Dipartimento di Informatica, Bioingegneria,
Robotica ed Ingegneria dei Sistemi

Automating the Quantification and Mitigation of Risks for Multiple Stakeholders

by

Majid Mollaeefar

Theses Series **DIBRIS-TH-2022-XX**

DIBRIS, Università di Genova
Via Opera Pia, 13 16145 Genova, Italy http://www.dibris.unige.it/

Università degli Studi di Genova

Dipartimento di Informatica, Bioingegneria,

Robotica ed Ingegneria dei Sistemi

Ph.D. Thesis in Computer Science and Systems Engineering
Computer Science Curriculum

Automating the Quantification and Mitigation of Risks for Multiple Stakeholders

by

Majid Mollaeefar

November, 2022

Dottorato di Ricerca in Informatica ed Ingegneria dei Sistemi
Indirizzo Informatica
Dipartimento di Informatica, Bioingegneria, Robotica ed Ingegneria dei Sistemi
Università degli Studi di Genova

DIBRIS, Univ. di Genova
Via Opera Pia, 13
I-16145 Genova, Italy
http://www.dibris.unige.it/

Ph.D. Thesis in Computer Science and Systems Engineering
Secure and Reliable Systems Curriculum
(S.S.D. ING-INF/05, INF/01)

Submitted by Majid Mollaeefar
DIBRIS, Univ. di Genova and Security & Trust Research Unit, Fondazione Bruno Kessler, Trento, Italy
mmollaeefar@fbk.eu, S4619475@studenti.unige.it

Date of submission: July 2022

Title: Automating the Quantification and Mitigation of Risks for Multiple Stakeholders

Advisor: Professor. Silvio Ranise
Department of Mathematics, University of Trento, Trento, Italy &
Director of the Cybersecurity Center, Fondazione Bruno Kessler, Trento, Italy
ranise@fbk.eu

Ext. Reviewers:
Professor. Joaquin Garcia-Alfaro
Institut Mines-Telecom & Institut Polytechnique de Paris, France &
Adjunct Research Professor, School of Computer Science, Carleton University, Canada
joaquin.garcia_alfaro@telecom-sudparis.eu

Federica Maria Francesca Paci
Associate Professor, Department of Computer Science, University of Verona, Italy
federicamariafrancesca.paci@univr.it

Abstract

Cybersecurity risk management consists of several steps including the selection of appropriate controls to minimize risks. This is a difficult task that requires searching through all possible subsets of a set of available controls and identifying those that minimize the risks of all stakeholders. Since stakeholders may have different perceptions of the risks (especially when considering the impact of threats), conflicting goals may arise that require finding the best possible trade-offs among the various needs such as costs and expertise needed to deploy controls. The ability to tackle this kind of problem is particularly relevant when considering privacy provisions deriving from national or international regulations (such as the General Data Protection Regulation, GDPR) whereby the organization offering a data processing activity should reduce the user's risk to an acceptable level while controlling costs and other business goals. In this context, being able to compute the subsets of controls that minimize the risks of both the organization of the system and its users is a necessary prerequisite to identify the most appropriate configuration of the controls that offer the best possible trade-off among the various objectives. The thesis proposes a quantitative and (semi)-automated approach to solve this problem based on the well-known notion of Pareto optimality. First, we describe a methodology to semi-automatically assist stakeholders in defining their objectives that measures how much risks are reduced by adopting a certain configuration of mitigation controls. Second, we define a decidable multi-objective optimization problem (based on the objectives previously identified)—called Multi-Stakeholder Risk Minimization Problem (MSRMP)—whose Pareto optimal solutions are the subsets of the controls for which no stakeholder's risk can be further reduced without increasing the risk of at least one of the other stakeholders. Third, we validate our approach by showing how a prototype tool based on it can assist in the Data Protection Impact Assessment mandated by the General Data Protection Regulation on different use case scenarios. Lastly, we evaluate the scalability of the approach by conducting an experimental evaluation.

Acknowledgements

During my Ph.D., I had the pleasure of working with and learning from some remarkable people in the Security & Trust Unit at Fondazione Bruno Kessler (FBK), where I acquired many great experiences.

First of all, I would like to express my sincere gratitude to my supervisor, Prof. Silvio Ranise, for all the support, advice, and encouragement he gave me throughout this long journey. Silvio guided me to plan my approach, identify the topic and the problem, and launch my research. I am grateful for all of his time, ideas, immense knowledge, and patience in making my Ph.D.

I wish to thank Prof. Federica Maria Francesca Paci and Prof. Joaquin Garcia-Alfaro for their efforts and time in reading my thesis and their positive and interesting remarks. I would also like to thank Prof. Giorgio Delzanno, coordinator of my Ph.D. program, and the technical committees.

I would like to thank Dr. Roberto Carbone for his kind advice and collaboration on the Trace4Safe project. I also wish to thank Dr. Alberto Siena for his help, advice, and cooperation at the beginning of my Ph.D.

I would like to thank my colleagues (Alex, Andrea, Amir, Biniam, Giada, Marco, Salimeh, Salvatore, Stefano, Tahir, and Umberto) in the Security & Trust Unit and all my friends in Trento.

Finally, I would like to thank my family for their love and support. My heartfelt gratitude goes to my love, Sevda, who has always been there for me over the last three years and supported me wholeheartedly in every challenging situation.

Table of Contents

List of Figures

List of Tables

Abbreviation

API	Application Programming Interface
CIA	It stands for Confidentiality, Availability, Integrity
EDPB	European Data Protection Board
ENISA	European Network and Information Security Agency
DPIA	Data Protection Impact Assessment
DPDD	Data Protection by Design and by Default
GDPR	General Data Protection Regulation
HSP	Health Service Provider
ISO	International Organization for Standardization
MOOP	Multi Objective Optimization Problem
MOCOP	Multi-Objective Combinatorial Optimization Problem
MSRMP	Multi-Stakeholder Risk Minimization Problem
NIST	National Institute of Standards and Technology
PIA	Privacy Impact Assessment
RMP	Risk Management Policy
PII	Personally Identifiable Information
SDM	Standard Data protection Model
SSP	Subset Sum Problem

Chapter 1

Introduction

It is becoming increasingly difficult for enterprises to protect themselves from cybersecurity threats as technology progresses, cyberspace evolves, and digitalization rises. Cybersecurity risks are ubiquitous, regardless of the size or industry of a firm. Therefore, organizations ought to embrace systematic and disciplined cybersecurity risk management to protect critical infrastructure and information systems. Incorporating a cybersecurity risk management strategy within an organization can have several positive consequences. For instance, it can help reduce the cost of security incidents while also minimizing data breaches, compliance difficulties, and attack vectors.

Identifying, evaluating, and prioritizing cyber risks and implementing controls to reduce them are essential components of any organization's risk management strategy. Several approaches are available to identify, evaluate, and prioritize cybersecurity threats, such as the NIST Risk Management Framework[1], ISO Information security risk management framework[2], ENISA Risk Management/Risk Assessment framework[3], consisting of several steps, including the selection of controls necessary to protect the system and organization that are commensurate with risk. These approaches are frequently employed in the literature on privacy, as security risks are analogous to privacy risks. Additionally, there are other risk assessment approaches that are either focused on technical failures (e.g., FMEA [T+05]) or the appearance of security risks (e.g., FAIR [FJ14a], CORAS [LSS10]), where the risk impact depends on the value of business assets or the degree of criticality of technical components or services in these approaches. However, a key distinction between security and privacy risk is that harm on the individuals is a primary consideration for privacy risks (even if organizations may translate that into reputational and regulatory risks), whereas it is of secondary importance for security risks [WB18]. Therefore, various privacy

[1] `https://csrc.nist.gov/projects/risk-management/`

[2] `https://iso.org/standard/75281.html`

[3] `https://enisa.europa.eu/topics/threat-risk-management/risk-management`

issues occur if privacy considerations during system development are not appropriately addressed [CF12].

On the other hand, the new data protection regulation of the EU—the General Data Protection Regulation (GDPR) [reg16]—has been presented to guarantee European citizens' fundamental rights concerning personal data protection and privacy. One of the primary objectives of the GDPR is to give individuals control over their data. Controllers and processors of personal data must provide appropriate technical and organizational measures to minimize the risks that personal data can be abused, for instance, in the case of a data breach. In May 2018, the GDPR took effect, repealing Directive 95/46/EC [Dir95]. According to article 83 of the GDPR, infringements of the GDPR may result in administrative fines of up to a maximum penalty of 20 million euros or 4% of global annual revenue [4]. Consequently, compliance with the GDPR has become a major priority for all organizations in the EU and those processing EU citizens' data [ARP18].

The GDPR requires that data subjects' risks be minimized while also taking other aspects into consideration, such as consent best practices and the budget constraints of the other stakeholders (e.g., the data controller, the data processor, and involved third parties). Additionally, the GDPR requires that a Data Protection Impact Assessment (DPIA) be conducted in order to evaluate the security and privacy measures that have been implemented and minimize the impact of threats on the rights and freedoms of individuals. This means that the organization offering a data processing activity should reduce the user's risk to an acceptable level while controlling costs and other business goals. Therefore, the data controllers must adhere to an approach, method, or framework that will assist them in making more informed decisions about which security mechanisms will provide a better trade-off between their requirements and those of the data subjects.

This task—providing a more favorable trade-off between organizations' needs and those of their users—is non-trivial, as it may require to search through a large set of available controls to mitigate the previously identified set of threats. It is important to note that, in an ideal situation, it is not sufficient to identify a solution(i.e., a subset of the available controls that reduces risk to the desired level); instead, it is desirable to identify those subsets that not only minimize risk but also satisfy other criteria, such as cost reduction or the availability of cybersecurity skills to correctly deploy the selected controls. To further complicate the situation, the stakeholders who are involved in the system or organization may have divergent objectives; for example, a customer of an online banking service may desire to have all threats to their financial transactions eliminated, whereas the bank may be willing to provide protection for the most common vulnerabilities while accepting the risk of more sophisticated attacks in order to maintain costs at an acceptable level.

[4] `https://gdpr-info.eu/art-83-gdpr/`

1.1 Research Context and Questions

Multi-stakeholder collaboration is common to design, develop, and deploy cyber-systems. Risk management is crucial in such modern systems because we must assess the risk from each stakeholder's perspective, which leads us to undertake a multi-stakeholder risk assessment. In this scenario, different stakeholders have various criteria to evaluate the potential impact of threats. Consequently, risk management policies (RMPs)—an RMP can be defined as a set of technical and organizational measures put in place to deal with risk—have different effects on the risk exposure of the stakeholders. An RMP should therefore be selected with the purpose of minimizing the risks for all the considered stakeholders while considering, at the same time, additional constraints, such as legal prescriptions or business requirements. This leads to our first research question:

***RQ* 1**

How can we conduct a risk assessment while considering the preferences of other stakeholders in order to investigate various risk management policies and provide auditability?

Basically, this question pertains to the *research context*—i.e., risk assessment in general and multi-stakeholder risk assessment in particular. The ultimate goal is to present a multi-stakeholder risk assessment technique that facilitates *auditability* by making each choice (i.e., the capability of selecting an RMP) *traceable*. This requires (i) a better understanding of the problem and formalizing it, (ii) developing a methodological approach capable of driving the risk analyst in the quantitative estimation of the risk exposure levels, (iii) evaluating the capability of the approach to explore alternative RMPs, by integrating an automated technique to discover all optimal solutions, and (iv) supporting in the decision-making process and guide risk analyst to select an RMP as the appropriate solution. Taking these considerations into account, we introduce our second research question:

***RQ* 2**

What are the procedures to collect the required parameters for quantifying the risk levels in a multi-stakeholder scenario?

Risk assessment is a critical task, and we need to focus on the design phase of the system development life cycle to determine which kind of RMPs must be put in place to reduce the risk. To achieve this—i.e., determine various RMPs and their risk reduction levels—it must be done in a scientific and principled manner and the best way is to make it as quantitative as possible. Therefore, we need to precisely identify the required *processes* and *parameters* to perform a quantitative multi-stakeholder risk assessment.

As we mentioned earlier, several approaches/methodologies/frameworks in the literature deal with security and privacy risks. Briefly, these procedures can be structured in two stages: (i) *configuration*, where the characteristics of the specific organization are captured and transformed into parameters of the risk evaluation method; and (ii) *execution*, in which the identified threats are estimated together with the mitigation controls, and their evaluated impact levels from each perspective, which in turn defines the risk exposure for the various stakeholders. An answer to **RQ2** can be divided into two parts; the former deals with defining the procedures necessary to gather all relevant data as input artifacts (e.g., threats, stakeholders' preferences, mitigation controls, etc.) for performing the risk assessment, while the latter deals with the required functions such as assigning the associations between input artifacts (e.g., the association between threats and mitigation controls), quantification processes, and risk level evaluations. Thus, to answer this question, we first need to take a quantitative approach and then find the required parameters and procedures to identify all the features that contribute to defining the problem.

Once all the processes needed to quantify risk levels have been specified, the next hurdle is to be in a position to investigate among the various risk management policies that could be adopted. Thus, our third research question is:

RQ 3

How may alternative risk management policies shall be considered to assist the decision-making process?

A possible approach to address **RQ3** is to perform a what-if analysis under various RMP combinations. However, it would be a time-consuming and tedious effort to manually evaluate numerous possible combinations of RMPs in order to find the most appropriate one. Furthermore, in terms of search space to explore all possible combinations, it would be impractical even if the process is done automatically. Indeed, in the presence of competing interests, the search for RMPs that simultaneously minimize the risk level for each stakeholder becomes a non-trivial task and needs the adoption of the concept of Pareto optimality. To understand the problem, consider a situation where, according to one RMP, the risk for stakeholder 1 is 1 and for the stakeholder 2 is 2 but for another RMP, the risks for the two stakeholders are switched; the former is better than the latter with respect to the risk of the first stakeholder, but it is worse with respect to that of the second. An obvious question arises: which solution should be preferred? The answer is to use the notion of Pareto optimality (see, e.g., [MA04]).

From a theoretical perspective, to address **RQ3**, we need to develop a solution to suggest using a multi-objective optimization problem that can be solved by available automated techniques for identifying Pareto optimal solutions whereby, from a stakeholder's perspective, each objective indicates how much the selected RMP minimizes the risk. Then, we can

formalize the problem in mathematical terms and transform it into a multi-objective optimization problem that enables us to explore Pareto optimal solutions in the search space. Note that, here, a Pareto optimal solution means a set of controls that no stakeholder's risk could be reduced more without at least one stakeholder's risk getting worse.

By defining and formalizing the problem and transforming it into a multi-objective optimization problem, the next research question deals with how to solve such a problem in an automated way:

RQ 4

How can we explore and find the optimum solutions among all conceivable risk management policies?

From a pragmatic perspective, to address this question, we need to solve the problem of identifying optimal RMPs, and since the search spaces in these kinds of problems are large, they should be complemented by a suitable heuristic to reduce the search space. A summary of the discussion above is illustrated in Figure 1.1, where we have sketched a conceptual flow of the research questions and the required tasks to answer them.

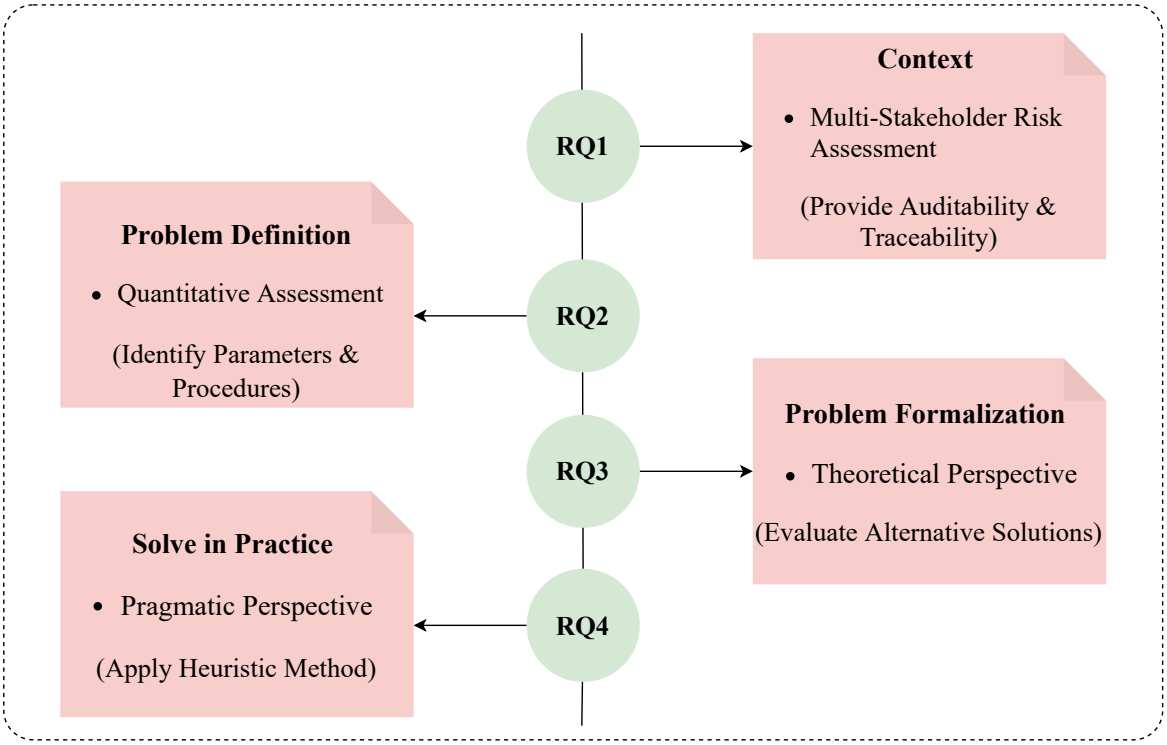

Figure 1.1: Conceptual flow of research questions along with the required tasks for each.

