is** [External Entity])

Figure 5.21: Rule 12: System sharing user data to non-compliant third party

6
Evaluation of Results

This section describes the results gathered during the evaluation of LINDDUN+, as well as the steps and work performed to determine the significance and scientific contribution of this study. For the evaluation of the results the same system used in the case study for the first iteration was utilized.

6.1 Pilot Workshop

Before performing the evaluation of the proposed extensions, a pilot study was conducted. The pilot workshop consisted of a workshop where the LINDDUN+ methodology was applied. The participant of the pilot workshop is a security engineer from Volvo Group Trucks Technology working with cyber-security and risk assessment. The participant has a great working experience with security, but does not possess any expertise in terms of privacy. Also, the participant did not have any previous knowledge regarding LINDDUN, however STRIDE sessions are part of the participant work activities. As the steps in STRIDE and LINDDUN are considered to be similar, specially in the initial steps, the participant was considered to have relevant knowledge to take part in the evaluation.

The purpose with the pilot workshop was not to analyze the performance of the a participant in terms of found privacy threats and their significance or correctness. Instead the purpose was to understand what parts of an empirical evaluation of the LINDDUN+ that was harder, for a practitioner, to understand. In other words, to reveal unexpected difficulties of the LINDDUN+, for the participant, that would affect the significance of the results. Another purpose behind the pilot workshop was to reflect on the assumptions, that was made by the authors, for the developed business-oriented DFD. Assumptions regarding default functionality of how software systems are build in the automotive domain. As has been explained earlier, the authors of this study are considered to possess limited knowledge in this area. Therefore the importance to understand this by conducting the pilot workshop with the security-engineer at Volvo Group Trucks Technology. Also, by understanding what things of the given material was harder for the participant to grasp, the authors could prepare for this in the future workshop. Hence, to focus more on these aspects during the real workshop, the workshop which is explained in section 6.2.3.

6.1.1 Preparation

The preparation for the pilot workshop consisted of four steps: 1) Development of the DFD, 2) Defining a baseline for the results, 3) Creation of LINDDUN+ Guide 1 and Guide 2, GDPR principles description, and System description document, and 4) Introduction session. All material and introductory sessions explained in this section are created and held by the authors of this study.

1. **Development of the DFD**: the DFD was developed based on the initial system description provided by Volvo Group Trucks Technology. The system used was the same platooning system that was used for the first iteration. From the system description a component diagram was created by the authors and utilized to derive the final business-oriented DFD. Note, the business-oriented DFD was developed by the authors of this study and hence provided to the participant of the pilot workshop. A domain expert was also consulted to validate the correctness, in terms of the functionality, of the developed DFD prior the workshop.

2. **Defining a baseline for the results**: based on the business-oriented DFD, that was developed by the authors, and the system domain, a baseline for the results was created. This baseline reflects the amount of threats considered to be found by a privacy and domain expert in the system under analysis. It was defined after a systematic analysis of each individual component of the developed business-oriented DFD, together with a closer look in the flows containing personal data, and assumptions of the system behaviour. The baseline is further explained in section 6.2.2.

3. **Creation of LINDDUN+ Guide 1 and Guide 2, GDPR principles description, and System description document**:

    - **LINDDUN+ Guide 1**: the guide 1 contains a step-by-step guide of the LINDDUN+ methodology. The guide provides a description of the steps, together with concrete examples for each step. Furthermore, the guide also includes the two extended threat trees: Content Unawareness and Policy and Consent-Non Compliance.

    - **LINDDUN+ Guide 2**: the guide 2 provides a detailed description of the threat tress and all individual nodes included in guide 1. The intention with this guide is that the practitioner himself or herself can read up on what and how each threat node implies.

    - **GDPR principles description**: this document provides an overall description of the GDPR provisions, together with the principles and user rights used in this study.

    - **System description document**: this document is a summarized version of the system description that was provided by Volvo Group Truck

Technology. The same system description that was used for the first iteration. The summarized version contains a brief description of the platooning system. The authors were needed to create such summarized version of the system because of the limited time frame of the upcoming workshops and also due to the system being classified. Hence, technical details and sensitive information, from Volvo's perspective, as well as unnecessary use cases were excluded from the original version of the system description. Further the summarized system description explains the type of data that is sent, the main components present in the system, the four use-cases, the definition of the terms used, and finally the developed DFD.

4. **Introduction session**: the introduction session was performed two days prior the workshop session. It consisted of a presentation of the LINDDUN and the LINDDUN+ steps. The session lasted in total an hour and the LINDDUN+ guide 1 and 2, and the GDPR principles description document was provided to the participant. The introduction session aimed to provide the participant with a base knowledge necessary to conduct the pilot workshop, and be introduced to the LINDDUN+ methodology.

6.1.2 The Pilot Workshop at Volvo Group Trucks Technology

The pilot workshop took place two days after the introduction session and lasted for four hours in total. The workshop included a brief recapitulation of the introduction session, where LINDDUN+ steps were presented once again. This provided the participant an opportunity to revise the steps and ask any possible questions regarding the methodology. Thereafter, the platooning system was presented to the participant together with the system description document.

Different observations were made during the pilot workshop:

The first observation was regarding the system itself. During the pilot workshop, a lot of time and efforts were spent on relevant assumptions regarding the automotive domain. The assumptions defined prior the pilot workshop over the automotive system, made by the authors, was not considered enough according to the participant. Additional assumptions were made and could only be possible due the great domain expertise and work experience of the participant with risk assessments techniques. These assumptions helped the authors of this study to clarify characteristics of the system that were not present in the architecture. Note, the architecture that was derived from the system description provided by Volvo Group Trucks Technology and also the architecture developed by the authors of this study. Further, such characteristics was not clear to the authors upon creation of the architecture due to their limitation of knowledge with the automotive domain. Examples of assumptions over such system characteristics are security mechanisms that are considered as a standard for establishing V2V communication. Since security provides confidentiality and thus enforces privacy in this way such mechanisms where considered relevant.

The second observation was related to the scope for the evaluation. Initially, the recap of the introduction session was planned to last thirty minutes including the presentation of the system. However, in practice it required more time than expected. Specially, because the assumptions made before the actual performance of the methodology. Also, the participant had a tendency to keep his focus at finding privacy threats on an implementation level rather than on an overview or design level. This resulted in the participant easily got stuck on analyzing a one specific DFD element which resulted in not all areas of the software system had the same amount of analysis. Hence, the authors gained knowledge from this was to emphasize the importance of having the mindset of eliciting privacy threats on an overview, and not focus on specific details that might concern a more technical implementation.

The third observation was regarding how to introduce the LINDDUN+. Since the time was limited for conducting the workshop and a lot of time was spent with the recapitulation and assumptions, the time for applying LINDDUN+ and for finding threats was also limited. This affected the results of the pilot workshop and limited the participant in discovering only one relevant threat. Hence, the importance to be clear and concise in how to introduce the LINDDUN+ methodology to other inexperienced participants, in a future empirical setting, was clear to the authors.

6.2 Evaluation of LINDDUN+

The evaluation of LINDDUN+ was conducted with three students from both Chalmers and University of Gothenburg. It was performed after the pilot study and also consisted of two sessions: 1) introduction session and 2) the workshop session. This section describes in full detail the steps of the final evaluation of the LINDDUN+ methodology.

6.2.1 Expected coverage to the GDPR

The LINDDUN+ extensions were developed with an aim to provide a higher coverage in terms of compliance with the GDPR. Thus, before proceeding with the final workshop, a second mapping table mapping the principles and user rights to LINDDUN+ threat trees was created. The table was mapped according to the table created for the mapping of LINDDUN and PIA in the first iteration. The table, presented in table 6.1 shows the results of the comparison of LINDDUN+, LINDDUN and PIA.

As presented in section 4.1.1, the original version of LINDDUN had a coverage rate of 14 percent, while PIA showed a coverage rate of 71 percent. Looking at Table 6.1, LINDDUN+ does address 12 principles and user rights. In theory, this result demonstrates that LINDDUN+ possess a coverage rate of 86 percent of the GDPR provisions. This implies a coverage increase of 42 percent units compared to the original version of LINDDUN. According to these results, a practical evaluation of

LINDDUN+ shall be able to provide a better coverage of the GDPR provisions than LINDDUN and PIA.

Table 6.1: LINDDUN versus PIA (GDPR mapping table).

GDPR Principles and User Rights	LINDDUN	LINDDUN+	PIA
1. Lawfulness, fairness and transparency	No	Yes	No
1.1 User consent for processing personal data	x	x	x
1.2 User consent for sharing data to third parties	-	x	x
1.3 User is able to view/update/withdraw consent	-	x	-
1.4 Lawful processing due child's consent	-	x	-
2. Purpose	No	Yes	Yes
2.1 Data must be collected for specified, explicit and legitimate purposes	-	x	x
2.2 Collected data for one purpose is not used for another purpose	-	x	x
3. Adequacy and data minimization	Yes	Yes	Yes
3.1 Data is limited to what is necessary to the purpose	x	x	x
4. Accurate and up-to-date processing	Yes	Yes	Yes
5. Storage limitation and retention time	Yes	Yes	Yes
6. Accountability	No	Yes	No
6.1 Notify data subject about data breach	-	x	-
6.2 Notify supervisory authority about data breach	-	x	x
6.3 Perform logs	-	x	x
7. Right to be informed	Yes	Yes	Yes
7.1 Inform the user how data is processed/stored	x	x	x
7.2 Inform the user how/where data is disseminated	x	x	x
8. Right of access	No	Yes	Yes
9. Right to rectification	No	Yes	Yes
10. Right to erasure	No	Yes	Yes
11. Right to restrict processing	No	No	No
12. Right to data portability	No	Yes	No
13. Right to object	No	Yes	Yes
14. Rights related to automated decisions	No	No	Yes

6.2.2 The Ground Truth

The ground truth is a term used in research to provide empirical evidence of information gathered by direct observation. In this study the ground truth was used in the validation of LINDDUN+ and to identify which threats were present in the system domain and had a direct relation with the GDPR principles and user rights. Further, the ground truth defines the expected result that a privacy and domain expert would end with after a session with LINDDUN+.

To define the ground of truth, the authors of this study created a DFD, the same DFD handed to the participants in the pilot workshop and the workshop with the students. This DFD was derived from the architectural views extracted from the system documentation provided by Volvo GTT. After creating the DFD, the authors analyzed each of the DFD elements individually and identified all the possible threats that had any relation with the GDPR provisions. Furthermore, the ground truth was developed through the experience and knowledge gained through this study by the authors and was validated during the pilot workshop conducted with the cyber-security specialist from Volvo GTT.

As preparation for the pilot workshop, a baseline for the results was created. However, after the pilot study and from the observations and assumptions made, this baseline was re-adapted. Specifically, the baseline tries to demonstrate the solid Ground Truth of all the correct threats that, in a perfect scenario, would be found by a practitioner when performing a system evaluation with LINDDUN+. The baseline was developed after a systematic and analysis of each individual element of the DFD, taking into consideration all the assumptions discussed under pilot study.

The development of the Ground Truth was only possible due to the knowledge acquired throughout the development of this study and the feedback received from the domain expert after the pilot study. Moreover, before conducting the final workshop a second validation of the derived system architecture was performed with a domain expert.

In total 34 privacy-related threats are considered to be present in the CACC system, where four principles constitutes the majority of the threats. The Accountability principle contains the highest number of threats, 11 threats in total, where most of the threats are related to insufficient logging, since logs are considered to be used to provide accountability. The Lawfulness Fairness and Transparency principle consists of 9 threats. The user right to be informed consists of 6 threats. The Purpose principle consists of 2 threats. One threat is related to data minimization. One threat related to retention time. One threat related to data erasure. Two threats related to the right to object. The number of threats for each principle and user right can be seen in Figure 6.1 below.

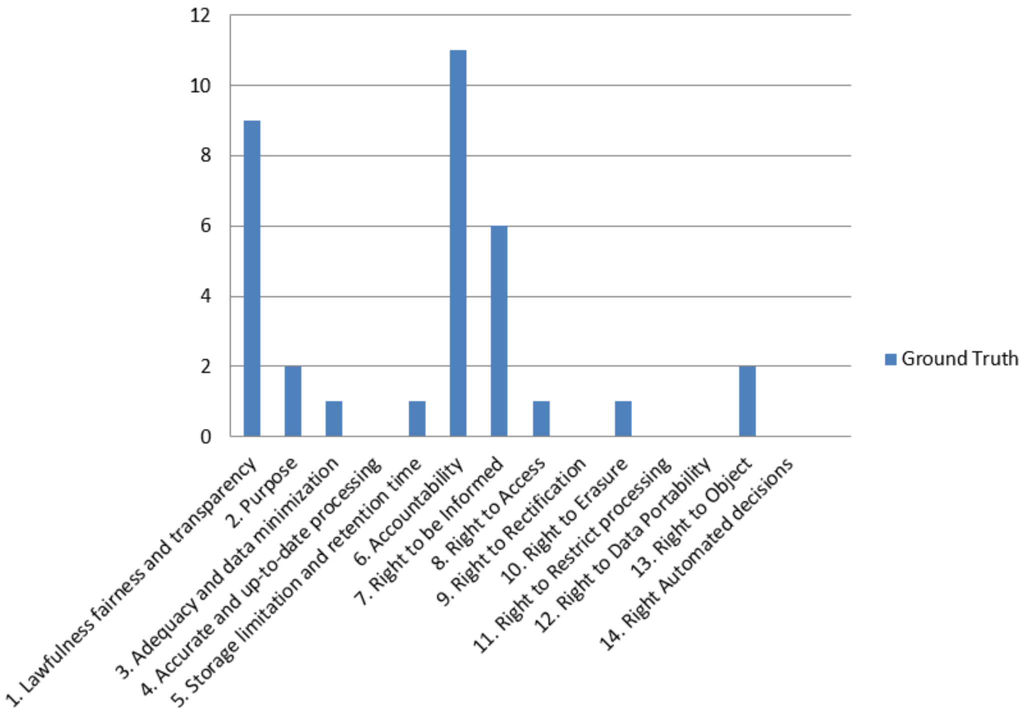


Figure 6.1: The Ground Truth of the CACC system

6.2.3 Workshop with Students

The final evaluation of LINDDUN+ was performed through a workshop conducted with students from Chalmers University of Technology and University of Gothenburg. In total three participants were part of the workshop. All the participants are master students in their final year of studies in Software Engineering. The participants have successfully completed the Advance Software Architecture and Requirements Engineering courses, therefore they were considered to be eligible to participate on the workshop. Since they were considered to have relevant base knowledge for applying a threat elicitation methodology in a software architecture. From this point on, the participants will be referred as Student 1, Student 2 and Student 3.

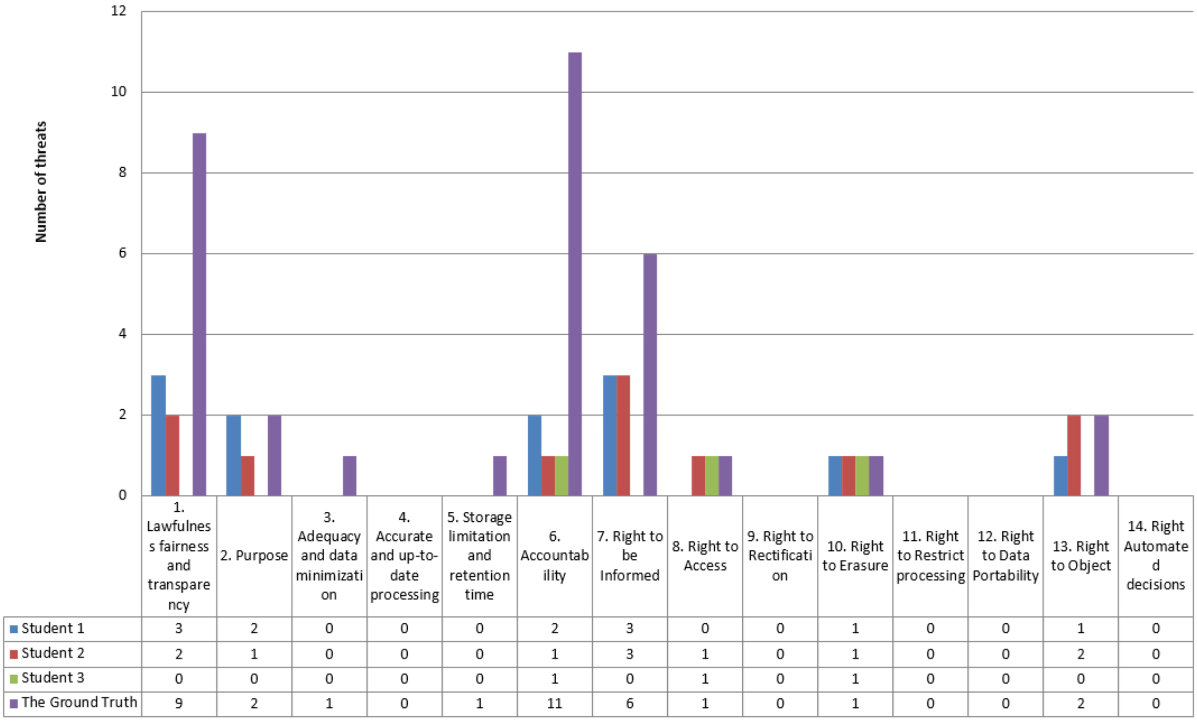
The final workshop was also divided in two sessions: 1) introduction session and 2) workshop session. The material provided was the same as the material used for the pilot study, with only few modifications on the system description document, since the assumptions made under the pilot study were also included. The workshop was conducted with the same time plan as the pilot workshop, with half an hour for revising the information held two days prior the workshop, as well as presenting the system, and three hours assigned for applying LINDDUN+. Hence, in total four hours were assigned for the total workshop.

6.2.3.1 Results from Students

The results obtained from the students showed that the participants were not able to identify all the threats. In this section only threats that were considered as a correct threat and possessed any relation to any of principles or user rights were included. A threat is considered to be correct when the documented threat is defined in a relevant context with an association to the right node of the specific threat tree. Also, the documented threat needs to be associated to the right DFD element.

The results from the workshop are presented in Figure 6.2. It shows the amount of threats found according to each principle. Furthermore, the number of threats for each principle and user rights found by the students were compared with the defined Ground Truth. Each of the students managed to accomplish the following result:

- **Student 1**: found 12 threats which were related to 6 principles or user rights.
- **Student 2**: found 11 threats which were related to 6 principles or user rights.
- **Student 3**: found 3 threats which were related to 3 principles or user rights.