1.2 Contributions

Figure 1.2 depicts a high-level overview of the main contributions made in this thesis. The contributions address the research questions outlined above and are as follows:

- ✦ We introduce the Multi-Stakeholder Risk Minimization Problem (MSRMP), provide a formalization as a multi-objective optimization problem, and present an approach to reduce its search space. For concreteness, we propose a running example to illustrate the main ideas underlying the problem (Chapter 3).
- ✦ We propose an automated technique to solve MSRMP instances to find all optimal solutions and help the various stakeholders to know under which risk management policies the risk exposure is minimized (Chapter 4). The proposed technique is meant to be integrated in privacy impact assessment methodologies.
- ✦ We demonstrate and validate the applicability of the proposed methodology by developing a tool to assist in defining an instance of the MSRMP and then conduct several experiments to evaluate the practicality of the methodology (Chapter 5).
- ✦ We further validate our approach by applying it to a real-world use-case scenario in the framework of the General Data Protection Regulation (Chapter 6).

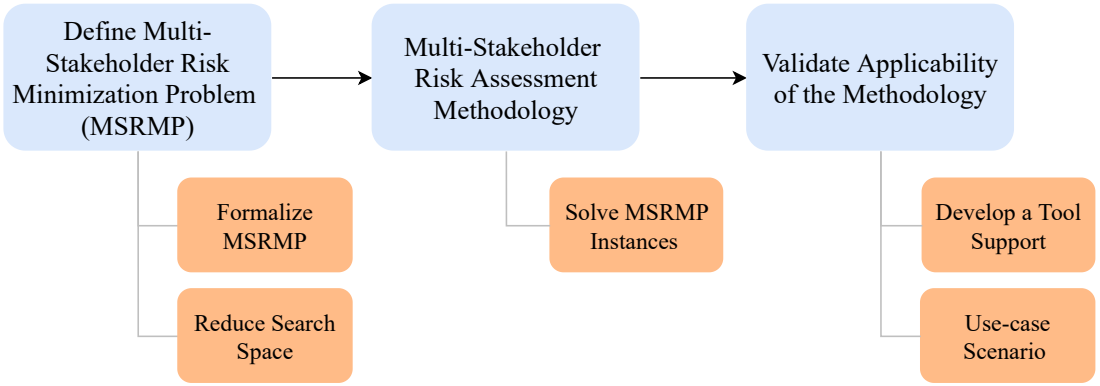

Figure 1.2: Overview of the main contributions of this thesis.

1.3 Thesis Outline

The outline of the thesis is as follows:

- In Chapter 2, we provide some background information.
- In Chapter 3, we introduce the Multi-Stakeholder Risk Minimization Problem (MSR MP), provide a formalization as a multi-objective optimization problem, and present an approach to reduce the search space. For concreteness, we propose a running example to illustrate the main ideas underlying the problem.
- To find all Pareto optimal solutions and help the various stakeholders to know under which risk management policies the risk exposure is minimized, we propose an automated technique to solve MSRMP instances in Chapter 4.
- To demonstrate the applicability of the proposed methodology, we developed a tool to assist in defining an instance of the MSRMP and then conducted several tests to experimentally evaluate the practicality of the methodology in Chapter 5.
- In Chapter 6, we apply our approach to a real-world use-case scenario. This chapter includes our contributions to a European research project, called `Trace4Safe`.
- In Chapter 7, we discuss related work.
- Finally, in Chapter 8, we conclude this dissertation with a summary of the main contributions and identify some directions for future work.

1.4 Publications

This thesis is based on the following three research papers and one research project written:

`Journal`:

1. **Majid Mollaeefar**, and Silvio Ranise. Identifying and Quantifying Trade-offs in Multi-Stakeholder Risk Evaluation with Applications to the Data Protection Impact Assessment of the GDPR. (*Under-Revision - "Journal of Computers and Security", September 2022*)

`Conference`:

1. **Majid Mollaeefar**, Alberto Siena, and Silvio Ranise. Multi-stakeholder cybersecurity risk assessment for data protection. In *Proceedings of the 17th International Joint Conference on e-Business and Telecommunications - Volume 3: SECRYPT*, pages 349–356. INSTICC, SciTePress, 2020.
2. **Majid Mollaeefar**, Marco Pernpruner, and Silvio Ranise. Identifying and Quantifying Risk Trade-offs in Enrollment Procedures. (*To be submitted, 2022*)

`Trace4Safe Project`:

1. **Majid Mollaeefar**, Silvio Ranise, and Roberto Carbone. Data Protection Impact Assessment within Trace4Safe "a Hybrid Contact Tracing Monitoring and Detection for a Safe Workplace". (*EIT Contact Tracing Project, 2021*)

Chapter 2

Background

This chapter provides some background notions required for this thesis to define the context, problem, and solution. In Section 2.1, we briefly explain the role of cybersecurity in businesses and outline some crucial concepts connected to the cybersecurity risk assessment and highlight the processes that need to be taken into consideration to conduct a risk assessment. Section 2.2 is introduced the GDPR along with some fundamental concepts in this context (see Section 2.2.1). The Data Protection Impact Assessment (DPIA) as a requirement under the GDPR introduces in Section 2.2.2. Then, we list the data protection goals (also known as privacy goals) along with a short description of each (see Section 2.2.3). Section 2.3 dedicates to the required fundamental knowledge in the field of multi-objective optimization. Lastly, we introduce a sub-class of multi-objective optimization called multiple objective combinatorial optimization in Section 2.3.3.

2.1 Cybersecurity Risk Assessment

Cybersecurity as a discipline originated in 1972 with a research study on ARPANET (The Advanced Research Projects Agency Network), the internet's forerunner[1]. There are many definitions for the "*cybersecurity*" term, for example, by searching the term in the NIST glossary[2] in [HN+15] it is defined as "*the prevention of damage to, unauthorized use of, exploitation of, and—if needed—the restoration of electronic information and communications systems, and the information they contain, in order to strengthen the confidentiality, integrity and availability of these systems.*"

With an increasing number of people, devices, and programs in today's organizations,

[1] `https://blog.avast.com/history-of-cybersecurity-avast`
[2] `https://csrc.nist.gov/glossary`

coupled with an increasing deluge of data, the majority of which is sensitive or secret, the demand for cybersecurity continues to expand. The rising volume and competence of cyber attackers, as well as their attack techniques, exacerbate the situation. Despite the rising importance of cybersecurity, many organizations continue to address the problem as a technological issue, just as they did in the mid-1990s. The goal of identifying and managing cyber-threats is universally shared by all businesses, and cyber risk management is the critical practice that may be used to achieve this goal. The following are some of the advantages of developing and maintaining cybersecurity practices:

- Protect businesses against cyber-attacks and data breaches.
- Protect data and network security.
- Prevent unauthorized accesses.
- Reduce time required to recover from a breach.
- Protect end user and endpoint device.
- Compliance with applicable regulations.

2.1.1 Cybersecurity Risk Assessment Terms

A glossary that represents all terms and definitions that are used in this thesis is provided in Appendix A. For the purposes of this thesis, here, we refer to the NIST publications such as [R$^+$11, A-116, SP812, SBP$^+$10, RMO16] to elucidate the terms that are frequently used in this thesis. In the following, we recall some of them that are related to the risk assessment procedure.

The term *risk* is a measure of the extent to which an entity is threatened by a potential circumstance or event, and typically is a function of: (i) the adverse impact, or magnitude of harm, that would arise if the circumstance or event occurs; and (ii) the likelihood of occurrence [A-116].

The term *impact* is the magnitude of harm that can be expected to result from the consequences of unauthorized disclosure of information, unauthorized modification of information, unauthorized destruction of information, or loss of information or information system availability [SBP$^+$10].

Likelihood or *likelihood of occurrence* is a weighted factor based on a subjective analysis of the probability that a given threat is capable of exploiting a given vulnerability or a set of vulnerabilities [SP812].

The term *threat* represents any circumstance or event with the potential to adversely impact organizational operations, organizational assets, individuals, other organizations, or the Nation through a system via unauthorized access, destruction, disclosure, modification of information, and/or denial of service [SP812].

The term *risk assessment* defined as the process of identifying risks to organizational operations (including mission, functions, image, reputation), organizational assets, individuals, other organizations, and the Nation, resulting from the operation of a system [R+11].

The term *risk evaluation* is the process of comparing the results of risk analysis with risk criteria to determine whether the risk and/or its magnitude is acceptable or tolerable [RMO16].

2.1.2 Risk Management Process under ISO-31000

ISO-31000 is a series of risk management standards developed by the International Organization for Standardization. It establishes principles and generic rules for organizations to follow when managing risks. For practitioners and enterprises using risk management procedures, ISO-31000 attempts to establish a universally recognized paradigm that replaces the variety of existing industry-specific standards, methodologies, and paradigms. Therefore, ISO-31000 recommendations can be tailored to any organization and its context. Figure 2.1 shows the risk management process presented in ISO-31000 [Pur10] which contains several activities. In the following, we briefly review all:

Establishing the context. This activity entails defining the risk management process's scope, as well as the organization's goals and risk evaluation criteria.

Monitoring and review. This work involves assessing risk management performance against predetermined metrics. Reporting on risk, progress with the risk management strategy, how effectively the risk management policy is being implemented, and analyzing the efficacy of the risk management framework.

Communication and consultation. This task assists in the understanding of stakeholders' interests and concerns, in ensuring that the risk management process is focusing on the appropriate elements, and in explaining the reasoning for decisions and specific risk treatment solutions.

Risk treatment. It is the process of enhancing existing controls or developing and implementing new controls.

Risk assessment. This process is concerned with identifying risks unique to the environment and estimating the severity of those risks.

We will learn about this process and its sub-processes in the following sections, which will

serve as a foundation for this thesis.

2.1.3 Risk Assessment Process

The evaluation of cybersecurity risks (also known as the risk assessment process) is an essential component of an organization's enterprise risk management strategy. Organizations would be able to do the following by undertaking a risk assessment [Sho14]:

1. Identify "what could go wrong" incidents, which are frequently the result of malicious behaviors by threat actors and have the potential to have unintended business effects in the future.
2. Determine the extent to which they are exposed to various degrees of cybersecurity

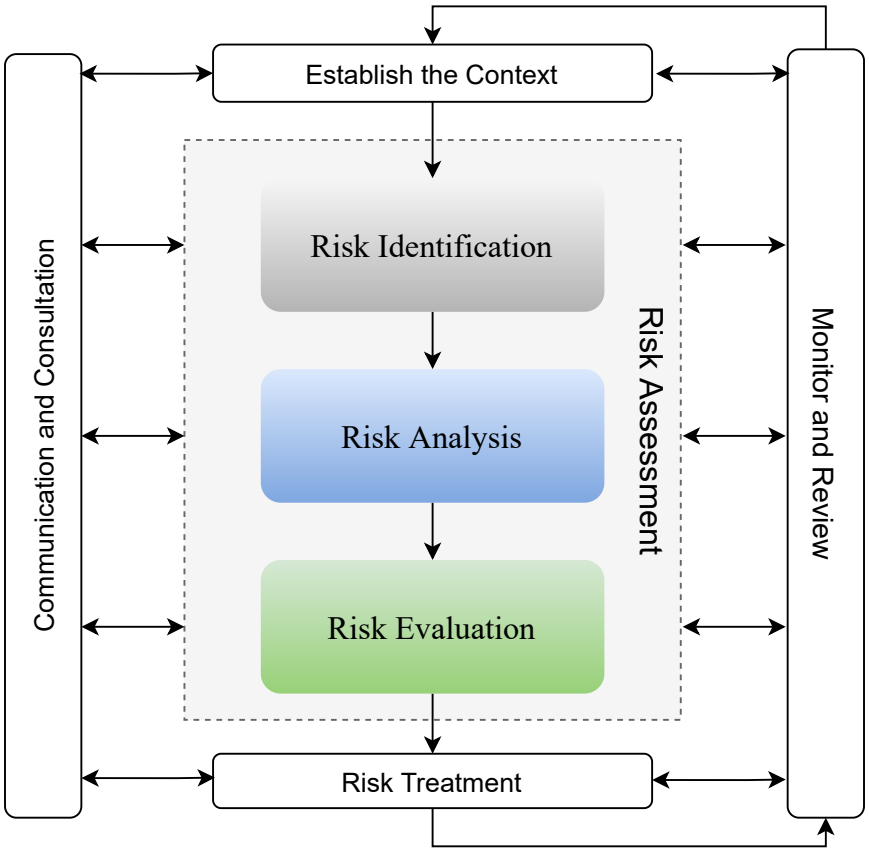

Figure 2.1: The risk management process from ISO-31000 [Pur10].

risk. A thorough awareness of risk levels enables an organization to prioritize and allocate sufficient action and resources to the most serious concerns.

3. Ensure that the organization has a risk-aware culture. Risk assessment is an iterative approach that entails including people in thinking about technological risks and how they relate to organizational objectives in order to complete the evaluation properly.

As shown and highlighted in Figure 2.1, the risk assessment process under ISO-31000 is the heart of a risk management process and entails three primary steps: risk identification, risk analysis, and risk evaluation.

Risk identification is the process of identifying and categorizing sources of risk in order to determine what has to be handled throughout a building project. Generally, in order to identify risks, there are various fundamental approaches to use, including documentation review, information collecting, checklists and risk catalogs, assumption analysis, and diagramming techniques. Modeling threats is one strategy that can be used in the process of identifying cybersecurity risks. Threat modeling has a vital role in building a secure software system by considering how an adversary might compromise the system. Indeed, it supports understanding security and privacy threats in a given system, how those threats impact data subjects and the organization as the data controller and identifying possible security and privacy countermeasures to mitigate potential attacks. It is highly recommended to apply threat modeling early in the development cycle, where potential issues can be detected and remedied early to prevent possible later consequences. There are several approaches that may be used while undertaking threat modeling, such as Microsoft STRIDE [Sho14], PASTA [UM15], LINDDUN [WJ15], and Trike [SLE05]. Each threat modeling technique follows a somewhat different set of phases. Not all methods are comprehensive, or in other words, a particular threat modeling method is not suggested over the others. The decision to select a method depends on the defined project's needs and specific concerns [SCO$^+$18]. In fact, the best model for your requirements is determined by the kind of threats you are attempting to model and the purpose for which you are doing so.

Risk analysis is focused with establishing an understanding of each risk, its effects, and the likelihood of those consequences [Pur10] and the ultimate result can be expressed as a qualitative, semi-quantitative, or quantitative form. *Qualitative* assessments use non-numerical categories or levels to measure risk (e.g., very low, low, moderate, high, very high). This type of assessment supports communicating risk outcomes to decision-makers. Furthermore, unless each value is clearly defined or accompanied by compelling examples, experts who use their own experiences could come up with very different assessments. Unlike qualitative assessments, *quantitative* assessments often involve a set of procedures, principles, or guidelines for assessing risk based on the use of numbers where the meanings and proportionality of values are preserved inside and outside the context of the assessment.

This form of analysis best supports cost-benefit assessments of alternative risk responses. In some cases, interpreting and explaining quantitative results is necessary, especially to clarify assumptions and limitations. Lastly, *semi-quantitative* is a type of assessment that can provide the benefits of quantitative and qualitative assessments [SP812].

Risk evaluation comprises making a decision about the level of risks and then prioritize them based on their significance levels.

2.2 General Data Protection Regulation

General Data Protection Regulation (GDPR) intends to tighten and unify data protection regarding the processing of personal data. It demands consistent and high-level protection of the natural person, eliminates impediments to flowing personal data across the European Union and handles the export of personal data beyond the EU to safeguard personal data of persons resident in the Union regardless of location.