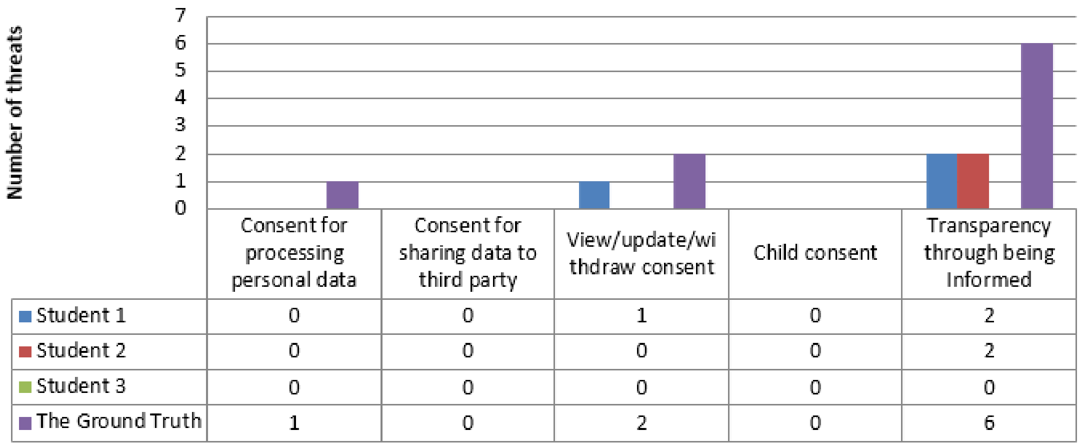


	1. Lawfulness fairness and transparency	2. Purpose	3. Adequacy and data minimization	4. Accurate and up-to-date processing	5. Storage limitation and retention time	6. Accountability	7. Right to be Informed	8. Right to Access	9. Right to Rectification	10. Right to Erasure	11. Right to Restrict processing	12. Right to Data Portability	13. Right to Object	14. Right Automated decisions
Student 1	3	2	0	0	0	2	3	0	0	1	0	0	1	0
Student 2	2	1	0	0	0	1	3	1	0	1	0	0	2	0
Student 3	0	0	0	0	0	1	0	1	0	1	0	0	0	0
The Ground Truth	9	2	1	0	1	11	6	1	0	1	0	0	2	0

Figure 6.2: Results from the workshop with the students

As was explained, in section 4.1.1 for the mapping of principles and user rights covered by LINDDUN in the first iteration, some of the principles were divided into smaller sub-principles. In order for a principle or user right, that contained smaller sub-principles, to be considered compliant all the sub-principles needed to be addressed by the methodology. The same reasoning holds for the results in this section. The difference is that the results are based on the Ground Truth. In this case, in order to the principle to be considered covered, all the sub principles included in the Ground Truth should also be addressed, accordingly. However, one cannot tell, by looking at Figure 6.2, if any of the principles that contains more specific sub-principles are covered by the methodology or not. To this reason a more detailed analysis of the results were made. The Figure 6.3, Figure 6.4, Figure 6.5 and Figure 6.6 present the more detailed result for each of the principles that contains sub-principles as shown in Table 6.1.

Figure 6.3 shows a detailed view of the threats found by each student that are related to the principle lawfulness fairness and transparency. For instance, Student 1 found 1 threat related to withdrawal of consent and 2 threats regarding transparency through the user being informed. Student 2 only found threats regarding transparency through being informed. And Student 3 did not find any threats regarding this principle. As seen, no participant managed to find all threats defined as the Ground Truth nor all sub-principles were addressed by combining the results from all the participants. To this reason, the lawfulness fairness and transparency principle is considered to be not covered.



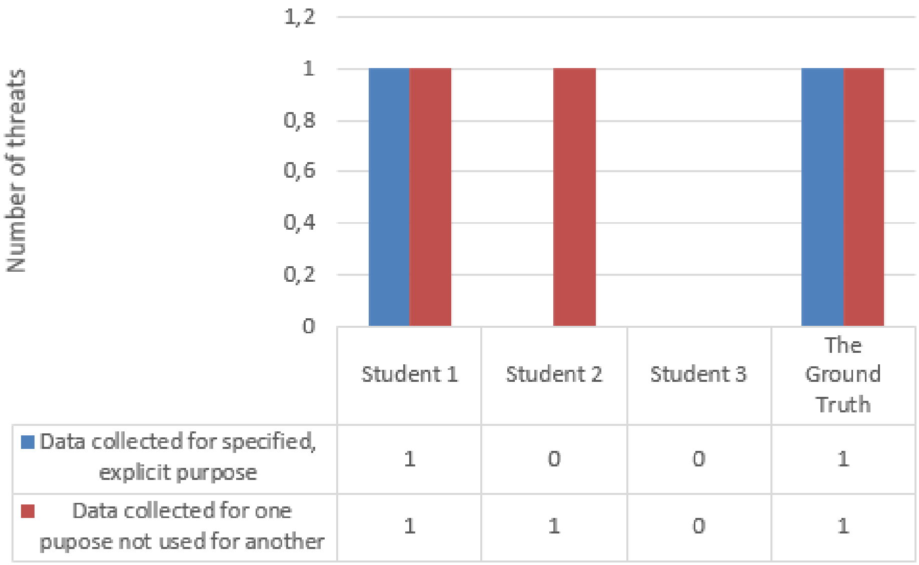
	Consent for processing personal data	Consent for sharing data to third party	View/update/withdraw consent	Child consent	Transparency through being Informed
🟦 Student 1	0	0	1	0	2
🟥 Student 2	0	0	0	0	2
🟩 Student 3	0	0	0	0	0
🟪 The Ground Truth	1	0	2	0	6

Figure 6.3: Detailed results regarding the lawfulness fairness and transparency principle

Figure 6.4 presents in more detail the threats found by the participants for the two sub-principles of the purpose principle. Student 1 found 1 threat each for the both sub-principles. Student 2 found 1 threat relating to purpose limitation. Student 3 did not find any threats regarding defined purposes in the software system. Since, the combined results from all the participants addressed threats for all the sub-principles, defined as the Ground Truth, this principle is considered covered.

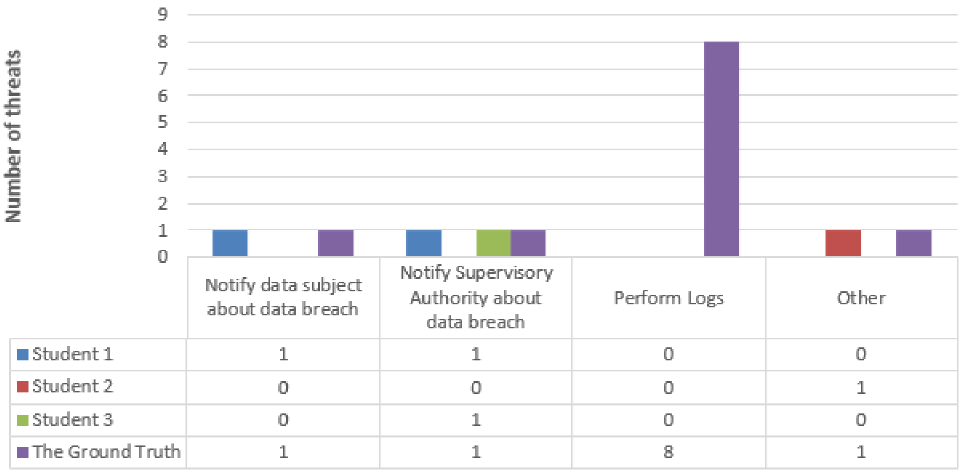
Looking at Figure 6.5 one can see all of the sub-principles for accountability were addressed by the participants except threats related to performing logs for data processing or storage. No participant managed to cover all the sub-principles defined as the Ground Truth. Since not all sub-principles for accountability were addressed by the participants the accountability is not considered to be covered.

For threats related to the right to be informed Student 1 and Student 2 managed to find three threats related to informing users about processing or storage of data by the system. Student 3 did not manage to identify any threats related to this user right. Since all the sub-principles, defined as the Ground Truth, is addressed by the



	Student 1	Student 2	Student 3	The Ground Truth
🟦 Data collected for specified, explicit purpose	1	0	0	1
🟥 Data collected for one pupose not used for another	1	1	0	1

Figure 6.4: Detailed results regarding the purpose principle



	Notify data subject about data breach	Notify Supervisory Authority about data breach	Perform Logs	Other
🟦 Student 1	1	1	0	0
🟥 Student 2	0	0	0	1
🟩 Student 3	0	1	0	0
🟪 The Ground Truth	1	1	8	1

Figure 6.5: Detailed results regarding accountability

participants, this user right is considered to be covered.

The empirical evaluation of LINDDUN+ resulted in a total coverage of 5 principles and user rights. This results in a 36 percent coverage of the principles and user rights under this study. Note, not all principles or user rights were considered applicable in the software system used in the workshop. All the principles and user rights covered can be seen in Table 6.2. When comparing the results from the participants with only the principles and user rights defined as the Ground Truth of the software system LINDDUN+ resulted in a coverage rate of 56 percent.

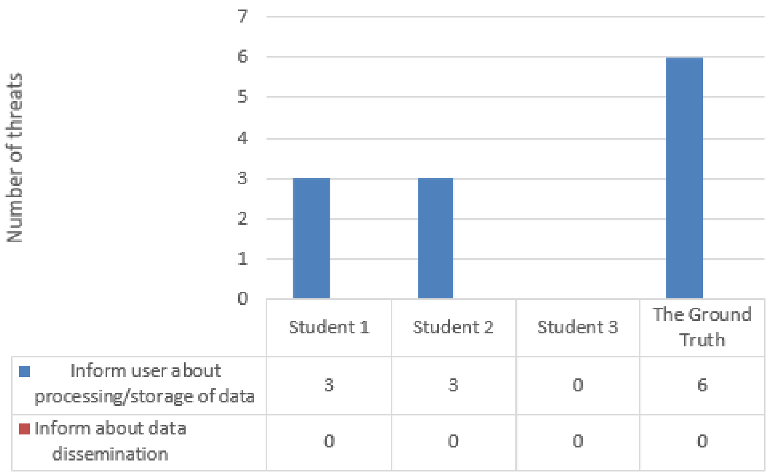


Figure 6.6: Detailed results regarding the right to be informed

Table 6.2: The results of LINDDUN+ workshop with students

Principle or User Right	Yes	No
1. Lawfulness fairness and transparency		X
2. Purpose	X	
3. Adequacy and data minimization		X
4. Accurate and up-to-date processing		X
5. Storage limitation and retention time		X
6. Accountability		X
7. Right to be Informed	X	
8. Right to Access	X	
9. Right to Rectification		X
10. Right to Erasure	X	
11. Right to Restrict Processing		X
12. Right to Data Portability		X
13. Right to Object	X	
14. Right related to automated decision making		X

Table 6.3 shows the overall results from both the first and second iteration of this

study. A total of 2 principles were covered by the LINDDUN methodology resulting in a coverage rate of 14 percent from the analysis in the first iteration. LINDDUN+ did cover 5 principles resulting in a coverage rate of 56 percent.