2.2.1 GDPR Concepts and Principles

Figure 2.2 illustrates—albeit simple—the key roles and their relationship and connections from the GDPR perspective. In the following, we provide some definitions related to key

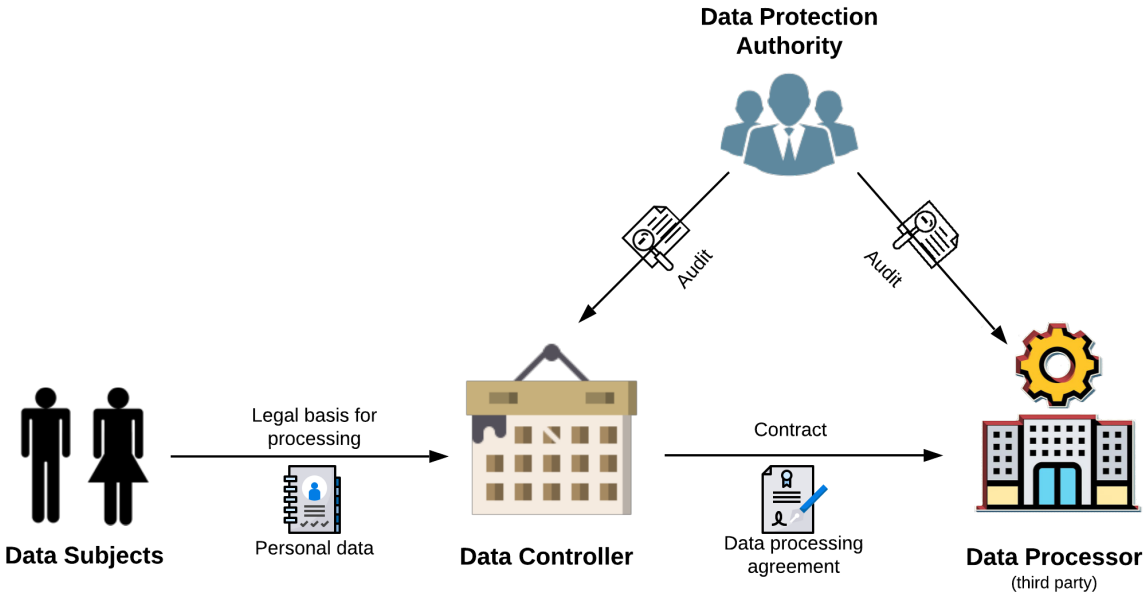

Figure 2.2: Key roles and their relationships under the GDPR.

roles and their responsibilities under the GDPR:

- "*Data subject*" refers to an identified or identifiable natural person whose personal data is being collected, stored and/or processed;
- "*Personal data*" denotes any information that directly or indirectly allows identifying the data subject, in particular by reference to information to an identifier such as name, identification number, location data, or to one or more factors specific to the physical, physiological, genetic, mental economics, culture or social identity of that natural person;
- "*Data controller*" means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law;
- "*Data processing*" means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;
- "*Data processor*" means a natural or legal person, public authority, agency or other body which processes personal data–by following documented instruction by data controller–on behalf of the controller.; unless required to do so by Union or Member State law to which processor subjects to;
- "*Data processing agreement*", is an agreement between a data controller (such as a company) and a data processor (such as a third-party service provider). It regulates any personal data processing conducted for business purposes.
- "*Data protection authority*", or in short DPA, is an independent public authority that supervises, through investigative and corrective powers, the application of the data protection law. In each EU Member State, there is one DPA. They provide expert advice on data protection issues and handle complaints lodged against violations of the GDPR and the relevant national laws.

The GDPR allows data processing only if the data controller and data processor are able to comply with the regulation. They shall: (i) take appropriate measures to protect data against unlawful processing, (ii) provide relative information to data subjects and supervisor authority, (iii) to exercise data subject right upon request, without undue delay.

The GDPR concentrates on identifying and analyzing threats to data subjects' rights and freedoms, and correspondingly to adopt actions to avoid or mitigate their effect on data subjects. For instance, Recital 83 states that "*to maintain security and to prevent processing in infringement of this Regulation, the controller or processor should evaluate the risks inherent in the processing and implement measures to mitigate those risks, such as encryption*"; and Article 24 states that "*taking into account the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons, the controller shall implement appropriate technical and organizational measures to ensure and to be able to demonstrate that processing is performed in accordance with this Regulation* [...]".

Article 5 of the GDPR sets out the fundamental principles underlying the data protection framework. The following is a concise outline of these principles of data protection mentioned in Article 5 GDPR:

(i) **Lawfulness, fairness and transparency.** Any processing of personal data should be processed lawfully, fairly and in a transparent manner in relation to the data subject.

(ii) **Purpose limitation.** Personal data should only be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes.

(iii) **Data minimization.** Processing of personal data must be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.

(iv) **Accuracy.** Controllers must ensure that personal data are accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay.

(v) **Storage limitation.** Personal data should only be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organizational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject.

(vi) **Integrity and confidentiality.** Personal data should be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures.

(vii) **Accountability.** The controller is responsible for, and must be able to demonstrate, their compliance with all of the above-named principles of data protection.

2.2.2 Data Protection Impact Assessment

A data protection impact assessment (DPIA) is a privacy impact assessment aimed at identifying and analyzing how specific data processing activities influence the privacy of data subjects. Under the GDPR, a DPIA will be mandatory for any new high-risk processing projects or when profiling activities are carried out using personal data. A DPIA can enable organizations to identify, mitigate, and prepare for the execution of any risk-related remedies against data protection risks and evaluate the sustainability of a project early. To conduct a DPIA, the GDPR calls for having the following components as described in Article 35:

(a) *A systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller.*

(b) *An assessment of the necessity and proportionality of the processing operations in relation to the purposes.*

(c) *An assessment of the risks to the rights and freedoms of data subjects.*

(d) *The measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned.*

2.2.3 Data Protection Goals

Protection goals play an essential role in analyzing information security in terms of concepts or implementations of data processing systems and guiding the selection of the proper technological or organizational protections for each application [ZH11]. As IT security protection goals, the classical triad of Confidentiality, Integrity, and Availability—first employed in the early 1980s—would classify the system's security capabilities and controls

to achieve a secure outcome [BBG+17]. In recent years, even without any legislative obligations, organizations have taken these three protection goals more into account of their own interests. Initially, they developed only for IT security and outline criteria for secure operation, namely the operation of organizational procedures in relation to their business operations. Apart from these IT security protection goals, the current data protection legislation (e.g., GDPR) has provided other data protection-specific goals. From a data protection perspective, organizations must also protect their business processes against potential threats when such business processes influence personal data. In this regard, in comparison to the IT security goals, the protection goals of data protection require a wider understanding.

The Standard Data protection Model (SDM) [fD17] uses the term "data protection goals" to describe certain categories of requirements derived from data protection law. These requirements are aimed at properties of lawful processing operations, which have to be ensured by technical and organizational measures. The SDM specifies CIA triad, *Unlinkability* & *Data minimization*, *Transparency*, and *Intervenability* as six protection goals of data protection. The latter three goals aim at the specific protection requirements of data subjects and reflect the data protection requirements in an operational form.

To systematize data protection requirements of the GDPR, the SDM employs "protection goals". The data protection requirements seek to ensure legal compliance processing, which technological and organizational safeguards must ensure. The assurance consists in lowering the risk of deviations from legally compliant processes to a suitable degree. Unauthorized processing by third parties and the failure to carry out mandatory processing procedures are examples of deviations to avoid. The data protection goals combine and arrange the criteria for data protection requirements and can be operationalized through integrated, scalable measures [fD17]. These protection goals are as follows:

G1. ***Confidentiality*** refers to the requirement that no person is allowed to access personal data without authorization.

G2. ***Integrity*** refers, on the one hand, to the requirement that information technology processes and systems continuously comply with the specifications that have been determined for the execution of their intended functions. On the other hand, integrity means that the data to be processed remain intact, complete, and up-to-date.

G3. ***Availability*** is the requirement that personal data must be available and can be used properly in the intended process. Thus, the data must be accessible to authorized parties and the methods intended for their processing must be applied.

G4. ***Unlinkability*** & ***Data minimization*** where the *Unlinkability* goal refers to the requirement that data shall be processed and analyzed only for the purpose for which

they were collected, while the *data minimization* goal covers the fundamental requirement under data protection law to limit the processing of personal data to what is appropriate, substantial and necessary for the purpose.

G5. ***Transparency*** refers to the requirement that the data subject as well as the system operators and the competent supervisory authorities can identify to a varying extent, which data are collected and processed for a particular purpose, and which systems and processes are used for this purpose, where the data flow to which purpose, and who is legally responsible for the data and systems in the various phases of data processing.

G6. ***Intervenability*** refers to the requirement that the data subjects are effectively granted the right to notification, information, rectification, blocking and erasure at any time.

The SDM has provided precise mappings between the GDPR requirements and these protection goals (for more details on the SDM and the mappings between protection goals and the GDPR requirements, we dedicated a section in Appendix B, and the mappings are reported in Table B.1). These mappings can be interpreted as if threats adversely affecting these protection goals mean non-compliance with the GDPR requirements. Working with protection goals simplifies the modeling of functional requirements in use cases and the visualization of conflicts. They also enable the methodical application of legal requirements into technological and organizational measures and are therefore "optimization requirements".

2.3 Multi Objective Optimization

Multi-objective optimization (also referred to as multi-objective programming, vector optimization, multi-criteria optimization, multi-attribute optimization, or Pareto optimization) is an area of multiple-criteria decision-making concerned with mathematical optimization issues requiring the simultaneous optimization of many objective functions. It has been applied in many domains of research (e.g., engineering, economics, etc.), where optimal decisions need to be taken in the context of trade-offs between two or more conflicting objectives.

2.3.1 Multi-Objective Optimization Problem

A multi-objective optimization problem (MOOP) deals with more than one objective function, and these objectives are to be minimized or maximized [SD94]. In most cases, one

solution would not satisfy all objective functions, and the optimal solution of one objective will not certainly be the most desirable solution for other objectives. Therefore, different solutions will produce trade-offs between different objectives, and a set of solutions is required to represent the optimal solutions of all objectives. A MOOP can be stated as follows:

$$min(max)_{\overline{x}} \ \langle f_1(\overline{x}), ..., f_n(\overline{x})\rangle \qquad \text{subject to } \overline{x} \in \mathcal{D} \tag{2.1}$$

where $\overline{x}$ is the vector of design variables, f_i is an objective function for $i = 1, ..., n$, and $\mathcal{D}$ is the feasible design space, i.e. the set of all possible values among which to search for the optimal solutions.

2.3.2 Multi-Objective Optimization Definitions

To fully perceive MOOP and the algorithms that solve these kinds of problems, some concepts and definitions must be clarified. In the following, some most important of these concepts are mentioned:

- **Decision variables and objective space**

  The variable bounds of an optimization problem restrict each decision variable to a lower and upper limit, which defines a space called decision variable space. In multi-objective optimization, values of objective functions create a multi-dimensional space called objective space. Each decision variable in variable space corresponds to a point in objective space.

- **Dominance relation**

  Optimizing a solution with respect to one objective will not result in an optimal solution regarding the other objectives. Thus, a vector $\langle f_1(\overline{x}^*), ..., f_n(\overline{x}^*)\rangle$ of objective functions is non-dominated iff there does not exist another vector $\langle f_1(\overline{x}), ..., f_n(\overline{x})\rangle$ such that $f_i(\overline{x}) \leq f_i(\overline{x}^*)$ for each $i = 1, ..., n$ with at least one $f_j(\overline{x}) < f_j(\overline{x}^*)$; otherwise, $\langle f_1(\overline{x}^*), ..., f_n(\overline{x}^*)\rangle$ is dominated. For example, in Figure 2.3, points A, B, and D are non-dominated, whereas points B or D dominate point C and it is not a Pareto optimal solution.

- **Pareto-optimal set (non-dominated set)**

  A solution is Pareto-optimal if it is not dominated by any other solution in the variable space. The Pareto-optimal is the best known (optimal) solution with respect to all

objectives, and cannot be improved in any objective without worsening in another objective. The set of all feasible solutions that are non-dominated by any other solution is called the Pareto-optimal or non-dominated set [MA04]. Therefore, with respect to Formula 2.1, a point $\overline{x}^* \in \mathcal{D}$ is Pareto optimal iff there does not exist another point $\overline{x} \in \mathcal{D}$ such that $f_i(\overline{x}) \leq f_i(\overline{x}^*)$ for each $i = 1, ..., n$ and $f_j(\overline{x}) < f_j(\overline{x}^*)$ for at least one $j \in \{1, ..., n\}$. In other words, a point is Pareto optimal if there is no other point that improves at least one objective function without detriment to another function. It is important to note that the feasible objective space contains not only Pareto-optimal solutions, but also solutions that are not optimal. In fact, feasible objective space can divide into two set solutions, a Pareto-optimal and a non-Pareto-optimal set [SD94]. For example, in Figure 2.3, the feasible objective space comprises all the points.

- **Pareto-front**

  The values of objective functions related to each solution of a Pareto-optimal set in objective space is called Pareto-front. For example, in Figure 2.3, Pareto-front values are specified in the dashed-line and each point in this line consider as a Pareto solution.

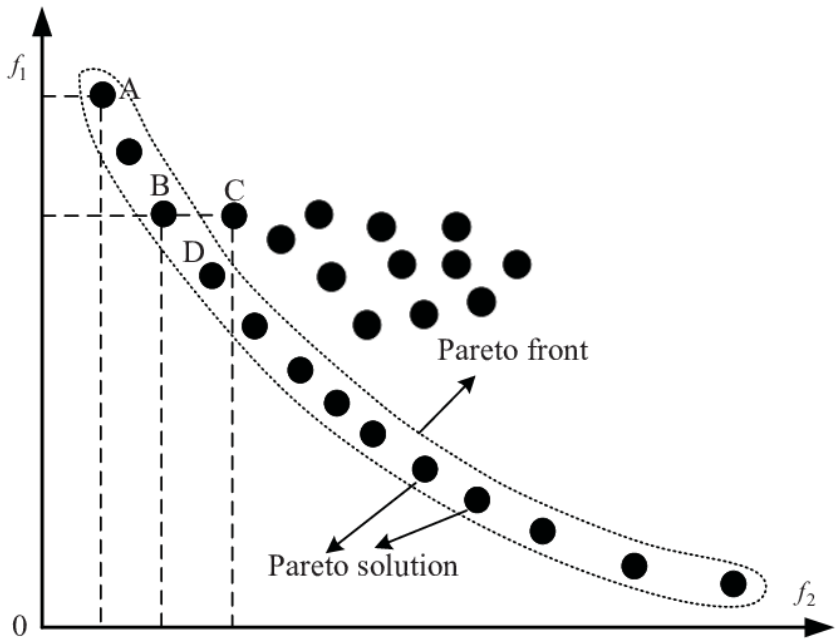

Figure 2.3: Example of Pareto-optimal solutions and dominated solution in a two-objective search space.

2.3.3 Multiple Objective Combinatorial Optimization

Multiple Objective Combinatorial Optimization (MOCO) is a topic that consists of finding an optimal object from a finite set of objects [Sch03]. In many such problems, exhaustive search is not tractable. It operates on the domain of those optimization problems in which the set of feasible solutions is discrete or can be reduced to discrete, and in which the goal is to find the best solution. Typical problems are the travelling salesman problem (TSP), the minimum spanning tree problem (MST), and the knapsack problem. Formally, a general MOCO problem can be stated as $min\ \{f(x) = \langle f_1(x), ..., f_p(x)\rangle \quad T : x \in X\}$, where the decision space X is a given discrete feasible set that usually has some additional combinatorial structure. Due to the fact that the set of feasible solutions to a MOCO problem is discrete and typically finite, it can theoretically be enumerated in order to find all Pareto optimal solutions. However, due to the exponentially growing number of feasible (and sometimes also Pareto optimal) solutions, this is often impractical [Kla09].

Chapter 3

Cyber-Risk Trade-offs in Multi-Stakeholder Scenarios

Cybersecurity risk management entails a series of procedures, the most important of which is the selection of suitable controls to minimize risk. This is a challenging task since it necessitates searching through all potential subsets of a set of available controls in order to select those that minimize the risks to all stakeholders. Due to the fact that stakeholders may have divergent perspectives on risks (particularly when assessing the effect of threats), stakeholders may have opposing objectives. This problem becomes more apparent when organizations realize the fact of compulsory compliance with regulations such as the GDPR. Thus, they must provide a more favorable trade-off between organizations' needs and those of their users.