Table 6.3: Overview of Results from Iteration 1 and Iteration 2

	LINDDUN		**LINDDUN+**	
	Mapping Table	**Analysis**	**Mapping Table**	**Analysis**
No of identified principles	2	2	12	5
Percentage of Coverage	14%	14%	86%	56%
No of threats found	-	7	-	26

7

Discussion

The provisions and requirements of the GDPR are at times very abstract. The legislation requires organizations to implement these principles and user rights but provides very little guidance in how organizations should implement them. Neither a clear definition of what is considered enough in terms of compliance with the regulation is given. This leads organizations to evaluate and decide, based in their own interpretation of the regulation, their own means to comply with it. The same interpretation was needed in this study in order to access the effectiveness of LINDDUN and PIA and propose extensions to LINDDUN, resulting in the LINDDUN+. With that said, this study aimed to answer the following research questions:

RQ1: *How effective are the state-of-art threat analysis techniques like LINDDUN and PIA in identifying design issues related to GDPR compliance?*

RQ2: *Does an extended version of LINDDUN provide a more complete coverage of said issues?*

7.1 The Results

The first iteration, the analysis of LINDDUN versus PIA presented in Chapter 3, was used to answer the **RQ1**. The first part of the iteration, the comparison of LINDDUN and PIA that resulted in the mapped table 4.1, showed that when mapping the two methodologies with the provisions stated in the GDPR, LINDDUN presented a 14 percent of coverage rate out of the 14 analyzed provisions. While PIA presented a 71 percentage of coverage rate. In theory, since the mapping table was a conceptual effort to identify which of the provisions stated by GDPR were addressed by each of the methodologies, PIA can be considered more effective when compared with LINDDUN. This difference in performance was later confirmed after the empirical evaluation of LINDDUN and PIA, as presented in the second part of the analysis of LINDDUN versus PIA. The results of the empirical evaluation also showed 14 percent of coverage rate for LINDDUN and 71 percentage coverage rate for PIA. Under the empirical evaluation of the two methodologies, 7 threats were identified with LINDDUN, where 2 of them were linked to GDPR. With PIA 44 threats were identified, and 11 of them were related to GDPR. Thus, based on the results obtained from the first iteration, it is possible to say that PIA can be considered more effective than LINDDUN in identifying design issues related to GDPR compliance.

With the 7 threats from LINDDUN and the 44 threats from PIA have been discussed one can understand that there is a vast difference in the amount of threats found and documented between the two methodologies. However, it is of great interest to compare the results from LINDDUN and PIA in more detail. More specifically, it is of interest to discuss the extend of similarity of the threats found by the two methodologies, or if the two methodologies managed to find different types of threats. Looking at the threats found by LINDDUN they all relate to the first GDPR principle, Lawfulness Fairness and Transparency. The threats addressed the principle through specifying the unawareness, from the perspective of the data subject, of how the personal data is used by the system. This unawareness was present due to the inability to review the personal data (User Right to Access) that had been collected by the software system or the inability to review the system's privacy notice (User Right to be Informed). Although these threats are considered to be relevant, the type of threats are not different than the threats found by PIA for the same GDPR principle. However, the threats found and documented in LINDDUN can be considered to differ in one way, that each threat is related to a specific context of where in the software system the threat exists. This since each threat in LINDDUN is documented through misuse cases where each misuse case have a reference to a specific DFD element and threat tree node. The same degree of detailed context was not possible for the threats found in the PIA. Instead, it is believed that, in PIA, the practitioner is dependent on either his/her own expertise or possibility for including stakeholders with sufficient expertise in order to identify the same degree of the specific context of where in system such threat is present. In the other words, all found threats with LINDDUN are also found by the PIA, but the threats in LINDDUN have a closer connection to the specific areas of the system where it applies. This can be seen as being in favour for the LINDDUN methodology.

The work and results generated from the second iteration and presented in Chapter 4 and 5 were used to answer the **RQ2**. In order to answer this question LINDDUN+ was developed. As presented, LINDDUN+ extends LINDDUN in three ways: 1) with a PA-DFD, 2) extended threat trees, and 3) LINDDUN+ threat tree rules. All of the extensions aimed to provide a more complete coverage of the methodology in terms of compliance with GDPR. Before the empirical evaluation of LINDDUN+, another mapping table, Table 6.1 was created with a second attempt to map the extensions with the extracted principles and user rights from GDPR. In theory, LINDDUN+ showed a coverage rate of 86 percent, demonstrating that, again in theory, LINDDUN+ can be considered more effective than LINDDUN and PIA. The results gathered during the empirical evaluation and validation of LINDDUN+ showed a coverage rate of 56 percent, contributing to an increased coverage rate of 42 percent units. Important to note is, the coverage of 56 percent is calculated on the principles and user rights defined as the Ground Truth presented in section 6.2.2 and based on the specific domain of the case study. Hence, not all the 14 principles and user rights, addressed in this study, were included in this result, as they were not present in the case study domain. However, when applying LINDDUN+ on a architectural design of a system which includes threats to all the 14 GDPR provisions, it is believed that it can provide a 86 percentage of coverage rate.

Again, not all principles and user rights were present in the case study, due to the specific domain. Also, Student 1 managed to find 12 threats that covers 3 principles and 3 user rights. Student 2 found 11 threats, which also relates to 3 principles and 3 user rights. Student 3 found 3 threats. The threats from Student 3 are correlated to 1 principle and 2 user rights. Why the participants only found a subset of the real threats in the system is believed to be because 1) the limited time frame of the workshop 2) the combination of limited prior knowledge in terms of privacy, the LINDDUN methodology as well as the automotive domain. If the students were given an increased time frame it is highly believed that more threats would have been found, and thus a believed increase in the amount of principles and user rights addressed. Although, the work and results obtained in Chapter 4 and 5 still confirms that LINDDUN+ does provide a more complete coverage of the design issues related to GDPR compliance.

7.2 Sustainability of LINDDUN+ in other domains

It is interesting to discuss how sustainable the extended work of this study is when applied in another setting than in this study. In other words, how well the LINDDUN+ performs in domains other than the automotive. In the section 3.2.1 typical personally sensitive data attributes that all can, either alone or combined with each other, be classified as PII has been explained. Not all attributes defined in the Table 3.1 of section 3.2.1 such as medical record numbers or health plan beneficiary number might not be present in the automotive domain but more in the health care domain. Hence, the work of LINDDUN+ has not been evaluated with a use of such data attributes but only with data attributes that is relevant to the automotive domain. Instead what is commonly sensitive in the automotive domain is the heavy use of location based positioning of the trucks and the possibility for data controllers (organizations) to understand driver behavior. For instance, that a driver repeatedly visits a certain place at a certain time or driver violating speed limits at certain routes. However, if LINDDUN+ would be applied in a domain, other than the automotive, with data attributes that do not existt in the automotive domain, LINDDUN+ is believed to yield the same degree of significant result as if it would be applied in an automotive context. This due to, what is considered as PII is not defined by the LINDDUN+ methodology itself, instead this responsibility is given to the practitioner of the methodology. For instance, during the development of the PA-DFD, explained in 5.3, the practitioner has to have an understanding of what is considered as sensitive data in the given domain which the software system belongs to. This could for instance be achieved by having a data classification at hand. Of course the defined sensitive data in such data classification will vary between different domains. However, the content of such data classification serves one purpose; to give the practitioner an understanding of how to prioritize more privacy sensitive areas before others. Since the responsibility to define the scope of the sensitive data is given to engineers that is familiar with the given domain, it is believed that the work of LINDDUN+ would perform equally good into other domains other than the automotive as well. However, it is still of interest to perform further analysis of the performance of LINDDUN+ in such other domains.

The main purpose with two of the contributions of this study, the extended threat trees and the developed threat tree rules, are not to provide all possible attack paths or threats in a software system. Instead, their purpose is to showcase the most common threats that a software system might be vulnerable to. Hence, they has been developed with the intention of being as generic as possible without any context specific details correlated to them. Therefore, it is believed that the defined attack paths of the extended threat trees and the developed threat tree rules are not domain specific but are also sustainable for other domains.

7.3 Interpretation of the GDPR

The first principle of GDPR, lawfulness fairness and transparency, covers a high amount of content. This principle often includes provisions that can be hard to understand, more specifically where the principle applies, what it means, or how one comply with it. The lawfulness part of this principle relates to the general sense as the usage of data is not violating any law, whether criminal or civil. This part has not been addressed in this study due to the amount of work this would require, specially to understand the jurisprudence context. Additionally, this would also lead to work exceeding the software engineering subject domain of this study. Instead, the focus of this principle is related to the fairness and transparency to the data subject. The fairness and transparency have been considered to be closely tied to each other. Often, when providing transparency to the data subjects fairness is also included. To conclude, how consents are implemented and provided to the users have been considered as a way to provide fairness. Moreover, in order to cover the transparency principle the user's right to be informed was also considered. Although, threats found under this user right have not only been considered correlated to the transparency provision but also fairness.

To have clearly defined purposes over processing and not processing for another purpose than was originally defined of personal data regarding the users, as the second principle of GDPR stipulates, is not as obvious as one might think at first glance. Various exceptions exist which leads organizations to pursue processing of personal data even though it might, from the data subjects point of view, be considered unlawful. Processing for the purpose to serve the public interest, scientific or statistical purposes are three exceptions of provisions from the GDPR which get prioritized over other provisions which enforces the user's rights. Where the line is drawn to determine if a certain processing serves such interests or not is a debate that can last for ages. Thus, throughout this study, to understand the boundaries for if a processing is classified into any of such exceptions was considered a hassle. This study describes the different general exceptions where processing of data can be allowed, but does not describe at what specific scenario this falls into. This due to the variety of input parameters, which differ between contexts, that can affect such classification. Instead, this has to be decided together with experts from the system's domain and stakeholders, as well as with legal experts.