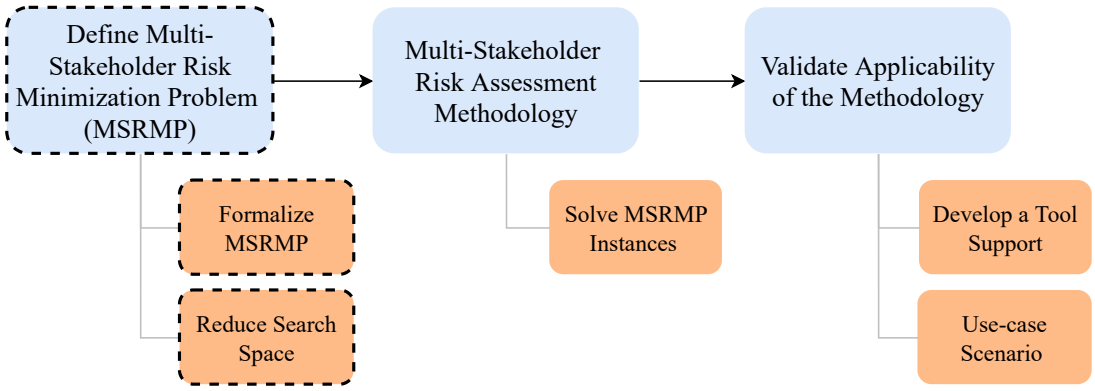

Figure 3.1: The highlighted part (dashed lines) illustrates the contribution of Chapter 3 in accordance with the contribution flow outlined in Section 1.2.

Figure 3.1 shows the contributions of this chapter, where we first provide a brief introduction regarding the problem we aim to address (Section 3.1), afterwards in Section 3.2, we introduce Multi-Stakeholder Risk Minimization Problem (MSRMP), and to better grasp the problem, we present a simplified but realistic running example in Section 3.2.1. Finally, in Section 3.2.2, we formalize the problem in the framework of multi-objective optimization, where we bring a series of interconnected examples to streamline the formalization and present an approach to reduce the search space.

3.1 Introduction

Cyber-risk is a measure of the likelihood and the impact of threats, i.e. circumstances or events with the potential to harm a cyber-system such as the unauthorized disclosure, destruction, modification, or interruption of system assets. Cyber-risk management is the *identification* and *assessment* of risks, followed by the *definition* and *enforcement* of appropriate *mitigation measures* for risk minimization. The identification of risks depends on the assets of the system to be protected and requires performing threat modeling, i.e. to understand and describe how an adversary might compromise a system. The assessment of risks amounts to evaluating the impact and the likelihood of the various threats. For instance, a backdoor in a certain version of an operating system may have a dramatic impact. The risk may be severe if patches are applied late as the likelihood that an adversary exploits the vulnerability is high, whereas the risk becomes small when patches are quickly applied as the time-window during which an attacker can exploit the vulnerability is substantially reduced. The balance between impact and likelihood is key to risk assessment. Once risks have been identified and assessed, suitable Risk Management Policies (RMPs) should be defined and enforced. RMPs comprise both technical (e.g., deploy the latest version of the Transport Layer Security protocol) and organizational (e.g., a cyber security awareness and training program for employees) measures to minimize risks. Indeed, the ultimate goal of risk management is to minimize risks while maximizing the chances to reach business objectives and complying with legal provisions, such as the GDPR. Indeed, failing to do this may bring in additional risks and costs due to an unsatisfactory return on investment or fines for lack of compliance.

As we stated earlier (Section 1.1), we seek an approach/technique that at the end of the process (i.e., performing a multi-stakeholder risk assessment) provides this facility for the data controller to audit and trace all possible solutions (i.e., various RMPs). This recalls **RQ1.** *How can we conduct a risk assessment while considering the preferences of other stakeholders in order to investigate various risk management policies and provide auditability?* In the following section, to address this question, we define the Multi-Stakeholder Risk Minimization Problem (MSRMP), and additionally, we provide a formalization of the problem as a multi-objective optimization problem.

3.2 Multi-Stakeholder Risk Minimization Problem

Given the increasing complexity of cyber-systems, it is routine that several stakeholders cooperate in their design, development, and deployment. This further complicates risk management. For instance, according to the GDPR, in case a system processes personal data, its data controller shall guarantee that the risk of violating the rights and freedom of the data subjects is low. The data controller must do this by considering state-of-the-art RMPs and budget constraints. When the data controller involves a data processor, the latter may have strict computational constraints for scalability and efficiency that, in turn, guarantee economy of scale. While the various stakeholders may agree on a common set of threats for a given system together with their likelihood, they will have diverging criteria to evaluate the potential impact of the identified threats. For instance, data subjects will favor comprehensive RMPs to reduce the risk of data breaches, while a data controller or a data processor may be more interested in cheap and easy to enforce RMPs that cover most threats while neglecting those that are less likely to occur. Besides making the definition of the impact of threats dependent on each stakeholder, this greatly complicates the search for RMPs that minimize risks. Indeed, the search for RMPs that simultaneously minimize the risk level for each stakeholder becomes a non-trivial task in the presence of conflicting objectives and requires the adoption of the notion of Pareto optimality. To understand the problem, consider the situation in which we have two RMPs *rpm1* and *rpm2* with risk vectors $\langle 1, 2, 1\rangle$ and $\langle 1, 1, 2\rangle$, respectively, where the first component is the risk of the data subject, the second is that of the data controller, and the third is that of the data processor. The data subject has no preference between the two RMPs, the data controller prefers *rpm1* over *rpm2*, and the data processor *rpm2* over *rpm1*. In other words, no RMP minimizes the risk for all the stakeholders; so, which one between *rpm1* over *rpm2* should be preferred? According to the notion of Pareto optimality (as we introduced in Section 2.3), both *rpm1* and *rpm2* are to be considered optimal and further aspects need to be considered to select one of the two, such as the fact that one of the two promises to provide a higher return on investment or that it is easier to show its compliance with the GDPR or other legal provisions. Because vectors cannot be ordered completely, all the Pareto optimal solutions can be regarded as equally desirable in the mathematical sense, and we need a decision maker to select the preferred one among them. To enable the decision maker to do this, we need to be able to compute the set of Pareto optimal solutions.

The explanation above reminds us **RQ3.** *How may alternative risk management policies shall be considered to assist the decision-making process?*, and to address this question below in Section 3.2.2, we formalize the problem of finding Pareto optimal configurations of RMPs— i.e., configurations that minimize stakeholders' risks—in the framework of multi-objective optimization and show how it can be solved by using general purpose algorithms under reasonable assumptions. Preliminary, we introduce a simplified but realistic running

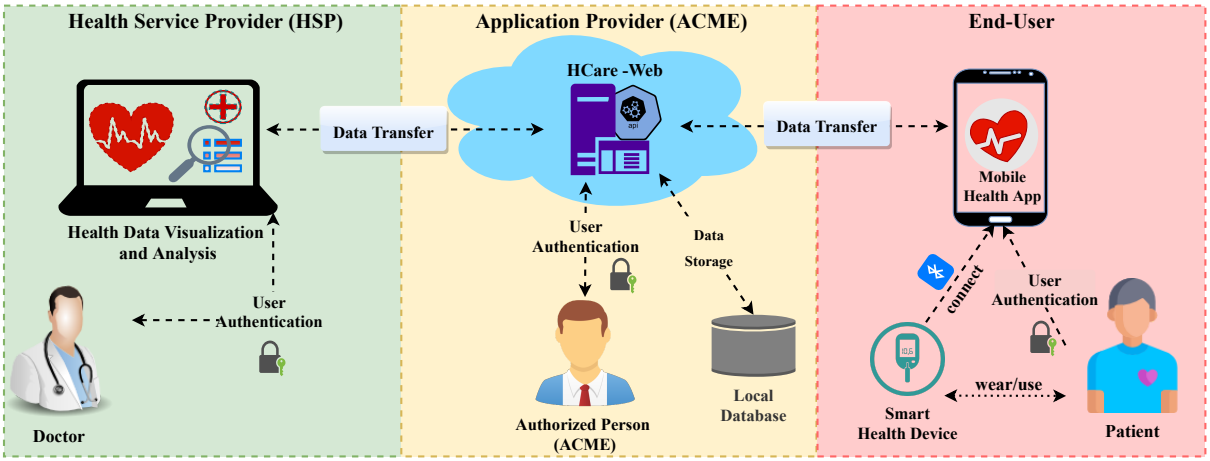

Figure 3.2: Overview on the main stakeholders in the scenario and their interaction with the system's components.

example to better grasp the problem.

3.2.1 Running Example: An Application of the GDPR's DPIA

We consider the situation in which an Italian company, called ACME below for the sake of anonymity, that must perform a Data Protection Impact Assessment (DPIA, see Section 2.2.2) for one of its software applications, as required by Article 35 of the GDPR.[1] The goal of a DPIA is to protect the rights and freedom of EU citizens with particular relevance to those related to their privacy. For this, it is crucial to perform an appropriate privacy risk assessments. From the GDPR perspective, there are three main stakeholders involved in the process, namely (i) the *Data Subject*, the patient whose data are being collected, stored and processed by the application, (ii) the *Data controller*, ACME which is responsible for offering the data processing activities implemented by the software application, and (iii) the *Data processor*, a company mandated by the Data Controller to design and implement the application deploying the various data processing activities. The data processor is a third party organization, possibly external to the data controller. In the rest of this section, we focus on the problem of identifying appropriate security controls among a set of available ones that minimize the risks of all three stakeholders. A peculiarity of this risk assessment is that the data controller must perform it to make the risk of the data subjects acceptable. Indeed, this may give rise to conflicts with the data controller's and data processor's requirements on budgets and skill's shortage. In Table 3.1, we later provide an assessment under the GDPR—as an indication—to highlight some potential

[1] https://gdpr-info.eu/art-35-gdpr/

conflicts in the ACME scenario.

ACME develops a software application, called HCare, exposing an API service to allow its clients to work together, as illustrated in Figure 3.2. Through the API, HCare connects three main stakeholders: the Health Service Provider (HSP), the API provider (ACME), and the patients which are the data controller, the data processor, and the data subjects in the context of the GDPR, respectively. Notice that an HSP in our case can also be an independent developer who provides IT-only services without offering actual health care support; for example, providing data visualization tools. Finally, the end-user is typically the patient using the app to send biometric data or user-initiated requests and receive responses from the HSP, e.g., prescriptions from a doctor, medical alerts, etc. HSPs use the APIs to perform some operations such as create, read, update, and delete (CRUD operations) in a compliant way – i.e., by considering proper roles and permissions and storing and accessing the data accordingly. The health data is stored in a cloud environment, controlled, and monitored by ACME. Consequently, from a legal perspective, ACME acts as the data processor. However, due to the nature of its offered services, ACME has also to support data controllers to comply suitably. Therefore, it looks at the issue of GDPR compliance from both perspectives, of the data processor and data controllers. This is handled by a service level agreement between ACME and the HSP. ACME, as data controller, must be aware of how to properly process the patients' data because there could be a variety of harmful or threat events that could put even the patients' life at risk. Table 3.1 shows an example of the impact assessment for the case of ACME. The table has been built through a process modeled on the GDPR data processing impact assessment procedure of ACME. It describes, with a certain level of abstraction, the identified risks with respect to the various principles. In this table, the GDPR principles (described in Section 2.2.1) are listed in the first column. In the second column, some potential risks from the data subject perspective associated with the related principle are reported, potential mitigation solutions for these risks are mentioned in the third column, and in the last column, the consequences and risks for the data controller are identified. For example, the first row reports a risk for the *Confidentiality and integrity* of data (column 1). Data (such as the medical history of the patient) could be lost or corrupted because of a *hardware failure*. The consequences for the patient can be extremely high, because the healthcare data in this scenario is used for providing healthcare services such as medical prescriptions and missing or corrupted data may result in wrong diagnoses or in the impossibility to provide the service (column 2). For this reason, data storage must be reliable, by introducing more frequent backups or data replication (column 3). But these solutions consequently change the risk exposure for the company. In particular, data replication introduces the need for a complex network architecture, with all its associated risks. For example, business risks due to the rising costs, but also process risks (due to the difficulty of decision-making and network configuration). Under the same principle, health data leakages could happen because of *unauthorized access* to the sensitive data, which

Table 3.1: An example of risk assessment under GDPR.

GDPR Principle	Data Subject Risks	Potential Solutions	Data Controller Risks
Confidentiality and Integrity	Patient data losses or data corruption; wrong diagnoses by doctors;	Patient data backup; patient data replication;	Higher inside job possibility; time-consuming; rising costs; recovery procedure;
	Unauthorized access to patient health data; identity theft; loss of reputation;	Anonymization, pseudonymization and obfuscation; access control; encryption;	System slowdown; complexity; Possible implementation faults; functionality degradation;
Purpose limitation	Unintended permission; Unauthorized data disclosure;	Documenting the purposes in a transparent manner; restrict access to users' data;	Loss of public reputation;
Accuracy, Storage limitation, Data minimization	Disclosing undeleted inaccurate or medical history data; incorrect data may drive to discrimination or social pressure for patients;	Ensuring data accuracy; data cleaning algorithms; automated enforcement of deletion policies; regularly checking data collection;	Rising costs; possible implementation faults

may have bad consequences for the patient, such as *social stigma* and *discrimination*. Because of these consequences, *anonymization* or *pseudonymization* techniques may be required to be applied, but this in turn can introduce additional risks for the data controller, such as degradation of functionalities due to the fact that data can no longer be treated transparently (column 4).

This example demonstrates the consequences of the law: given that the data subject has certain fundamental rights, it is the data controller's responsibility to put in place the appropriate technical and organizational means to ensure that the rights of the data subject are respected. The endeavor to reduce the risks for the data subject, on the other hand, may result in an increase in the risk exposure for ACME, which may include risks other than those related to personal data. From these considerations we can see that it is likely that each stakeholder has different preferences for the various RMPs yielding different threat impact levels for each threat. Therefore, we must solve the problem of selecting the optimal risk management policy, which we formalize in the framework of multi-objective optimization in the next section.

To summarize, for the running example, we consider a set $\mathcal{S}$ containing two stakeholders, namely the Data Controller and the Data Subject, a list of 5 threats $T_1, ..., T_5$ shown in Table 4.2 (at page 52) and a list of associated security controls $c_1, ..., c_{25}$ shown in (the first two columns of) Table 4.3 (at page 54). Thus, we have 5 threats and 25 controls; the latter are associated to each threat as follows: $c_1, ..., c_5$ to T_1, $c_6, ..., c_{15}$ to T_2, $c_{16}, ..., c_{19}$ to T_3, $c_{20}, ..., c_{22}$ to T_4, and $c_{23}, ..., c_{25}$ to T_5. In the next section, we use the running example to illustrate the formal notions we introduce, albeit in a simplified form for the sake of simplicity and space. So, for instance, we will consider only 3 threats instead of 5 and only 5 security controls instead of 25. We observe that we use $c_1, ..., c_5$ as identifiers of the security controls in the following section for the sake of simplicity, but they have been

renamed in Table 4.3 where the whole set of controls is listed. The solution of the multi-objective optimization problem in its full generality is discussed later in Section 5.1.1.

3.2.2 Problem Formalization

Let $\mathcal{S}$ be a finite set of stakeholders and $\mathcal{T}$ a finite set of threats. For each stakeholder s in $\mathcal{S}$, we assume a mapping $i_s : \mathcal{T} \to \mathcal{I}$ that computes the impact level of the harmful events generated by a threat T when it occurs, where $\mathcal{I}$ is a sub-set of the reals denoting impact levels, intuitively $il_1 < il_2$ implies that the impact level il_1 is less severe than the impact level il_2.

Example . 1

Referring to the example in Section 3.2.1, the set $\mathcal{S}$ of stakeholders contains $s_1 =$ *Data controller (ACME)* and $s_2 =$ *Data subject (the patient)*. Consider the set $\mathcal{T}$ of threats to contain $T_1 =$ *Unlimited data storage*, $T_2 =$ *Unauthorized access*, and $T_3 =$ *Linkage attack*, as three potential threats. We may define the mappings i_{S_1} and $i_{S_2} : \mathcal{T} \to \mathcal{I}$ by means of a table as follows:

	T_1	T_2	T_3
i_{S_1}	0.6	0.2	0.3
i_{S_2}	0.3	0.5	0.6

The values in the first and second rows of the table denote the impact levels for each threat from the point of view of the data controller and data subject, respectively. For instance, the associated impact level to threat T_1 from the data controller point of view is 0.6 whereas from the data subject point of view is 0.3.

As shown in the example above, i_s is typically specified by using a tabular format. This is also the case for other mappings that we consider below.

Let $\mathcal{C}$ be a finite set of controls and $\{\mathcal{C}_T\}_{T \in \mathcal{T}}$ a family of finite set of controls; intuitively, $\mathcal{C}_T$ is the set of controls that, alone or in combination, may mitigate a threat T.

Example . 2

To mitigate the risk of threats in Example 1, we identify a family of set of controls $\{\mathcal{C}_{T_1}, \mathcal{C}_{T_2}, \mathcal{C}_{T_3}\}$ where $\mathcal{C}_{T_1} = \{c_1, c_2\}$, $\mathcal{C}_{T_2} = \{c_3, c_4\}$, and $\mathcal{C}_{T_3} = \{c_5\}$. For instance, c_1 can be *(Ensuring data minimization)*, c_2 *(Enabling data deletion)*, c_3 *(Ensuring secure storage)*, c_4 *(Logging access to personal data)*, and c_5 *(Ensuring data anonymization)*.

For each threat T in $\mathcal{T}$, we assume a mapping $\mu_T : \mathcal{C}_T \to [0..1)$ that quantifies the mitigation by a control in $\mathcal{C}_T$ on the impact of a threat T. Intuitively, $\mu_T(c)$ can have three possible statuses: (i) $\mu_T(c) = 0$ clarifies that the control c is not adopted and thus can not contribute in mitigating threat T, (ii) $0 < \mu_T(c) < 1$ means that the control c is adopted and partially mitigates the threat T, and (iii) $\mu_T(c) = 1$ represents that the control is adopted and fully mitigates T.

We are now in the position to define the impact residue of the threat T under a given mitigation mapping μ_T as:

$$ir_s(T) = i_s(T) \cdot (1 - \frac{\Sigma_{c \in \mathcal{C}_T} \mu_T(c)}{|\mathcal{C}_T|}) \tag{3.1}$$

We observe that the expression between parentheses is the mitigation obtained by adopting some of the controls in $\mathcal{C}_T$ associated to T and that the degree of effectiveness of a control c in mitigating a threat T is given by $\mu_T(c)$. Because of its importance, we introduce the following abbreviation:

$$m(T) = \frac{\Sigma_{c \in \mathcal{C}_T} \mu_T(c)}{|\mathcal{C}_T|} \tag{3.2}$$

that depends on the mitigation mapping μ_T (and since $ir_s(T) = i_s(T) \cdot (1 - m(T))$ also $ir_s(T)$ depends on μ_T), but we avoid making such a dependence explicit to simplify notation. Given a family $\{\mu_T\}_{T \in \mathcal{T}}$ of mitigation mappings, the overall impact residue for a given stakeholder $s \in \mathcal{S}$ is defined as $oir(s) = \Sigma_{T \in \mathcal{T}} ir_s(T)$, where $ir_s(T)$ is evaluated under the mitigation mapping μ_T. In other words, $oir(s)$ is the sum, over the set $\mathcal{T}$ of threats, of all impact residues, each one evaluated under the associated mitigation mapping in $\{\mu_T\}_{T \in \mathcal{T}}$.

Example . 3

For simplicity, we consider three possible values in the co-domain of μ_{T_1}, μ_{T_2}, and μ_{T_3}, namely 0 (the control does not mitigate the threat), 0.5 (the control partially mitigates the threat), and 1 (the control eliminates the threat). Continuing the previous examples, the mitigation mappings for T_1, T_2, and T_3 can be defined as follows:

$\langle \mu_{T_1}(c_1), \mu_{T_1}(c_2) \rangle$	$m(T_1)$
$\langle 0, 0 \rangle$	0
$\langle 0, 0.5 \rangle$	0.25
$\langle 0.5, 0 \rangle$	0.25
$\langle 0.5, 0.5 \rangle$	0.5
$\langle 1, 0 \rangle$	0.5
$\langle 0, 1 \rangle$	0.5
$\langle 1, 0.5 \rangle$	0.75
$\langle 0.5, 1 \rangle$	0.75

$\langle \mu_{T_2}(c_3), \mu_{T_2}(c_4) \rangle$	$m(T_2)$
$\langle 0, 0 \rangle$	0
$\langle 0, 0.5 \rangle$	0.25
$\langle 0.5, 0 \rangle$	0.25
$\langle 0.5, 0.5 \rangle$	0.5
$\langle 1, 0 \rangle$	0.5
$\langle 0, 1 \rangle$	0.5
$\langle 1, 0.5 \rangle$	0.75
$\langle 0.5, 1 \rangle$	0.75

$\langle \mu_{T_3}(c_5) \rangle$	$m(T_3)$
$\langle 0 \rangle$	0
$\langle 0.5 \rangle$	0.5

where the first column of each table lists all possible mitigation vectors that are assigned to the controls of $\mathcal{C}_{T_1}$, $\mathcal{C}_{T_2}$, and $\mathcal{C}_{T_3}$, respectively, when considering an arbitrary total order on the controls (in our case c_i comes before c_j if $i < j$ for $i, j \in \{1, ..., 5\}$. For instance, the vector $\langle 0.5, 0 \rangle$ means that c_1 partially mitigates T_1 whereas c_2 has no mitigation effect on T_1.

Example . 4

From the definitions of $i_s(T)$ and $m(T)$ in Examples 1 and 3, respectively, we can compute the impact residue $ir_s(T) = i_s(T) \cdot (1 - m(T))$ for each mitigation vector in Example 3 as follows:

T	$ir_{s_1}(T)$	$ir_{s_2}(T)$
T_1	$0.6 \times (1-0) = 0.6$	$0.3 \times (1-0) = 0.3$
	$0.6 \times (1-0.25) = 0.45$	$0.3 \times (1-0.25) = 0.225$
	$0.6 \times (1-0.25) = 0.45$	$0.3 \times (1-0.25) = 0.225$
	$0.6 \times (1-0.5) = 0.3$	$0.3 \times (1-0.5) = 0.15$
	$0.6 \times (1-0.5) = 0.3$	$0.3 \times (1-0.5) = 0.15$
	$0.6 \times (1-0.5) = 0.3$	$0.3 \times (1-0.5) = 0.15$
	$0.6 \times (1-0.75) = 0.15$	$0.3 \times (1-0.75) = 0.06$
	$0.6 \times (1-0.75) = 0.15$	$0.3 \times (1-0.75) = 0.06$
T_2	$0.2 \times (1-0) = 0.2$	$0.5 \times (1-0) = 0.5$
	$0.2 \times (1-0.25) = 0.15$	$0.5 \times (1-0.25) = 0.375$
	$0.2 \times (1-0.25) = 0.15$	$0.5 \times (1-0.25) = 0.375$
	$0.2 \times (1-0.5) = 0.1$	$0.5 \times (1-0.5) = 0.25$
	$0.2 \times (1-0.5) = 0.1$	$0.5 \times (1-0.5) = 0.25$
	$0.2 \times (1-0.5) = 0.1$	$0.5 \times (1-0.5) = 0.25$
	$0.2 \times (1-0.75) = 0.05$	$0.5 \times (1-0.75) = 0.125$
	$0.2 \times (1-0.75) = 0.05$	$0.5 \times (1-0.75) = 0.125$
T_3	$0.3 \times (1-0) = 0.3$	$0.6 \times (1-0) = 0.6$
	$0.3 \times (1-0.5) = 0.15$	$0.6 \times (1-0.5) = 0.3$

where the second and third columns represent the computed impact residues under all possible mitigation mappings and the corresponding threat (in the rows) for s_1 and s_2, respectively. For instance, the impact residue under the mitigation vector $\langle 0.5, 1\rangle$ for T_1 from the point of view of s_1 is 0.15 whereas it is 0.06 for s_2. Recalling that $oir(s) = \Sigma_{T\in\mathcal{T}} ir_s(T)$, the overall impact residue for s_1 is $oir(s_1) = 0.6+0.2+0.3 = 1.1$ and for s_2 is $oir(s_2) = 0.3+0.5+0.6 = 1.4$ where $\langle \mu_{T_1}(c_1), \mu_{T1}(c_2)\rangle = \langle 0, 0\rangle$, $\langle \mu_{T_2}(c_3), \mu_{T_2}(c_4)\rangle = \langle 0, 0\rangle$, and $\langle \mu_{T_3}(c_5)\rangle = \langle 0\rangle$.

The *Multi-Stakeholder Risk Minimization Problem* (MSRMP) amounts to solve the following multi-objective optimization problem:

$$min_{\langle \mu_T\rangle_{T\in\mathcal{T}}} \quad \langle oir(s)\rangle_{s\in\mathcal{S}} \tag{3.3}$$

where $\langle\ \rangle_{T\in\mathcal{T}}$ and $\langle\ \rangle_{s\in\mathcal{S}}$ are the vectors of all mitigation mappings and overall impact residues (under the associated mitigation mappings) according to arbitrary total orders over $\mathcal{T}$ and $\mathcal{S}$, respectively. In other words, the MSRMP consists of finding the vector of mitigation mappings that allows for minimizing the overall impact residues of the

stakeholders. A solution of (3.3) is a vector $\langle \mu_T \rangle_{T \in \mathcal{T}}$ of mitigation mappings that is Pareto optimal (see, e.g., [MA04]), i.e. it is such that if there does not exist another vector $\langle \mu'_T \rangle_{T \in \mathcal{T}}$ of mitigation mappings such that $oir(s) \leq oir'(\overline{s})$ for each $s \in \mathcal{S}$ and $oir'(\overline{s}) < oir(s)$ for at least one $\overline{s} \in \mathcal{S}$ where oir and oir' are the overall impact residues under the family $\{\mu_T\}_{T \in \mathcal{T}}$ and $\{\mu'_T\}_{T \in \mathcal{T}}$ of mitigation mappings, respectively.

We make two observations. First, (3.3) considers only the impact and not the likelihood since, as already discussed earlier, we assume that the stakeholders in $\mathcal{S}$ agree on both the set $\mathcal{T}$ of threats and their likelihood. As a consequence, minimizing the impact is equivalent to minimizing the risk since the latter is the product of impact and likelihood, and it is a constant and positive value for each stakeholder in $\mathcal{S}$. This is a natural assumption to make in the context of the GDPR, whereby the data controller is accountable for the risk assessment and needs to guarantee that the risks of the data subject are kept to a minimum. The second observation is about solving (3.3). Indeed, it is possible to re-use the cornucopia of techniques available for Multi-Objective Optimization Problem (MOOP); see, e.g., [MA04]. However, for some of the techniques to be applicable, it is crucial to have a definition of the functions i_s and μ_T for $T \in \mathcal{T}$ in closed form. This is rarely the case for the use case scenarios we have in mind. Instead, experts are typically able to define both i_s and μ_T as discrete functions, i.e. by associating a given impact level with a certain threat for i_s and quantifying the amplitude of the mitigation associated to a given control in $\mathcal{C}_T$ for μ_T. The examples above present this kind of definitions for such functions by using tables.

As a consequence of the two observations above, we make the following assumptions. First, each stakeholder s in $\mathcal{S}$ provides a definition of the mapping i_s as a finite set of pairs of the form (T, il) where T is a threat in $\mathcal{T}$ and il is an impact level in a finite set $\mathcal{I}$ of values (i.e., $\mathcal{I} = \{0, 1, 2, 3, 4\}$ where 0 denotes a negligible impact, 4 a dramatic impact, and the values in between increasing values). Second, for each threat T in $\mathcal{T}$, the stakeholder in charge of the risk management process (i.e., the data controller in the case of the GDPR) defines the mapping $\mu_T : \mathcal{C}_T \to \mathcal{A}$ with $\mathcal{A}$ a finite set of values in the interval [0..1]; in other words, μ_T is specified as a finite set of pairs of the form (c, p) where c is a control in $\mathcal{C}$ and p is the amplitude of the mitigation of the impact of the threat T when adopting the control c. For instance, we can take $\mathcal{A} = \{0, 0.5, 1\}$, so that $\mu_T(c) = 0$ means that control c has no effect in mitigating the threat T, $\mu_T(c) = 0.5$ has partial effect on T, and $\mu_T(c) = 1$ has full effect. Under these assumptions, we obtain an instance of (3.3) that belongs to a particular class of MOOP called Multi-Objective Combinatorial Optimization Problems (MOCOPs); see, e.g., [Kla09]. We observe that finding all Pareto optimal solutions of such instances of (3.3) requires, in the worst case, to search among $\Pi_{T \in \mathcal{T}}(k^{|\mathcal{C}_T|} - 1)$ candidate sets of controls for $k = |\mathcal{A}|$ the number of distinct real values in the co-domain of the mappings μ_T for all T in $\mathcal{T}$. The -1 in the expression considers that it is never the case that all controls in $\mathcal{C}_T$ will be adopted; this is a reasonable assumption because of multiple reasons, including lack of skills to manage several different technologies on which the controls are based and

constraints in costs. Indeed, this implies the decidability of the instances of the MSRMP that we consider in the rest of the work. We observe that, despite their decidability, solving these instances of the MSRMP may be quite a challenge from a computational point of view because the number of possible solutions in which to search for the optimal ones is exponential in the size of $\mathcal{C}_T$ for $T \in \mathcal{T}$. In the rest of this section, we describe a strategy to manage this problem and in Section 5.2, we propose an experimental evaluation of some refinements and study the scalability of the proposed approach in practice.

Example . 5

As described above, by considering $k = 3$ possible values for the mappings μ_{T_1}, μ_{T_2}, and μ_{T_3} introduced in Example 3, the search for finding optimal solutions is among $\Pi_{T\in\mathcal{T}}(k^{|\mathcal{C}_T|} - 1) = (3^2 - 1) \times (3^2 - 1) \times (3^1 - 1) = 128$ candidates. Note that we do not consider the situation in which all controls are in place, as this would yield a risk equal to zero, thereby making the search for optimal solutions trivial. This is reasonable in practice since it is unlikely that the stakeholders will be able to adopt all security controls in $\{\mathcal{C}_T\}_{T\in\mathcal{T}}$ because of other constraints such as those related to budget and required security skills for their deployment.

The example above and the concern in the previous paragraph recalls our **RQ4.** *How can we explore and find the optimum solutions among all conceivable risk management policies?* To address this question and simplify the solution of the instances of (3.3), we consider an associated problem derived from (3.3), by introducing a variable x_T to replace $1 - m(T)$ and obtain:

$$\begin{array}{l} min_{\langle x_T\rangle_{T\in\mathcal{T}}} \ \ \langle \frac{1}{|\mathcal{T}|}\Sigma_{T\in\mathcal{T}} \ (i_s(T) * x_T)\rangle_{s\in\mathcal{S}} \\ \text{subject to } x_T \in \{1 - m(T)\} \ \text{ for each } T \in \mathcal{T} \end{array} \tag{3.4}$$

where $m(T)$ is the expression defined in 3.2, $\langle x_T\rangle_{T\in\mathcal{T}}$ is the vector of variables representing mitigation amplitudes when considering an arbitrary total order over $\mathcal{T}$. For each threat T in $\mathcal{T}$, we have that $|\{1 - m(T)\}|$ is the number of distinct sum values, divided by the number of controls in $\mathcal{C}_T$, that can be obtained by adding values in $\mathcal{I}$ (that, in our examples, is the set $\{0, 0.5, 1\}$) according to a μ_T that induces a value $m(T)$. The space of solutions of the modified version of (3.4), is thus $\Pi_{T\in\mathcal{T}}|\{1 - m(T)\}|$ which may be remarkably less than $\Pi_{T\in\mathcal{T}}(k^{|\mathcal{C}_T|} - 1)$. For instance, consider Example 3, the first two tables contain 8 different mitigation vectors with only 4 different values for the function $m(\cdot)$.

Example . 6

Recall Example 3, consider only the values of $m(T)$ that are distinct, and derive the values $x_T = 1 - m(T)$ for each $T \in \{T_1, T_2, T_3\}$:

$m(T_1)$	X_{T_1}	$m(T_2)$	X_{T_2}	$m(T_3)$	X_{T_3}
0	1	0	1	0	1
0.25	0.75	0.25	0.75	0.5	0.5
0.5	0.5	0.5	0.5		
0.75	0.25	0.75	0.25		

The set of possible solutions of (3.4) is the set of all triples of the form $\langle x_{T_1}, x_{T_2}, x_{T_3}\rangle$ whose values are taken from the three tables above, and thus the size of such a set is $4 \times 4 \times 2 = 32$. Observe that this is one-fourth of the size of the set of potential solutions to the original problem (3.3), namely $\Pi_{T\in\{T_1,T_2,T_3\}}(k^{|\mathcal{C}_T|} - 1) = (3^2 - 1) \cdot (3^2 - 1) \cdot (3^1 - 1) = 128$. For larger problem instances, the reduction is much more substantial as we will see in Section 5.2 below. By considering the 32 triples $\langle x_{T_1}, x_{T_2}, x_{T_3}\rangle$, we can derive the values of the overall impact values for the two stakeholders by recalling that $oir(s) = ir_s(T_1) + ir_s(T_2) + ir_s(T_3)$, $ir_s(T) = i_s(T) \cdot (1 - m(T))$ from (3.1), (3.2, and $x_T = 1 - m(T)$ for $s \in \{s_1, s_2\}$ and for $T \in \{T_1, T_2, T_3\}$. Also, recall that the definition of $i_s(\cdot)$ can be found in Example 4. The pairs $(oir(s_1), oir(s_2))$ so computed are plotted in Figure 3.3 where the x-axis shows the values of $oir(s_1)$ and the y-axis those of $oir(s_2)$.