The accountability principle is another vague provision from the regulation. Fundamentally, it stipulates that organizations shall be able to provide enough evidence to show compliance to the remaining principles and user rights of the regulation. Not only to the supervisory authority but also to the data subjects. Again, the regulation does not define a clear and specific rule which tells enough measures and how an organization can be compliant to this principle. This resulted in the same interpretation, as was described for the first principle. Since this study relates to the informational technology context, the following measures have been considered relevant for complying with the accountability principle: logging, ensuring notification to authority and affected data subjects, introduction of a DPO, ensuring data subjects possibility for accessing the data stored about him or her, as well as access to privacy notice.

7.4 The significance of the PA-DFD

The aim of the privacy-aware DFD is to increase the awareness of a future practitioner of LINDDUN over how privacy data is flowing in the software system. The benefit of writing notes on each PA-DFD element is that the practitioner does end up with an explicit documentation over what data types are used where and also how this data might change at a later stage in the software system. Thus, the practitioner can, by using a data classification specification, understand at what areas of the software system does process, store or disseminate sensitive data about a data subject. It is believed, this contribution is promising in terms of providing significant value to the LINDDUN methodology. The reason is because the analyst does now perform a more explicit analysis towards privacy in the first step of the methodology. Also, a business-oriented DFD does not seek to contribute solely and specifically privacy issues, but instead contribute to a broader purpose. To explain how data flows in a software system. Hence, LINDDUN did suffer improvement potential in terms of considering privacy issues already in its first step. An improvement potential that is considered to be addressed with the PA-DFD presented in the work.

7.5 The Participants

With the results known from the section 6.2.3 it is possible to discuss the degree of significance of the extended work presented in this study. Although, these results represent the coverage of the methodology, by only looking at the data represented in that section from one perspective only. One cannot understand possible external factors that might have affected the outcome from the participants. To understand such other external factors, that could help clarify any unexpected results, a questionnaire was provided to the participants after they finished their LINDDUN+ session. In the questionnaire the participant could reflect on their experience of applying the methodology. The feedback from the students filling in the questionnaire is described below.

Student 1 posses 2-3 years of experience from working in industry but did not posses any prior experience from working with risk assessments.

Student 2 had 2-3 years of experience from working in industry. The student did have five years of experience from working with risk assessment prior the participation of the study.

Student 3 did have general work experience of five years or more. The student did posses no prior experience regarding risk assessment analysis had been conducted.

Thus, Student 1, Student 2 and Student 3 all had prior experience from industry. However, none of the students had prior experience in working with privacy within industry. Also, none of the students said they were unfamiliar with the provisions of the GDPR. The reason behind this is most likely due of the prior information session held two days in advance of the LINDDUN+ workshop session. All the students did explain they had experience with STRIDE. Since the initial steps of the LINDDUN+ methodology does have similarities to STRIDE this can be seen as an advantage for the students when applying the LINDDUN+.

Looking at the difference between the number of threats found by the students and the number of threats defined as the Ground Truth, the results can be considered a bit surprising. A higher number of found threats by the students was expected. Although, all the students did explain that they did find the methodology hard to understand. Specifically, all the students did share they were either uncertain or very uncertain that they had found all threats in the software system. The reason behind this is believed to be due the non-existing prior knowledge regarding privacy by all of the students. Hence, it is believed that if a similar study would be conducted with participants with a higher degree of prior knowledge regarding privacy, the participants would be more certain about their results being correct.

7.6 The Threat Tree Rules

As has been explained previously in this work, the threat tree rules serve the purpose to accomplish automation of LINDDUN, an interest that requires work which exceeds this study. Despite this, the rules were included to the students for exploratory reasons, to see if they could help the participants of this study, already, to relate privacy threats in the applied software system. None of the students did consider the third contribution of this study, the developed rules, helpful. Looking at the amount of new information given to the students, apart from the rules, it can be considered rather vast. Again, this due to the unfamiliarity of the students regarding privacy, the introduction of the LINDDUN+ methodology and also risk management in the software domain in general. If the participants would possess knowledge in these three areas prior the participation in this study a belief is they would have a better ground to stand on. Hence, the belief is they would consider the rules more helpful. Again, this was although not the main purpose behind the development of the threat tree rules.

For both the first and second iteration, a principle or user right is considered to be covered by the methodology as soon as a threat, from any of the participants, relates to it. This reasoning is legitimate since the purpose of this study is not to analyze the performance of the methodologies, but only to understand the number of the GDPR provisions addressed by the methodologies under analysis.

7.7 Threats to Validity

This section describes the threats to the validity of this study. They are divided into three subcategories: internal validity, external validity and construct validity. Threats to internal validity implies errors in the study that can jeopardize the significance and correctness to drawn conclusions from the generated results of a study. This could happen when a study is a victim of having biased participants, authors or control units. External validity is the threats related to incorrect generalizations of the drawn conclusions presented to the public reader of the study. In other words, the certainty to state that the results represents a generic context and not solely represents a context with the same restricted environment and fixed parameters as under this study. Construct validity implies the error of having incorrect measurements in a study. Hence, are the measurements chosen in the study able to answer the question the study tries to answer.

7.7.1 Threats to Internal Validity

The results from the first iteration was generated by the two authors of this study. Where one author performed a session of the original version of LINDDUN, while the other author conducted a session of PIA. Neither of the authors had working experience with the two methodologies prior the sessions. However, enough knowledge regarding what privacy is, what is sensitive regarding privacy and how violations to privacy can occur was gained. Since the authors already possessed this knowledge prior their evaluations of the methodologies most likely the results could have been biased compared to the results provided by external participants in the conducted sessions.

The use case which has been used for all the evaluation sessions in this study (both in the first iteration as well as in the second iteration) has, at the time of the study, no defined architecture. Hence, the authors of the study were compelled to create an architecture of the system in order to perform LINDDUN and LINDDUN+ analysis. The authors did not participate in prior development sessions of the system specification at Volvo Group Trucks Technology. It is possible that the authors possessed limited essential knowledge and information regarding key aspects for software systems in the automotive domain. A way to deal with this limited knowledge was to create assumptions and consider the developed system architecture from an overview and general view, since the system design was developed over the use cases and specifications provided in the system specification document. Another threat to internal validity was the use of the same DFD during both empirical evaluations, in the first and second iteration, since it was derived from the system architecture created by engineers (the authors of this study) with limited domain knowledge.

As was mentioned in section 6.2.2, for validating the results from the second iteration, a baseline was created. The baseline served as a definition of the Ground Truth that specified all applicable threats in the software system which the LINDDUN+ was applied on. Thus, the perfect scenario from the evaluation of LINDDUN+ would be that all participants did perform the exact same results as the Ground Truth. Ideally, the development of the Ground Truth should be made by a privacy and domain expert. This was however not possible, due to limited resources. Instead, the authors of this study was given this responsibility. Although, the creators of the Ground Truth did have some knowledge in privacy and automotive industry, from conducting this study, the knowledge can still be considered limited. Knowledge that, if the creators of the Ground Truth would have had, could change the perception and reasoning behind the included threats of the baseline. For this reason, one can argue the study has a threat to internal validity.

7.7.2 Threats to External Validity

The participants of this study are students and do not possess the most ideal knowledge background, as a privacy expert that has enough working experience as a software architect or is familiar with extensive risk assessments in a software architectural context. To this reason, it is not possible to draw general conclusions from the results. Instead the results of this study represent a specific setting with the given conditions and scope characteristic to the scope of this study.

Important to note, the software system which was used for applying the empirical evaluations of the LINDDUN, LINDDUN+ and PIA methodology did not include threats which are related to all the principles and user rights under study of this work. Hence, it is not possible to draw any conclusions that a certain principle or user right is not covered by the LINDDUN+ methodology just because the results presented indicates so. For instance, the software system did not contain any threats related to data portability or rights related to automated decision making. Thus, it is not possible to conclude that LINDDUN+ does not cover threats related to these user rights. Even though the results of this study shows a coverage rate of 56 percent over the principles and user rights, it is believed that an increased coverage over the principles and user rights would be experienced with a system containing threats relating to all principles and user rights. Moreover, this work includes threats that are not feasible to evaluate.

7.7.3 Threats to Construct Validity

Threats to construct validity of this study concerns the abstraction level of the principles and user rights of the GDPR. Since some of the principles of the regulation can be interpreted as very broad and abstract, it can be hard to determine if the methodologies under study in this work do cover these principles and user rights or not. To this reason, the authors of this study did manually go through the regulation and together determined the appropriate abstraction level. For instance, the first principle of the regulation, lawfulness fairness and transparency, does relate to several other provisions of the regulation. Also, the right to be informed does provide transparency through telling the data subject how their information is used. Of course, the possibility of other authors, conducting a similar mapping as in this study, could end up in a more or less detailed correlation between the principles and rights compared to this study.

8
Conclusion

The focus of this study has been to propose a methodology to validate compliance to the GDPR. In order to accomplish this, different state-of-the-art threat analysis techniques have been studied. Due to limited time and resources, two of these techniques have been chosen for a closer analysis: LINDDUN and PIA, which are two well known privacy methodologies among the privacy community. The methodologies were then analyzed and compared to the GDPR provisions, in order to identify design issues related to compliance with the regulation. Finally, these design issues were addressed, resulting in the LINDDUN+ methodology.