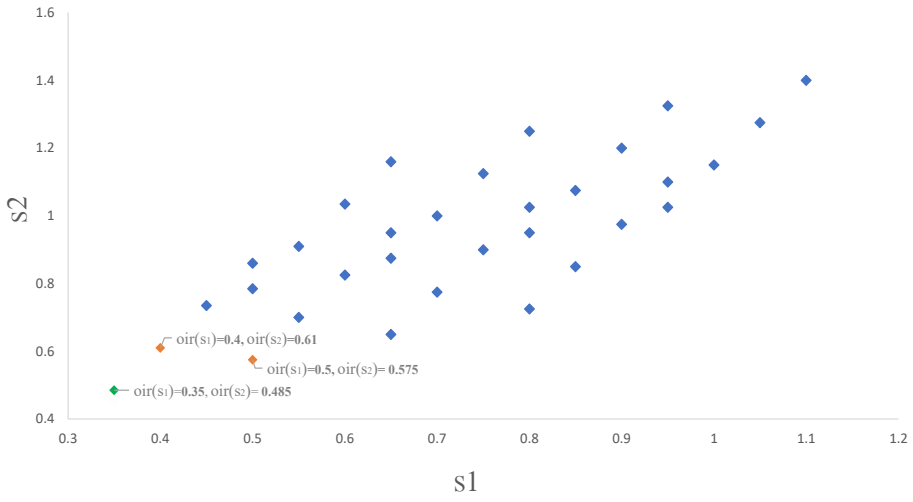

Figure 3.3: The solution points.

It is then immediate to see that the point $(0.35, 0.485)$ at the bottom left (in green) is the Pareto optimal solution. We also observe that the two points in orange are not dominated by any other points but the optimal one.

Indeed, it is possible to find solutions of (3.3) corresponding to those of the simplified version of (3.4) by adapting the procedure above. Let $\langle x_T^* \rangle_{T \in \mathcal{T}}$ be a solution for (3.4). By definition and the simplifying assumption above, there must exist μ_T^* such that $x_T^* = 1 - m(T) = 1 - \frac{\Sigma_{c \in \mathcal{C}_T} \ \mu_T^*(c)}{|\mathcal{C}_T|}$ for each $T \in \mathcal{T}$ and it is thus immediate to discover all the solutions of (3.3).

Example . 7

We explain how it is possible to derive the sets of controls associated to a certain triple $\langle x_{T_1}^*, x_{T_2}^*, x_{T_3}^* \rangle$. To illustrate, we consider the (orange) point in Figure 3.3 with coordinates $(0.4, 0.61)$ that is associated to the triple $\langle x_{T_1}^*, x_{T_2}^*, x_{T_3}^* \rangle = \langle 0.25, 0.5, 0.5 \rangle$. From (3.2) and $x_T = 1 - m(T)$, it is immediate to derive that

$$\frac{\mu_{T_1}(c_1) + \mu_{T_1}(c_2)}{2} = 1 - x_{T_1}^* \quad \frac{\mu_{T_2}(c_3) + \mu_{T_2}(c_4)}{2} = 1 - x_{T_2}^* \quad \frac{\mu_{T_3}(c_5)}{1} = 1 - x_{T_3}^*$$

so that we are left with the problem of enumerating all mitigation mappings $\mu_{T_1}(\cdot), \mu_{T_2}(\cdot), \mu_{T_3}(\cdot)$ satisfying the three equalities above. The following table lists all possible such mappings:

	$x_{T_1}^* = 0.25$		$x_{T_2}^* = 0.5$		$x_{T_3}^* = 0.5$
	$\mu_{T_1}(c_1)$	$\mu_{T_1}(c_2)$	$\mu_{T_2}(c_3)$	$\mu_{T_2}(c_4)$	$\mu_{T_3}(c_5)$
$\mathbb{S}_1$	1	0.5	0.5	0.5	0.5
$\mathbb{S}_2$	0.5	1	0.5	0.5	0.5
$\mathbb{S}_3$	1	0.5	1	0	0.5
$\mathbb{S}_4$	0.5	1	1	0	0.5
$\mathbb{S}_5$	1	0.5	0	1	0.5
$\mathbb{S}_6$	0.5	1	0	1	0.5

The obvious question is the computational complexity of enumerating all possible mitigation mappings $\mu_T(\cdot)$ such that

$$\frac{\Sigma_{c \in \mathcal{C}_T} \mu_T(c)}{|\mathcal{C}_T|} = 1 - x_T^* \tag{3.5}$$

for each $T \in \mathcal{T}$; notice that the three equalities in Example 7 are instances of (3.5). Indeed, if there exists a (practically) efficient algorithm to enumerate the mitigation mappings satisfying (3.5), we can hope that solving instances of (3.4) and then using such an algorithm to derive the corresponding solutions of (3.3) is an efficient alternative to solving directly the latter as the number of the possible solutions of (3.4) is smaller (as we have seen in Example 6 and even substantially so as we will see in Section 5.2) than those of (3.3).

To answer this question, we consider the Subset Sum Problem (SSP) with multiplici-

ties [CLRS01], i.e. given a multiset X of integers and an integer s, does any non-empty multisubset of X sum to s? Solving the instances of (3.5) for each $T \in \mathcal{T}$ is equivalent to solving an instance of the SSP under the natural assumption that x_T^* and the values in $\mathcal{A}$ are real numbers that can be represented as $v \cdot 10^{-d}$ for v and d positive integers such that $0 < v \cdot 10^{-d} < 1$. To see this, observe that all the values in $\mathcal{A} \cup \{1 - x_T^*\}$ can be transformed to integers by multiplying each one by their maximum exponent d when represented as $v \cdot 10^{-d}$, the integers so obtained from the values in $\mathcal{A}$ are added to the multiset X, each one with multiplicity equal to the number of controls in $\mathcal{C}_T$ for $T \in \mathcal{T}$, and the integer obtained from $1 - x_T^*$ is set to s. Several different algorithms are available to solve this problem with different complexities ranging from exponential to (pseudo-)polynomial (see, e.g., [CLRS01]). The most naive algorithm (with exponential worst-case complexity) amounts to cycling through all multisubsets of X and, for each one, check if it sums to s. To solve the SSP, it is possible to stop as soon as one solution is found, but in our case, we need to find all possible solutions. Indeed, the naive algorithm can be trivially adapted to do this, resulting in exponential best-case and worst-case complexity. Despite being in such a complexity class, the naive algorithm turns out to give satisfactory results in practice because the instances derived from (3.5) are typically small because the cardinality of $\mathcal{C}_T$ is relatively small for each $T \in \mathcal{T}$ or can be reduced by exploiting the knowledge of security experts. We will discuss this issue in Section 5.1 below.

In our first work [MSR20] in this context, we have considered a simpler instance of the MSRMP introduced above, called the Multi-Stakeholder Risk Trade-off Analysis Problem (MSRToAP). This work [MSR20] considers a similar—albeit simpler—optimization problem, allowing for finding the best possible solutions among a (finite and small) set of possible RMPs. Indeed, such solutions are not guaranteed to be Pareto optimal as those of the MSRMP considered in our journal paper [MR22]. Additionally, in [MSR20], no methodology to identify the set of possible RMPs is provided whereas this work provides a structured methodology for the definition of the whole set of RMPs via the notion of MSRMP.

Chapter 4

Multi-Stakeholder Risk Assessment Methodology

In the previous chapter, we introduced the Multi-Stakeholder Risk Minimization Problem (MSRMP) and formalized it in mathematical terms that enable us to perform a risk trade-off when we have multiple stakeholders in a scenario. As can be seen in Figure 4.1, this chapter is dedicated to proposing the multi-stakeholder risk assessment methodology in which the automated technique to solve MSRMP instances can be effectively integrated. Initially, Section 4.1 briefly describes and highlights the processes (activities) that we consider for conducting the multi-stakeholder risk assessment. Such a technique is described in Section 4.2. The definition of the risk impact levels relies on the two important fac-

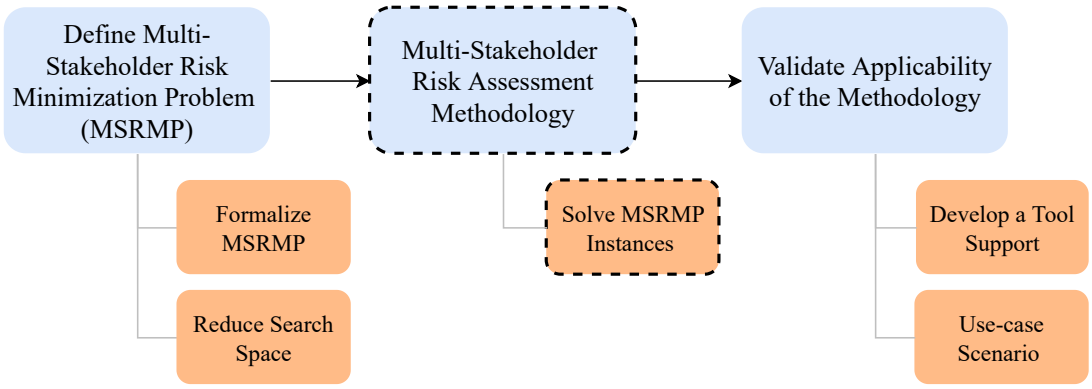

Figure 4.1: The highlighted part (dashed lines) illustrates the contribution of Chapter 4 in accordance with the contribution flow outlined in Section 1.2.

tors, namely *stakeholders' protection criteria* and *protection goals*. The former is used to determine the threat's aversion level from each stakeholder's perspective, while the latter measures how critical the threat is. We include these two factors to define and evaluate impact levels in Section 4.2.1 and Section 4.2.2. To conclude this section, we present a brief discussion explaining the use of these two factors in Section 4.3 to reduce the subjectivity of impact evaluation.

4.1 Introduction

It is imperative that we conduct a comprehensive risk assessment throughout the system design phase to determine which kind of RMP should be implemented to minimize the potential risks. The ideal approach to accomplish this—i.e., to determine various RMPs and their risk reduction levels—is to do so in a scientific and principled manner, and the most effective way to do so is to make the process as quantitative as possible. Therefore, in order to solve instances of the MSRMP, we initially need to specify the procedures—that contain parameters and processes—that need to be considered. This need recalls our **RQ2.** *What are the procedures to collect the required parameters for quantifying the risk levels in a multi-stakeholder scenario?* As we mentioned earlier in Section 1.1, this question deals with processes and parameters that need to be specified in order to quantify risk impact levels.

The contributions below are made in order to address the research question mentioned above:

- First, we propose (in Section 4.2.1) a way to define impact levels based on stakeholders' protection criteria.
- In the second attempt (in Section 4.2.2), we propose a less subjective definition for impact levels by considering the protection goals introduced in Section 2.2.3.

In the following, we will highlight all the activities that we consider for defining the impact levels according to both contributions mentioned above. An overview of these activities (sub-processes) is depicted in Figure 4.2. In this figure, the activities are specified within the three risk assessment steps, as discussed in Section 2.1.3. An abstract definition is assigned to the activities of each step of the risk assessment, which are `Artifact Preparation`, `Association Processes` and `Computation Processes`. In the *risk identification* step, some activities must be carried out in order to prepare artifacts for the next step of risk assessment. These activities include: identifying threats, identifying the mitigation controls, determining the stakeholders involved in the risk assessment and their protection criteria (preferences), and selecting the protection goals as we mentioned in

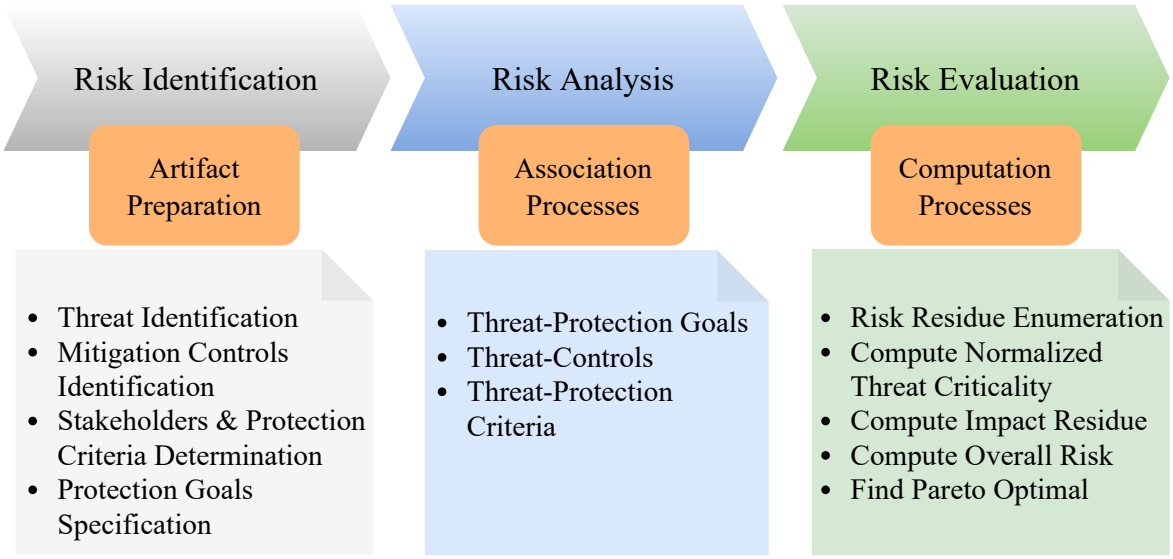

Figure 4.2: Overview of the activities in the proposed multi-stakeholder risk assessment.

Table 4.1: List of variables used in Chapter 4.

Variable	**Description**
$\mathcal{T}$	The set of threats
$\mathcal{S}$	The set of stakeholders
$\mathcal{C}$	The set of controls
$\mathcal{P}$	The protection criteria
$\mathcal{G}$	The set of protection goals
$\mathcal{C}_T$	A family of controls associated to $T \in \mathcal{T}$
x_T	The residual risk for threat T
i_s	The impact level for each stakeholder $s \in \mathcal{S}$
$\mu_T(c)$	The impact of T after applying control $c \in \mathcal{C}_T$
PW_p^s	The associated weight to the preference p of stakeholder s
il_{max}	Donates to the maximum impact level
$al_p^s(T)$	The aversion level of threat T for the preference p of stakeholder s
OW_T	The observation weight for threat T
NTC_T	The normalized threat criticality value for threat T
AG_T	The number of goals in $\mathcal{G}$ affected by a threat $T \in \mathcal{T}$
$oir(s)$	The overall impact residue for stakeholder s

Section 2.2.3. In the middle step—i.e., the *risk analysis*—some association mappings will require to be done manually. These mappings contain the association between threats and protection goals, the association between threats and the identified mitigation controls, and the association between threats and each stakeholder's protection criteria which are obtained from the previous step. Finally, in the *risk evaluation* step, some computation processes will need to be performed to obtain overall risk levels. Therefore, risk residues are automatically enumerated in this step, then threat criticality and impact levels will calculate. In the end, these processes lead to generating the feasible set—i.e., a list of overall risk levels—that needs to apply the Pareto optimality algorithm to find optimal solutions.

After providing an overview of the required tasks to conduct the multi-stakeholder risk assessment, the following section defines instances of our problem, the MSRMP, with the objective of identifying the best solutions that minimize risk for stakeholders. A summary of all used variables along with a brief description of each is provided in Table 4.1.

4.2 Defining Instances of the MSRMP

Our main goal is to assist in the identification of the best possible set of controls to minimize the risk for all stakeholders. This has been formalized as solving an appropriate instance of the MSRMP introduced in Section 3.2.2. To specify instances of the MSRMP in either statement (3.3) or (3.4), we consider additional information that is typically available in

Table 4.2: An example of possible threat scenarios and associated malicious activities in the ACME scenario.

Threats ($\mathcal{T}$)	Possible malicious activity
T_1- Unlimited data storage	Personal data is kept stored longer than necessary for the purposes by ACME.
T_2- Unauthorized access and disclosure	Due to over-privileged or inadequate controls, insiders (i.e., a medical practitioner or an ACME's staff) modify patients' data or disclose by mistake.
T_3- Linkage attack	Patients and their personal data can re-identify in de-identified data sets by outsiders' malicious.
T_4- Denial of service	Attackers can disrupt the communication channel between patients and the healthcare service provider to prevent data from being uploaded to the server.
T_5- Threat to intervenability	ACME does not implement a procedure (technical and /or processes) that allows the patients to rectify, erase, or block individual data.

many methodologies for risk assessment. In the rest of this section, we first (Section 4.2.1) consider the problem statement (3.4) and discuss an approach to derive the risk residue i_s for each stakeholder s that yields a problem with a reduced search space whose solutions can be used to derive optimal mitigation mappings as explained at the end of Section 3.2.2. We will see that this approach requires the stakeholder s to take several decisions that are highly subjective, and this may lead to bias. Then (Section 4.2.2), we propose an approach that aims to reduce the level of subjectivity in defining the risk residue i_s that requires to consider the general problem statement (3.4). We will discuss how also in this case it is possible to first solve a problem with a reduced search space and then to derive optimal mitigation mappings. Both approaches require to identify a set $\mathcal{S}$ of stakeholders, a set $\mathcal{T}$ of threats, a family $\{\mathcal{C}_\mathcal{T}\}_{T\in\mathcal{T}}$ of sets of controls (each one associated to a threat $T \in \mathcal{T}$), and be able to define the mapping i_s that quantifies the impact level for each stakeholder $s \in \mathcal{S}$ and the residual risk x_T for each threat T that results from applying a certain set of controls (or, equivalently, from selecting a certain mitigation mapping μ_T). The approaches presented in Sections 4.2.1 and 4.2.2 differ in the definition of i_s. For this reason, we preliminarily consider the definitions of the other parameters, namely $\mathcal{T}$, $\{\mathcal{C}_\mathcal{T}\}_{T\in\mathcal{T}}$, and x_T.

As reviewed earlier, the literature lists several approaches (e.g., [Sho14, WJ15]) dealing with threat identification together with appropriate mitigation controls that allow us to define the set $\mathcal{T}$ of threats and the family $\{\mathcal{C}_T\}_{T\in\mathcal{T}}$ of sets of controls associated to the threats in $\mathcal{T}$. The decision to select a method or another depends on the specific needs and specific concerns (see, e.g., the discussion in [SCO+18]). For instance, Microsoft STRIDE [Sho14] is a well-established threat modeling to identify security threats according to a predefined classification of threat types. It is an acronym for *Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service,* and *Elevation of privilege*. These threat types represent the violation of the primary security properties: authentication, integrity, non-repudiation, confidentiality, availability, and authorization. LINDDUN [WJ15] is another well-known threat modeling approach to identify privacy threats, and it is an acronym for *Linkability, Identifiability, Non-repudiation, Unawareness, Detectability, Disclosure of information,* and *Non-compliance*. Similar to STRIDE, also, these represent violations of properties characterizing different dimensions of privacy. For concreteness, an instance of the set $\mathcal{T}$ is shown in Table 4.2 and an instance of the family $\{\mathcal{C}_\mathcal{T}\}_{T\in\mathcal{T}}$ can be found in the first two columns of Table 4.3 (for instance, consider T_4 = Denial of service, $\mathcal{C}_{T_4}$ is associated with three controls, namely *Enabling off-line authentication*, *Network monitoring*, and *Prevention mechanisms for DoS attacks like firewalls, IDS, etc*); both are related to the running example introduced in Section 3.2.1. For the applicability of the method proposed in this work, any methodology that allows for the definition of $\mathcal{T}$ and $\{\mathcal{C}_\mathcal{T}\}_{T\in\mathcal{T}}$ can be used. We are left with the problem of defining i_s for $s \in \mathcal{S}$ and x_T for $T \in \mathcal{T}$. Concerning the latter, recall that according to the problem statement (3.4) and 3.1, the risk residue $x_T \in \{1-m(T)\}$ with $m(T) = 1 - \frac{\Sigma_{c\in\mathcal{C}}\mu_T(c)}{|\mathcal{C}_T|}$, i.e. x_T is the risk residue obtained

by applying a certain combination of the security controls available in $\mathcal{C}_T$ for the threat T according to the mitigation mapping μ_T. Recall also that $\mu_T(c)$ measures the impact of T after applying control $c \in \mathcal{C}_T$ and thus $m(T)$ measures the aggregated mitigating effect of selecting a given set of controls in $\mathcal{C}_T$ on the risk of T materializing (under the assumption that the mitigations are independent of each other). The third and fourth columns of Table 4.3 show a given mitigation mapping μ_T and the associated value x_T of the resulting

Table 4.3: Threats with associated security controls (first two columns) together with a mitigation mapping (third column) and the resulting risk residue (fourth column). Legend: each control is associated to a mitigation level among three possible values ○ = 0 (the control has not been selected for implementation), ◐ = 0.5 (the control has been selected for implementation, but it is only partially effective to mitigate T), or ● = 1 (the control has been selected for implementation, and it is fully effective to mitigate T).

Threats ($\mathcal{T}$)	Controls $\{\mathcal{C}_T\}_{T\in\{T_1,T_2,T_3,T_4,T_5\}}$	Mitigation Mapping μ_T	Risk Residue x_T
T_1	c_1) Purpose specification	●	0.4
	c_2) Ensuring limited data processing	●	
	c_3) Ensuring purpose related processing	◐	
	c_4) Ensuring data minimization	◐	
	c_5) Enabling data deletion	○	
T_2	c_6) Ensuring data subject authentication	●	0.35
	c_7) Ensuring staff authentication	●	
	c_8) Ensuring device authentication	◐	
	c_9) Logging access to personal data	◐	
	c_{10}) Performing regular privacy audits	○	
	c_{11}) Ensuring data anonymization	◐	
	c_{12}) Providing confidential communication	●	
	c_{13}) Providing usable access control	◐	
	c_{14}) Ensuring secure storage	●	
	c_{15}) Ensuring physical security	◐	
T_3	c_{16}) Providing confidential communication	●	0.25
	c_{17}) Logging access to personal data	◐	
	c_{18}) Ensuring data subject authentication	●	
	c_{19}) Ensuring data anonymization	◐	
T_4	c_{20}) Enabling offline authentication	○	0.83
	c_{21}) Network monitoring	◐	
	c_{22}) Prevention mechanisms for DoS attacks like firewalls, etc.	○	
T_5	c_{23}) Informing data subjects about data processing	◐	0.66
	c_{24}) Handling data subject's change requests	◐	
	c_{25}) Providing data export functionality	○	

risk residue. It will be the task of an automated solver to explore the space of all possible values of x_T and find those that are Pareto-optimal solutions of the MSRMP instance (3.4) so that it is possible to derive the optimal mitigation mappings as described at the end of Section 3.2.2; see Section 4.2.1). As already said above, we will see that finding optimal values for x_T is crucial also for solving instances of the general problem statement (3.3); see Section 4.2.2.

4.2.1 Defining Impacts Levels According to Stakeholders: A First Attempt

Different stakeholders have different criteria that define what they consider risky. Data controllers (e.g., companies) typically choose business impact criteria, such as financial impact or reputation, whereas data subjects (e.g., individuals) evaluate risk based on impact on their personal sphere. For the running example introduced in Section 3.2.1, we consider the *social situation, individual freedom, financial situation* [OS14], and *health condition* as the data subject protection criteria while for the data controller, *reputational situation* and *financial situation* are the protection criteria, which are linked to indirect or direct pecuniary losses. Additionally, each stakeholder has different preferences, which result in different importance given to different criteria; e.g., in the running example, the health condition criterion is more momentous than others for patients. We capture these high-level stakeholder preferences by assigning a weight to each stakeholder's protection criterion. The associations among stakeholders, protection criteria, and weights are shown in the first three columns of Table 4.4. Formally, we assume the availability of a set $\mathcal{P}$ of protection criteria, a family $\{PW_p^s\}_{p\in\mathcal{P},s\in\mathcal{S}}$ of weights associated to a preference p for each stakeholder s besides the definitions of $\mathcal{T}$, $\{\mathcal{C}\}_{T\in\mathcal{T}}$, and x_T for $T \in \mathcal{T}$ as discussed above in this section.

The additional information in $\mathcal{P}$ and $\{PW_p^s\}_{p\in\mathcal{P},s\in\mathcal{S}}$ are used to define the impact level i_s by giving a quantitative evaluation of the negative influence that a threat $T \in \mathcal{T}$ may have on a preference $p \in \mathcal{P}$ for a certain stakeholder $s \in \mathcal{S}$. The intuition is to characterize how each threat is perceived as more or less dangerous by each stakeholder in relation to his/her own protection criteria. For instance, in the context of the running example, it is very unlikely that excessive storage of patients' health data would damage the data controller's reputation; by increasing stored data, there is financial damage to the data controller, causing the cost of storage and management of the IT infrastructure. On the other hand, the reputation of patients is not affected by excessive storage of personal data; indeed, a larger amount of stored data increases the impact of data breaches and leaks on the rights and freedoms of patients. For this, we assign an impact value in $\mathcal{IL}$ (recall that this set typically contains a finite set of integer values from 0 to 4 included) to the level of aversion that each stakeholder s has for a threat T acting on a given protection

Table 4.4: The assigned impacts to each stakeholders' preferences for each threat in our scenario.

Stakeholders ($\mathcal{S}$)	Protection Criteria ($\mathcal{P}$)	Weights PW_p^s	Aversion level (al_p^s)				
			T_1	T_2	T_3	T_4	T_5
Data Subject	Health condition	0.4	0	4	0	3	4
	Individual freedom	0.2	0	2	4	3	3
	Social situation	0.3	1	2	3	0	3
	Financial situation	0.1	0	3	1	0	3
Data Controller	Reputational situation	0.4	1	2	3	2	2
	Financial situation	0.6	2	2	3	3	2

criterion p. Formally, we assume the definition of an *aversion mapping* $al_p^s : \mathcal{P} \to \mathcal{IL}$ for each preference $p \in \mathcal{P}$ and stakeholder $s \in \mathcal{S}$. At this point, we are in the position to define i_s by combining the weight PW_p^s and the mapping al_p^s as follows:

$$i_s(T) = \frac{1}{|il_{max}|} \sum_{p \in \mathcal{P}} al_p^s(T) \ \times \ PW_p^s \tag{4.1}$$

where $il_{max} \in \mathcal{IL}$ represents the maximum impact level (in our case, it is 4). The crux to specify i_s is thus to define the family $\{al_p^s\}_{p \in \mathcal{P}, s \in \mathcal{S}}$ of aversion mappings. This can be done as shown in the fourth column of Table 4.4 where each threat $T \in \mathcal{T}$ gets an aversion level al_p^s between 0 and 4 (recall that 0 means no, 1 low, 2 moderate, 3 critical, and 4 catastrophic impact) for each protection criterion p and stakeholder s. Intuitively, the values are assigned by answering the question "For the stakeholder s, what would be the impact level on the criterion p if the threat T happen?"

To illustrate, consider Table 4.4 in which the aversion level of the *health condition* for the second threat (T_2) according to the data subject ($s = DS$) is 4 and thus the value of $i_{DS}(T_2)$ will be $\frac{(0.4\times4)+(0.2\times2)+(0.3\times2)+(0.1\times3)}{4} = 0.725$ according to (4.1).

To summarize, we have described an approach to define i_s by assuming the capability of identifying protection criteria for each stakeholder (i.e. being able to define the set $\mathcal{P}$), of quantifying the relevance of each such criterion (in a scale between 0 and 1) for each stakeholder (i.e. being able to define the family $\{PW_p^s\}_{p \in \mathcal{P}, s \in \mathcal{S}}$), and assigning an aversion level of each stakeholder when a threat impacts a given protection criterion (i.e. defining the family $\{al_p^s\}_{p \in \mathcal{P}, s \in \mathcal{S}}$). This allows us to define an instance of the MSRMP (3.4) which, as we will see in the following, can be solved by using available techniques and then, as described at the end of Section 3.2.2, to identify the set of Pareto optimal mitigation mappings that minimize the risks with respect the various stakeholders. However, we observe that it may

be non-obvious to quantify the weights in $\{PW_p^s\}_{p\in\mathcal{P},s\in\mathcal{S}}$ and the aversion level mappings in $\{al_p^s\}_{p\in\mathcal{P},s\in\mathcal{S}}$ as their definitions are quite subjective for each stakeholder. This is somehow unavoidable because it is up to each stakeholder to define i_s, however it is important to mitigate possible bias that would make the solutions of the corresponding instance of the MSRMP (3.4) hardly useful in practice or even detrimental because of an over or under estimation of the risk levels with negative business or privacy impacts, respectively, on some stakeholders. We can consider assigning the definitions of $\{PW_p^s\}_{p\in\mathcal{P},s\in\mathcal{S}}$ and $\{al_p^s\}_{p\in\mathcal{P},s\in\mathcal{S}}$ to two independent groups of experts for each stakeholder, so to mitigate possible bias. In the next section, we describe a refined approach to define an instance of the MSRMP (3.3) that aims to further reduce the level of subjectivity of each stakeholder in defining i_s.

4.2.1.1 Support (non-Expert) Stakeholders in Impact Level Definition

Due to the ad-hoc nature and diversity of privacy breaches, only a qualitative approach can be used [OS14, ENI20]. For instance, it is difficult to evaluate the consequences of a leaked body scan quantitatively. However, defining impact levels is not easy, especially if the stakeholders are not experts in determining the appropriate impact level, which relies explicitly on situational awareness and domain experience. As we stated, since the consequences of privacy breaches are often softer than security breaches, to quantify the impact, we use the term "protection criteria" as protection demands required for privacy (as previously presented in [OS14]) and identify them for each stakeholder. The following steps guide the stakeholders in evaluating the potential impact on their previously identified protection criteria (also known as protection demands) related to a threat that might bring:

- First, for each threat, the stakeholder will be asked "What could be impacted on the protection criteria if the threat happens?". It aims to know the potential damages to the data subject and data the controller can be anticipated.
- Second, assign an aversion level based on the defined baseline (e.g., no, low, medium, high, and catastrophic).

Table 4.5 outlines the strategy used to evaluate the aversion level of each threat on the protection criteria for two stakeholders. This approach is similar to security assessment procedures recommended by the German Federal Office for Information Security (BSI) [OSG+11] and the ISO [ISO17a] for privacy impact assessments. Additional to these approaches, the taxonomy proposed by ENISA [DP16] can be a helpful reference in helping to understand impact levels on individuals (i.e., data subjects), where the lowest level of impact considers minor inconveniences on individuals. In contrast, individuals may encounter significant or even irreversible consequences at the highest level. The outcome of this task will produce

Table 4.5: Stakeholders' protection criteria and impact levels.

What could be impacted on the protection criteria (for each perspective) if the threat happens?					
Data controller		Data subject			
Financial situation	Reputational situation	Health condition	Individual freedom	Social situation	Financial situation
0= No impact.					
1= Low. The impact of any loss or damage is limited and calculable					
2= Medium. The impact of any loss or damage is considerable.					
3= High. The impact of any loss or damage is significant.					
4= Catastrophic. The impact of any loss or damage is devastating.					

the aversion levels for each threat from different perspectives (as reported in Table 4.4) that contribute to evaluating the impact levels (see Formula 4.1).

We later (in Chapter 6) give an in-depth analysis (the analysis is reported in Appendix C.4) of the potential impact association, which contributes to the application of our proposed methodology in a use case scenario.

4.2.2 A Less Subjective Definition of Impact Levels

Our goal is to reduce the level of subjectivity with which i_s is defined. The idea is to refine the definition of i_s given above by introducing a cross-weighting system to reduce bias resulting from stakeholders as much as possible. Besides the availability of a set $\mathcal{P}$ of protection criteria and a family $\{PW_p^s\}_{p\in\mathcal{P},s\in\mathcal{S}}$ of weights associated to a preference p for each stakeholder s, we consider a set $\mathcal{G}$ of protection goals which play a crucial role in identifying appropriate security controls (see, e.g., [ZH11]). Indeed, Confidentiality, Integrity, and Availability are obvious candidates to be included in the set $\mathcal{G}$ (see, e.g., [BBG+17]). However, these are not enough to consider the complex protection requirements deriving from national and international legal provisions such as those concerning data protection contained in the GDPR. For this reason, we assume the set $\mathcal{G}$ to contain the "data protection goals" introduced by the Standard Data protection Model (SDM) [fD17], namely, Confidentiality, Integrity, Availability, Unlinkability and Data minimization, Transparency, and Intervenability (see, Section 2.2.3). We observe that our approach can be applied with other protection goals, we consider those of [fD17] only for the sake of concreteness.

The goal of the approach discussed below is twofold: (i) identify how many goals each threat

Table 4.6: Affected protection goals by each threat and the observation weights in our scenario, G1= Confidentiality, G2= Integrity, G3= Availability, G4= Unlinkability & Data minimization, G5= Transparency, and G6= Intervenability.

Threat	Data Protection Goals						Observation Weights (OW)
	G1	G2	G3	G4	G5	G6	
T_1	✕	-	-	✕	-	-	2/10
T_2	✕	✕	✕	-	-	-	3/10
T_3	✕	-	-	✕	-	-	2/10
T_4	-	✕	✕	-	-	-	2/10
T_5	-	-	-	-	-	✕	1/10

is impacting, and (ii) measure the amplitude of the impact on each goal of a given threat. We start by considering (i). For example, a "*Denial of service*" threat will intuitively have more impact on the data availability goal rather than on the integrity goal; an "*Identity theft*" threat will have more impact on the data confidentiality goal. To keep track of this, we use a *Threat-Protection Goals association* as shown in the first two columns in Table 4.6 where the "✕" ("-") mark in a cell means the goal in the column is affected (not affected, respectively) by the threat in the row (the particular instance of the threat-protection goals association is related to the running example of Section 3.2.1). Intuitively, the more a threat impacts multiple goals, the more it is considered pervasive (e.g., threat T_2 is the most pervasive in Table 4.6 as it affects 3 goals); the more a goal is impacted by multiple threats, the more it is considered scattered (e.g., goal $G1$ is the most scattered in Table 4.6 as it impacts 3 threats). The third column of Table 4.6 shows the so-called Observation Weight and compute as follows:

$$OW_T = \frac{AG_T}{\sum_{T \in \mathcal{T}} AG_T} \tag{4.2}$$

that measures how much a threat T is pervasive for the goals in $\mathcal{G}$, where AG_T is the number of goals in $\mathcal{G}$ affected by a threat $T \in \mathcal{T}$. For example, in Table 4.6, the observation weight OW_{T_1} is 2/10, where $G1$ and $G4$ are the two affected goals by T_1, and the total number of affected goals is 10.