This study provides two main scientific contributions. The first contribution involves the analysis of LINDDUN and PIA consisting of two parts: the development of a mapping table where LINDDUN and PIA were systematically mapped to fourteen principles and user rights extracted from GDPR; and an empirical evaluation of the two methodologies. The second contribution includes the extension of the LINDDUN methodology in order to provide a more complete coverage of the regulation. To this contribution, three areas has been at focus; a proposal for using a more privacy-aware DFD in the initial steps of LINDDUN, an extension of two threat trees and development of threat tree rules for the LINDDUN methodology. Together they form an extended LINDDUN, referred to as LINDDUN+.

This study has been performed in collaboration with Volvo Group Trucks Technology, under the HoliSec project. The HoliSec project is a project that includes a vast and solid work regarding cybersecurity in the automotive domain. Until today, the degree of privacy-related work in the HoliSec project is limited, thus the outcome of this study serves as an entry point for introducing privacy to Volvo Group Trucks Technology and the HoliSec project.

This study aims to answer the two following research questions:

- **RQ1:** *How effective are state-of-the-art threat analysis techniques like LINDDUN and PIA in identifying design issues related to GDPR compliance?*

- **RQ2:** *Does an extended version of LINDDUN provide a more complete coverage of said issues?*

The results from the first iteration showed that LINDDUN provided a coverage rate of 14 percent of all the principles and user rights considered from GDPR. While PIA provided a coverage rate of 71 percent of all the principles and user rights. This showed that PIA can be considered more effective than LINDDUN, thus answering **RQ1**.

The results from second iteration showed that LINDDUN+ provided a theoretical coverage of 86 percent when mapped to the extended work of this study. From the empirical evaluation, the results showed that LINDDUN+ provided a coverage of 56 percent, implying an increase of 42 percent units compared to the original LINDDUN. Thus evidence has been provided that answers **RQ2**, that an extended version of LINDDUN does provide a higher coverage to GDPR provisions.

8.1 Future Work

The work of this study includes an assessment of the LINDDUN methodology for the purpose of increasing the coverage of principles and user rights of the GDPR. The evaluation of the work, developed in this study, was conducted together with students and the size of the participant group is considered limited. For this reason, it would be of interest to conduct another study for the purpose to validate the work developed in this study with an increased number of participants. Further, the participants should be privacy and domain experts with the system that is under evaluation with the methodology.

Since the software system used in this work for applying the methodologies under analysis did not contain threats related to all the principles and user rights, it is of interest to conduct a similar study with a software system containing these principles and user rights.

8.1.1 Development of the Threat Trees

The work regarding the extensions to the threat trees has been performed with the attempt of covering as many of the most fundamental provisions of the GDPR. The reason behind why to only include the two last threat trees of LINDDUN in this work has been explain in section 4.1.1. The reason was, they were considered the only trees in LINDDUN that aimed at addressing compliance to the provisions of GDPR. A subset of the remaining trees of LINDDUN does although address some of the provisions stated by GDPR. This is not for the purpose of being compliant to the regulation, but to enforce the privacy property it is related to, e.g. linkability. Interesting future work would be to consider adding such privacy enforcement to the policy and consent non-compliance tree. This could be to consider if concepts such as data minimization should also be included in the policy and consent compliance tree or not.

Not all the provisions of the regulation were possible to cover in this study. Previously, in this paper the argument was given that some of the provisions of the regulation, such as data minimization, were already included in the LINDDUN methodology. Although, they were not placed in the two threat trees under this study. For this reason it was considered that these provisions were not addressed by the methodology, because they were not included for the purpose of providing compliance to the regulation. The authors of this study advocate future work for continued development of the Policy and Consent-Non compliance threat tree to enhance even more its coverage.

8.1.2 Rules for the Threat Trees

The work which has been accomplished for serving the third proposal in section 5.2, development of rules for the threat tree nodes, can be considered limited. The intention behind their development was to contribute with value for a potential future automated software tool. As for the rules developed in this work, they cannot be used in an automated software context. In order to accomplish such interest they have to be developed with a parsable modeling query rule language, for instance OCL. Thus, this can be addressed in a potential future work. Ideally, each node of the threat trees of LINDDUN should be associated with at least one rule. This is another potential work which fits the scope for future work.

Bibliography

[1] DistriNet Research Group. LINDDUN Privacy Threat Modeling Website. https://linddun.org/index.php, 1998.

[2] M. Kaur, S. Kaur, and G. Singh. Vehicular AD HOC Networks. *Journal of Global Research in Computer Science*, 3(3), 2012.

[3] A. Pfitzmann and M. Hansen. A Terminology for Talking About Privacy by Data Minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management. 2010.

[4] A. F. Westin. Privacy and Freedom. *Washington and Lee Law Review*, 1968.

[5] The European Parliament. Regulations (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). , 26 April 2016.

[6] A. Cavoukian. Privacy by Design. https://www.ipc.on.ca/wp-content/uploads/Resources/pbd-implement-7found-principles.pdf, 2009.

[7] The Verge. Facebook and Google hit with $8.8 billion in lawsuits on day one of GDPR. https://www.theverge.com/2018/5/25/17393766/facebook-google-gdpr-lawsuit-max-schrems-europe, 2018.

[8] M. Deng, K. Wuyts, R. Scandariato, B. Preneel, and W. Joosen. A Privacy Threat Analysis Framework: Supporting the Elicitation and Fulfillment of Privacy Requirements. *Requirements Engineering*, 16(1):3–32, 2011.

[9] C. Kallaniatis, E. Kavakli, and S. Gritzalis. Addressing Privacy Requirements in System Design: The PriS Method. *The Journal of Systems and Software*, 96:241–255, 2008.

[10] M. C. Oetzel, S. Spiekermann, I. Grüning, H. Kelter, and S Mull. Privacy Impact Assessment Guideline for RFID Applications. *Bundesamt für Sicherheit in der Informationstechnik*, 2011.

[11] Volvo Technology AB. HoliSec: Holistic Approach to Improve Data Security. https://www.vinnova.se/en/p/holisec-holistiskt-angreppssatt-att-forbattra-datasakerhet/, 2016.

[12] T. Antignac, R. Scandariato, and G-Schneider. A Privacy-Aware Conceptual Model for Handling Personal Data. *Springer International Publishing*, 2016.

[13] T. Antignac, R. Scandariato, and G-Schneider. Privacy Compliance via Model Transformations. *International Workshop on Privacy Engineering - IWPE'18*, 2018.

[14] D. J. Solove. Conceptualizing Privacy. *California Law Review*, 90(4), 2002.

[15] S. D. Warren and L. D. Brandeis. The Right to Privacy. *Harvard Law Review*, 4(5):193–220, 1890.

[16] D. J. Solove. A Taxonomy of Privacy. *The University of Pensylvania Law Review*, 154(3):124–131, January 2006.

[17] Information Commissioner's Office. The Guide to Data Protection. `https://ico.org.uk/media/for-organisations/guide-to-data-protection-2-10.pdf`, 8 February 2018.

[18] B. Krishnamurthy and C. E. Wills. On the Leakage of Personally Identifiable Information Via Online Social Networks. *ACM SIGCOMM Computer Communication Review*, 40(1), 2009.

[19] E. McCallister, T. Grance, and K. Scarfone. Recommendations of the national institute of standards and technology. In *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)*. NIST Special Publication 800-122, 2010.

[20] U.S. Department of Health Human Services. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule . *Guidance on De-identification of Protected Health Information*, 2012.

[21] P. M. Schwartz and D. J. Solove. The PII Problem: Privacy and a New Concept of Personally Identifiable Information. *Berkeley Law Scholarship Repository*, 90(4), 2011.

[22] M. Barbaro and T. Zeller Jr. A Face Is Exposed for AOL Searcher No. 4417749. *New York Times*, 2006.

[23] A. Narayanan and V. Shmatikov. Privacy and Security Myths and Fallacies of "Personally Identifiable Information". *Communications of the ACM*, 53(6):24–26, 2010.

[24] Information Commissioner's Office. Data controllers and data processors: what the difference is and what the governance implications are. `https://ico.org.uk/media/for-organisations/documents/1546/data-controllers-and-data-processors-dp-guidance.pdf`, 2014.

[25] W. Y. Tham Y. B. Leau, W. K. Loo and S. F. T. Software Development Life Cycleb AGILE vs Traditional Approaches . *International Conference on Information and Network Technology (ICINT 2012)*, 2012.

[26] ISO 27001. Information Technology — Security Techniques — Information Security Management Systems — Requirements. , 2005.

[27] Office of Research. Privacy and Confidentiality. `https://www.research.uci.edu/compliance/human-research-protections/researchers/privacy-and-confidentiality.html#confidentiality`, Accessed: 16 February 2018.

[28] United Kingdom Parliament. Data Protection Act 1998. `http://www.legislation.gov.uk/ukpga/1998/29/data.pdf`, 1998.

[29] Allen Overy. The EU General Data Protection Regulation. `http://www.allenovery.com/SiteCollectionDocuments/Radical%20changes%20to%20European%20data%20protection%20legislation.pdf`, Accessed: 15 February.

[30] Information Commissioner's Office. Preparing for the General Data Protection Regulation (GDPR): 12 steps to take now. `https://ico.org.uk/media/1624219/preparing-for-the-gdpr-12-steps.pdf`, Accessed: 14 February 2018.

[31] Gabriel Maldof. The Risk-Based Approach in the GDPR: Interpretation and Implications. *iapp*, 2017.

[32] Information Commissioner's Office. Guide to the General Data Protection Regulation (GDPR). `https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/principles/`, Accessed: 27 February 2018.

[33] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian Network Classifiers. *Machine Learning*, 29(1):131–163, 1997.

[34] JH. Hoepman. Privacy Design Strategies. *ICT Systems Security and Privacy Protection*, pages 1–12, 2013.

[35] F. Buschmann, R. Meunier, H. Rohnert, and P. Sommerlad. *Pattern-Oriented Software Architecture Volume 1: A System of Patterns.* John Wiley Sons, 1996.

[36] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software.* Addison-Wesley, 1994.

[37] M. Hafiz. A collection of Privacy Desgin Patterns. *Proceedings of the 2006 conference on Pattern languages of programs (PLoP '06)*, 2006.

[38] A. Pfitzmann and M.Waidner. Advances in cryptology proceedings of a workshop on the theory and application of cryptographic techniques (eurocrypt '85). In *Networks Without User Observability — Design Options*, pages 245–253. Springer, 1985.

[39] M. Hafiz. A Pattern Language for Developing Privacy Enhancing Technologies. *Software: Practice and Experience*, 43(7):769–787, 2010.

[40] J. Borking. Der identity-protector. *Datenschutz und Daten-sicherheit*, 1996.

[41] G. W. van Blarkom, J. J. Borking, J. J. Borking P. Verhaar. PET. In G. W. van Blarkom, and J. G. E. Olk. Handbook of privacy and privacy-enhancing technologies - the case of intelligent software agents. In *Chapter 3*, pages 33–54. College bescherming persoonsgegevens, 2003.

[42] L. Bass, P. Clements, and R. Kazman. Relationship between tactics and patterns. In *Software Archtitcture in Practice*, Third generation, pages 200–209. Addison-Wesley, 2013.

[43] M. K. Reiter. Crowds: Anonymity for Web Transactions. *ACM Transactions on Information and System Security (TISSEC)*, 1(1):66–92, 1998.

[44] D. M. GoldSchlag, M. G. Reed, and P. F. Syverson. Hiding Routing Information. *International Workshop on Information Hiding*, 1174:137–150, 1996.

[45] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The Second-Generation Onion Router. *Proceedings of the Usenix Security Symposium*, 2004.

[46] The World Wide Web Consortium. Platform for Privacy Preferences (P3P) Project. `https://www.w3.org/P3P/`, Accessed: March 1 2018.

[47] L. F. Cranor. *Web privacy with p3p.* O'Reilly, 2002.

[48] L. Sweeney. k-ANONYMITY: A MODEL FOR PROTECTING PRIVACY. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):557–570, 2002.

[49] C. Dwork, K. Nissim F. McSherry, and A. Smith. Calibrating Noise to Sensitivity in Private Data Analysis. *Theory of Cryptography Conference*, 3876:265–284, 2006.

[50] B. Nelson and T. Olovsson. Introducing Differential Privacy to the Automotive Domain: Opportunities and Challenges. *Proceedings of the 2nd International Workshop on Vehicular Security (V-SEC 2017)*, 2017.

[51] D. L. Chaum. Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. *Communications of the ACM*, 24(2):84–90, 1981.

[52] M. L. Damiani, E. Bertino, and C. Silvestri. The PROBE Framework for the Personalized Cloaking of Private Locations. *Transactions on Data Privacy*, 3:123–148, 2010.

[53] K. Wuyts, R. Scandariato, and W. Joosen. Empirical evaluation of a privacy-focused threat modelingmethodology. *The Journal of Systems and Software*, 96:122–138, 2014.

[54] F. Swiderski and W. Snyder. *Threat Modeling.* Microsoft Press, 2004.

[55] R. Clark. Privacy Impact Assessment: Its Origins and Development. *Computer Law Security Review*, 25(2):225–239, 2009.

[56] Information Commissioner's Office. Conducting Privacy Impact Assessments Code of Practice. `https://ico.org.uk/media/for-organisations/documents/1595/pia-code-of-practice.pdf`, 2014.

[57] N. Notario, A. Crespo, Y-S Martín, J. M. Alamo, D. Métayer, T. Antignac, A. Kung, I. Kroener, and D. Wright. PRIPARE: Integrating Privacy Best Practices into a Privacy Engineering Methodology. *IEEE CS Security and Privacy Workshops*, 2015.

[58] Commission nationale de l'informatique et des libertés (CNIL). Methodology For Privacy Risk Management. , Paris, June 2012.

[59] National Bureau of Standards Federal Information Processing Standards Publication 31. Washington: U.S. Department of Commerce. Guidelines FOR AUTOMATIC DATA PROCESSING PHYSICAL SECURITY AND RISK MANAGEMENT. , June 1974.

[60] S. Spiekermann and L. C. Cranor. Engineering Privacy. *IEEE TRANSACTIONS ON SOFTWARE ENGINEERING*, 35(1), 2009.

[61] Education U.S. Department of Health and Welfare. The Code of Fair Information Practices. `https://simson.net/ref/2004/csg357/handouts/01_fips.pdf`, 1973.

[62] Federal Trade Comission. PRIVACY ONLINE: A REPORT TO CONGRESS. , 1998.

[63] C. Jensen, J. Tullio, C. Potts, and E. D. Mynatt. STRAP: A Structured Analysis Framework for Privacy. 35(1), 2009.

[64] J. Luna, N. Suri, and I. Krontiris. Privacy-by-Design Based on Quantitative Threat Modeling. *Risk and Security of Internet and Systems (CRiSIS) 7th International Conference*, 2012.

[65] P. Kleberger, T. Olovsson, and E. Jonsson. Security Aspects of the In-Vehicle Network in the Connected Car. *Intelligent Vehicles Symposium (IV)*, page 528–533, Baden Baden: IEEE, 2011.

[66] S. ZeadallyEmail, S. Hunt, Y-S. Chen, A. Irwin, and A. Hassan. Vehicular ad hoc networks (VANETS): status, results, and challenges. *Telecommunication Systems*, 50(4), 2010.

[67] J-P. Hubaux, S. Capkun, and J. Luon. The Security and Privacy of Smart Vehicles. *IEEE Security Privacy*, 2(3):49–55, 2004.

[68] S. Biswas, R. Tatchikou, and F. Dion. Vehicle-to-Vehicle Wireless Communication Protocols for Enhancing Highway Traffic Safety. *IEEE Communications Magazine*, 44(1):74–82, 2006.

[69] F. Schaub, Z. Ma, and F. Kargl. Privacy Requirements in Vehicular Communication Systems. *International Conference on Computational Science and Engineering*, 4:139–145, 2009.

[70] "Trucks talking to each-other | Volvo Group", Volvogroup.com, 2018. [Online]. Available: http://www.volvogroup.com/en-en/news/2017/oct/trucks-talking-to-each-other-in-multi-brand-platooning-project.html. [Accessed: 03- May-2018].

[71] The Impact of Cooperative Adaptive Cruise Control on Traffic-Flow Characteristics. *IEEE Transactions on Intelligent Transportation Systems*, 7(4), 2007.

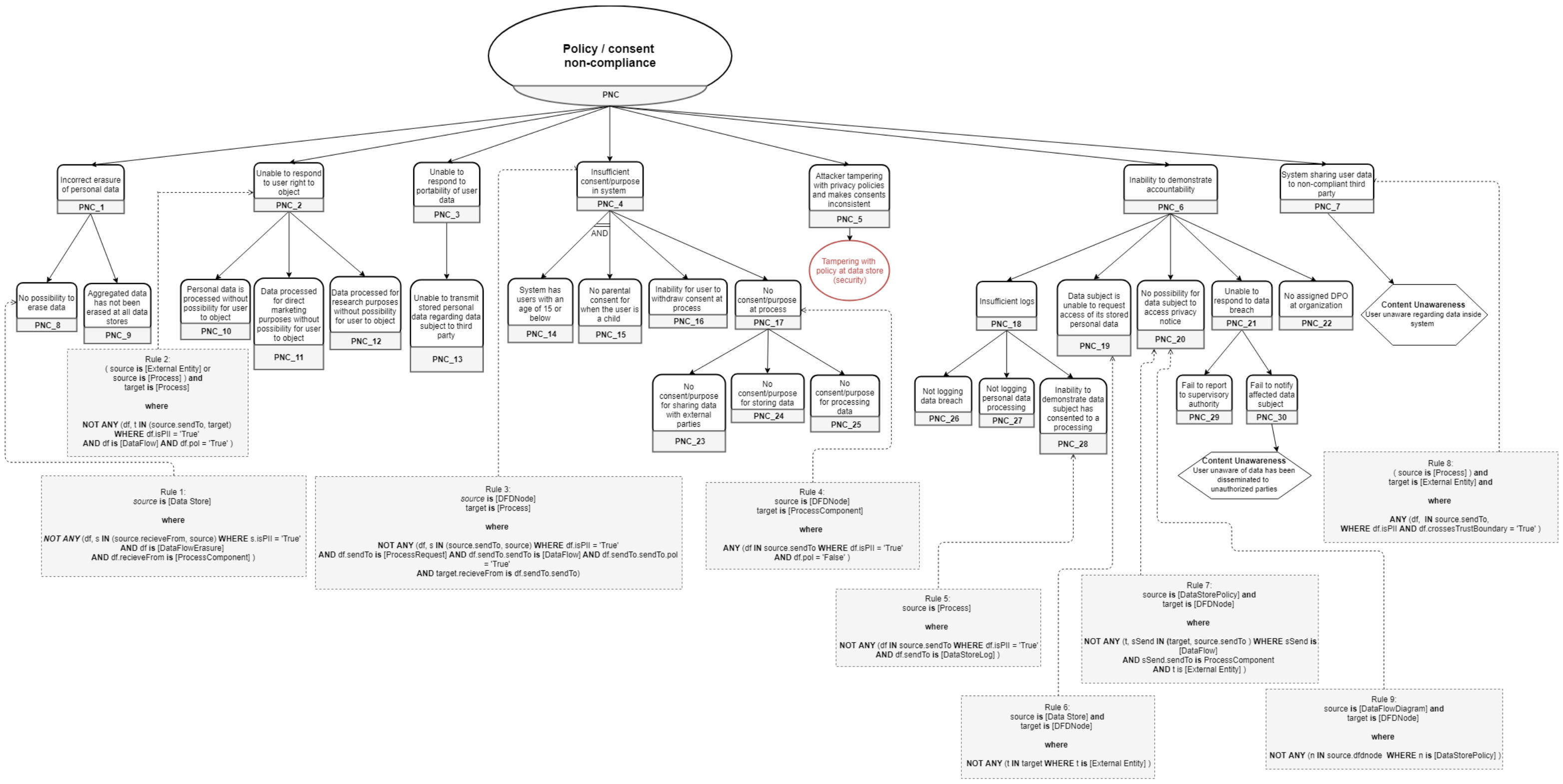
[72] C. Nowakowski, S. E. Shladover, X.-Y. Lu, D. Thompson and A. Kailas. Cooperative Adaptive Cruise Control (CACC) For Truck Platooning: Operational Concept Alternatives. *California PATH Program Institute of Transportation Studies University of California, Berkeley*, 2015.

[73] Information Commissioner's Office. Privacy notices, transparency and control: A code of practice on communicating privacy information to individuals. 2016.

A

Proposed extensions for LINDDUN

A.1 Policy and Consent Non-Compliance



Policy / consent non-compliance
PNC
Incorrect erasure of personal data
PNC_1
Unable to respond to user right to object
PNC_2
Unable to respond to portability of user data
PNC_3
Insufficient consent/purpose in system
PNC_4
AND
Attacker tampering with privacy policies and makes consents inconsistent
PNC_5
Tampering with policy at data store (security)
Inability to demonstrate accountability
PNC_6
System sharing user data to non-compliant third party
PNC_7
No possibility to erase data
PNC_8
Aggregated data has not been erased at all data stores
PNC_9
Personal data is processed without possibility for user to object
PNC_10
Data processed for direct marketing purposes without possibility for user to object
PNC_11
Data processed for research purposes without possibility for user to object
PNC_12
Unable to transmit stored personal data regarding data subject to third party
PNC_13
System has users with an age of 15 or below
PNC_14
No parental consent for when the user is a child
PNC_15
Inability for user to withdraw consent at process
PNC_16
No consent/purpose at process
PNC_17
No consent/purpose for sharing data with external parties
PNC_23
No consent/purpose for storing data
PNC_24
No consent/purpose for processing data
PNC_25
Insufficient logs
PNC_18
Not logging data breach
PNC_26
Not logging personal data processing
PNC_27
Inability to demonstrate data subject has consented to a processing
PNC_28
Data subject is unable to request access of its stored personal data
PNC_19
No possibility for data subject to access privacy notice
PNC_20
Unable to respond to data breach
PNC_21
Fail to report to supervisory authority
PNC_29
Fail to notify affected data subject
PNC_30
Content Unawareness
User unaware of data has been disseminated to unauthorized parties
No assigned DPO at organization
PNC_22
Content Unawareness
User unaware regarding data inside system
Rule 1:
source is [Data Store]
where
NOT ANY (df, s IN (source.recieveFrom, source) WHERE s.isPII = 'True'
AND df is [DataFlowErasure]
AND df.recieveFrom is [ProcessComponent])
Rule 2:
(source is [External Entity] or
source is [Process]) and
target is [Process]
where
NOT ANY (df, t IN (source.sendTo, target)
WHERE df.isPII = 'True'
AND df is [DataFlow] AND df.pol = 'True')
Rule 3:
source is [DFDNode]
target is [Process]
where
NOT ANY (df, s IN (source.sendTo, source) WHERE df.isPII = 'True'
AND df.sendTo is [ProcessRequest] AND df.sendTo.sendTo is [DataFlow] AND df.sendTo.sendTo.pol = 'True'
AND target.recieveFrom is df.sendTo.sendTo)
Rule 4:
source is [DFDNode]
target is [ProcessComponent]
where
ANY (df IN source.sendTo WHERE df.isPII = 'True'
AND df.pol = 'False')
Rule 5:
source is [Process]
where
NOT ANY (df IN source.sendTo WHERE df.isPII = 'True'
AND df.sendTo is [DataStoreLog])
Rule 6:
source is [Data Store] and
target is [DFDNode]
where
NOT ANY (t IN target WHERE t is [External Entity])
Rule 7:
source is [DataStorePolicy] and
target is [DFDNode]
where
NOT ANY (t, sSend IN (target, source.sendTo) WHERE sSend is [DataFlow]
AND sSend.sendTo is ProcessComponent
AND t is [External Entity])
Rule 8:
(source is [Process]) and
target is [External Entity] and
where
ANY (df, IN source.sendTo,
WHERE df.isPII AND df.crossesTrustBoundary = 'True')
Rule 9:
source is [DataFlowDiagram] and
target is [DFDNode]
where
NOT ANY (n IN source.dfdnode WHERE n is [DataStorePolicy])

Figure A.1: Extended Threat Tree Policy and Consent Non-Compliance

A.2 Content Unawareness

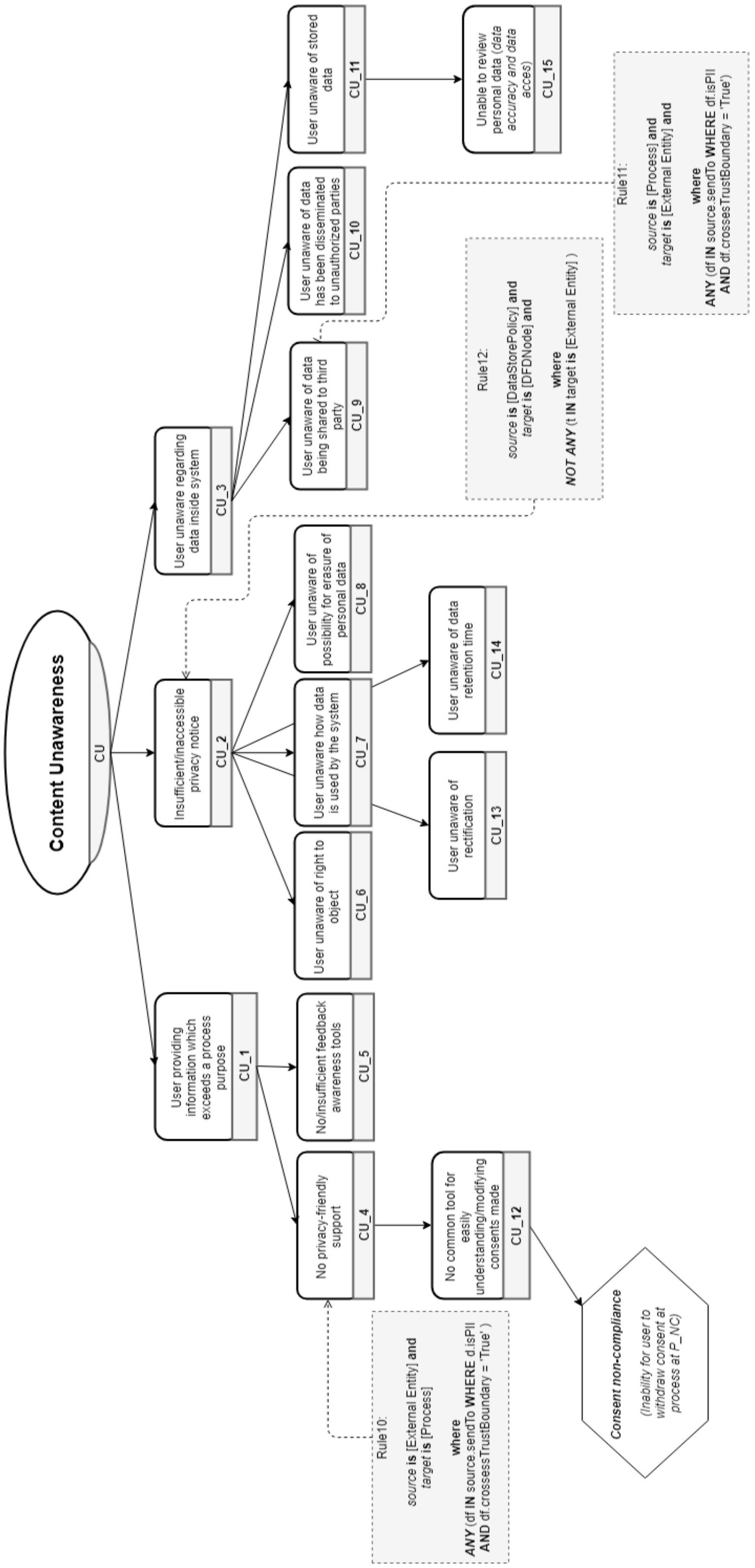

Content Unawareness
CU
User providing information which exceeds a process purpose
CU_1
Insufficient/inaccessible privacy notice
CU_2
User unaware regarding data inside system
CU_3
No privacy-friendly support
CU_4
No/insufficient feedback awareness tools
CU_5
User unaware of right to object
CU_6
User unaware how data is used by the system
CU_7
User unaware of possibility for erasure of personal data
CU_8
User unaware of data being shared to third party
CU_9
User unaware of data has been disseminated to unauthorized parties
CU_10
User unaware of stored data
CU_11
No common tool for easily understanding/modifying consents made
CU_12
User unaware of rectification
CU_13
User unaware of data retention time
CU_14
Unable to review personal data (data accuracy and data acces)
CU_15
Consent non-compliance
(Inability for user to withdraw consent at process at P_NC)
Rule10:
source is [External Entity] and
target is [Process]
where
ANY (df IN source.sendTo WHERE d.isPII
AND df.crossessTrustBoundary = 'True')
Rule11:
source is [Process] and
target is [External Entity] and
where
ANY (df IN source.sendTo WHERE df.isPII
AND df.crossesTrustBoundary = 'True')
Rule12:
source is [DataStorePolicy] and
target is [DFDNode] and
where
NOT ANY (t IN target is [External Entity])

Figure A.2: Extended Threat Tree Content Unawareness