We now consider objective (ii), namely to measure the amplitude of the impact on each goal of a given threat. This is necessary as soon as we realize that the information in Table 4.6 is not enough alone to define i_s because it may be the case that the impact value can be much higher when a goal is impacted severely by a single threat rather than when this is impacted by many threats but only lightly. We do this in two steps. First, we define

the *normalized threat criticality level* as follows:

$$NTC_T = \frac{OW_T \times x_T}{\sum_{T \in \mathcal{T}} (OW_T \times x_T)} \tag{4.3}$$

to quantify the severity of a threat $T \in \mathcal{T}$ (recall that x_T is the impact residue of the threat T after applying the security controls according to a mitigation mapping μ_T). Intuitively, NTC_T is the level of danger of a threat T among all threats in $\mathcal{T}$, or in other words, the relative importance of T with respect to all other threats in $\mathcal{T}$. By having obtained the observation weights (in Table 4.6) and the calculated x_T values (in Table 4.3), the computed normalized threat criticality values for $T \in \mathcal{T}$ are shown in the second column of Table 4.7.

The second step to achieve goal (ii) above is to use the normalized threat criticality level to weight the function i_s defined in Section 4.2.1 when considering a certain protection goal $G \in \mathcal{G}$ for a given stakeholder s so to define the overall impact residue as follows

$$oir(s) = \sum_{G \in \mathcal{G}} \left(\frac{\sum_{T \in \mathcal{T}} \xi_{T,G} \times NTC_T \times i_s(T)}{\#(\mathcal{T}, G)} \right) \tag{4.4}$$

where $\xi_{T,G}$ is 1 when the threat $T \in \mathcal{T}$ compromises the goal $G \in \mathcal{G}$ and 0 otherwise; $\#(\mathcal{T}, G)$ is the number of threats in $\mathcal{T}$ that have an impact on the goal G (this means that $\#(\mathcal{T}, G) = \sum_{T \in \mathcal{T}} \xi_{T,G}$). Observe that the expression between parentheses in (4.4) can be seen as the average impact on a given goal G with respect to the threats in $\mathcal{T}$ that are relevant to G. For instance, according to Table 4.6, the *intervenability* goal ($G6$) is affected only by T_5 which means that $\#(\mathcal{T}, G6)$ is 1. According to Table 4.7, the average impact of the *confidentiality* goal ($G1$) for the data subject is 0.074, while the same value for the data controller is 0.087. Finally, observe that since the *transparency* goal ($G5$) is not affected by anyone of the threats (according to Table 4.6), it is not mentioned in Table 4.7 neither used for calculating the overall impact residue. By aggregating the impact average of protection goals, the overall impact residue from the data subject's point of view is $oir(DS) = 0.549$, and for the data controller is $oir(DC) = 0.576$.

At this point, we are in the position to define instances of the MSRMP statement (3.3) by using (4.4) as the definition of the overall impact residue rather than those proposed in Section 3.2.2. We also observe that by substituting the definition (4.3) to NTC_T in the expression of $oir(s)$, it is easy to see that we can derive a MSRMP similar to (3.4), i.e. considering x_T as variables rather than μ_T for $T \in \mathcal{T}$, for which it is possible to apply the same technique discussed at the end of Section 3.2.2 that allows us to solve an optimization problem over a smaller search space and then derive optimal solutions for the original problem.

4.3 Discussion

In the first attempt above, we evaluated the impact levels only on the basis of the assumption that all threats have an equal importance level (apart from the mitigation controls that state their presence). In addition to this assumption, we observed the fact that the aversion level association is a subjective task. Therefore, assessing overall risk levels in this way may become too subjective if only impact levels (i.e., the level of aversion in Table 4.4) are going to be considered. Thus, by defining one further association for threats, we are able to evaluate better and then prioritize them, as we proposed in the second attempt. As a consequence of this association (i.e., threat-protection goal association in Table 4.6), the subjectivity level of the evaluation could be fairly reduced. To understand these associations better, consider the following instances to illustrate the distinction between these two:

(i) where according to Table 4.6, threat T_5 has impacted on only one protection goal (i.e., $G6$). However, according to Table 4.4, T_5 is the most significant threat from the data subject point of view with a high aversion level.

(ii) where according to Table 4.4, threat T_3 has the high aversion level to compare with other threats from the data controller point of view, while the aversion level for the data subject is even less than T_2 and T_5. However, according to Table 4.6, threat T_3 is the most dangerous threat by affecting three protection goals. Whereas in this table, T_1, T_3, and T_4 are impacting on two protection goals.

Therefore, combining these two associations helps us have a less subjective impact assessment.

Table 4.7: Threat criticality and impact level values, together with the computed protection goals' impacts for each threat for the data subject (DS) and the data controller (DC).

Threats (T)	Normalized Threat Criticality (NTC)	Normalized Impact Level $i_s(T)$		Protection Goals' Impacts									
				G1		G2		G3		G4		G6	
		DS	DC	DS	DC	DS	DC	DS	DC	DS	DC	DS	DC
T_1	0.17	0.075	0.4	0.013	0.068	0.000	0.000	0.000	0.000	0.013	0.068	0.000	0.000
T_2	0.22	0.725	0.5	0.163	0.112	0.163	0.112	0.163	0.112	0.000	0.000	0.000	0.000
T_3	0.11	0.45	0.75	0.048	0.080	0.000	0.000	0.000	0.000	0.048	0.080	0.000	0.000
T_4	0.36	0.45	0.65	0.000	0.000	0.160	0.231	0.160	0.231	0.000	0.000	0.000	0.000
T_5	0.14	0.85	0.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.121	0.071
Average impact	-	-	-	0.074	0.087	0.161	0.172	0.161	0.172	0.030	0.074	0.121	0.071

Chapter 5

Implementation and Experimental Evaluation

The previous chapters' contributions dealt with performing a cyber-risk assessment in multi-stakeholder scenarios. As a quick recap; in Chapter 3, we initially introduced the Multi-Stakeholder Risk Minimization Problem (MSRMP) to assist in the definition of the best (for all the stakeholders involved in the system) Risk Management Policies (RMPs), and then formalized the MSRMP as a multi-objective optimization problem that can be solved by using state-of-the-art techniques for Pareto Optimality. In Chapter 4, we proposed a semi-automated methodology to define and solve instances of the MSRMP. As depicted in Figure 5.1, we validate the applicability of the proposed methodology in two

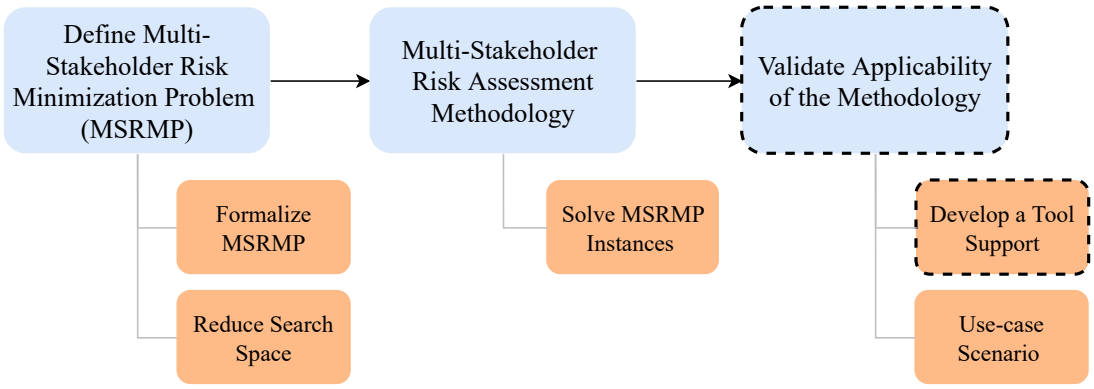

Figure 5.1: The highlighted part (dashed lines) illustrates the contribution of Chapter 5 in accordance with the contribution flow outlined in Section 1.2.

separated contributions namely "Develop a tool support", and "Use-case Scenario". This chapter examines the first contribution, where we demonstrate the applicability of the proposed methodology by developing a tool for defining an instance of the MSRMP in Section 5.1. Furthermore, in Section 5.2, we conduct several tests to evaluate the practicality of the methodology experimentally.

5.1 Tool Support

To validate the applicability of the proposed methodology, we have implemented a tool (in Java) able to assist in defining an instance of the MSRMP as discussed in Section 4.2 and performed two sets of tests in order to experimentally evaluate the practicality of our approach[1].

The goal of the tool is two-fold, namely (i) assisting in the definition of an instance of the MSRMP and (ii) automatically solving the resulting instance. The architecture of the tool is illustrated in Figure 5.2. The tool operates in two phases (see outer boxes in the figure) and assumes the availability of the sets of stakeholders $\mathcal{S}$, threats $\mathcal{T}$, security controls $\mathcal{C}$, protection criteria $\mathcal{P}$ together with their weights $\{PW_p^s\}_{p\in\mathcal{P},s\in\mathcal{S}}$, and goals $\mathcal{G}$; the first three are discussed in Section 3.2.2, the fourth in Section 4.2.1, and the last in Section 4.2.2. The architecture also reports how tabular definitions of the various entities can be given; for instance, the set $\mathcal{T}$ of threats can be defined as in Table 4.2 and the set $\mathcal{P}$ of protection goals together with their weights $\{PW_p^s\}_{p\in\mathcal{P},s\in\mathcal{S}}$ as in Table 4.4. We assume that these inputs are derived from the application of available and well-known techniques for risk assessment, as already discussed above; our approach is agnostic with respect to the particular methodology used. The tables specifying the inputs above are encoded in JSON format.

The first phase is semi-automated and can be seen as a preparatory to the definition of an instance of the MSRMP. More precisely, it defines the association between controls and threats $\{\mathcal{C}_T\}_{T\in\mathcal{T}}$ (see Section 3.2.2), the aversion level mapping al_p^s for each protection criteria $p \in \mathcal{P}$ and stakeholder $s \in \mathcal{S}$ (see Table 4.4 in Section 4.2.1), and the observation weight OW_T for each threat $T \in \mathcal{T}$ (see last column of Table 4.6 whose value is derived according to the expression (4.2)). The first two outputs of this phase are obtained with human intervention as the user needs to identify which security controls are effective for each threat and which is the level of aversion of each stakeholder for a given protection criteria to be violated, whereas the last one is automatically derived after the user has specified which goals are affected by each threat. Also, the outputs of the phase are represented in JSON format.

[1]The code of the tool and the material to replicate the experiments are available at `https://github.com/stfbk/MSRMP`

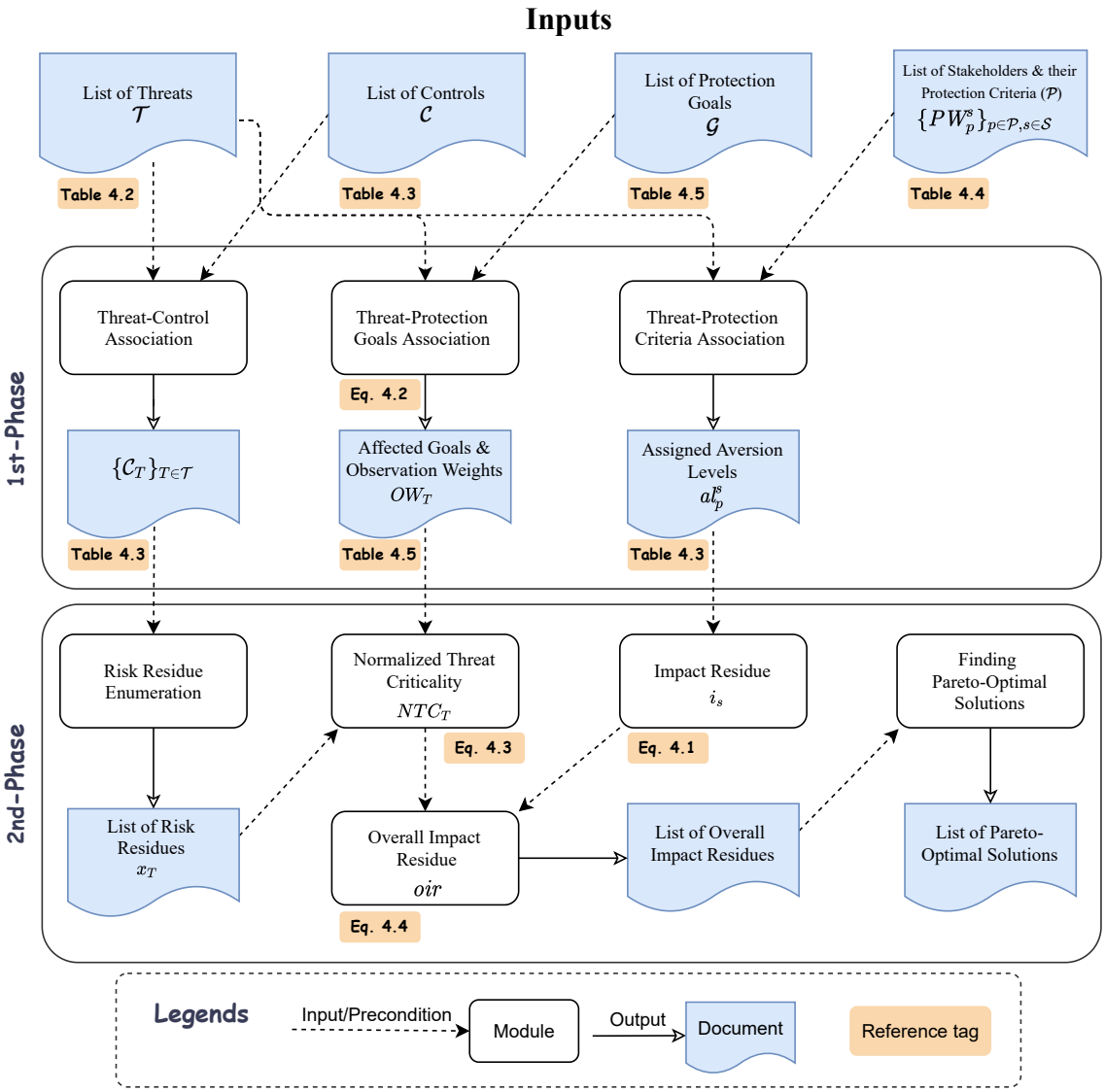

Figure 5.2: Architecture of the implemented tool.

The second phase is fully automated and aims to define and solve an instance of the MSRMP. This requires to use the outputs of the first phase to define the impact level mapping i_s for each stakeholder $s \in \mathcal{S}$ (see Section 3.2.2) along the lines of Section 4.2.1 and then the overall impact residue oir as discussed in Section 4.2.2. At this point, the tool has fully defined an instance of the MSRMP (3.3) and it is left with the task of solving it. For this, it needs to enumerate all risk residues x_T for each threat $T \in \mathcal{T}$ by using the approach in Section 4.2 for defining Table 4.3 together with deriving the Normalized Threat Criticality values for the various threats and then adapting the strategy discussed

at the end of Section 3.2.2 to identify the mitigation mappings that are Pareto optimal.

We observe that there are multiple possible strategies to combine the numeration of risk residues and the identification of Pareto optimal values ranging. For instance, one can first compute the entire set of feasible solutions and only after look for Pareto optimal ones or one can imagine interleaving the two activities by computing the Pareto optimal values in different subsets of the whole set of feasible solutions and then select those solutions that are Pareto optimal for the entire search space. Below, we first discuss the computational behavior of the second phase on the running example in Section 3.2.1, and we design two sets of tests to understand which is the most promising strategy to identify the set of Pareto Optimal risk residues or, equivalently, mitigation mappings.

5.1.1 Applying the Prototype Tool on the Running Example

In this section, we discuss the results of applying the second phase of our methodology, as implemented in the prototype tool, on the running example of Section 3.2.1. First, the tool computes the whole set of possible solutions whose cardinality is $57,600$; this is as expected from the formula $\Pi_{T\in\mathcal{T}}|X_T| = |X_{T_1}| \times |X_{T_2}| \times |X_{T_3}| \times |X_{T_4}| \times |X_{T_5}| = 10 \times 20 \times 8 \times 6 \times 6 = 57,600$ presented in Section 3.2.2 (see Example 6).

This takes around 2.1 seconds on a machine with 16 GB of RAM and a 1.90 GHz CPU. Each solution is a pair containing the risk residue values for the Data Subject (DS) and the Data Controller (DC). Figure 5.3 shows the set of possible solutions plotted on a Cartesian plane, whose x-axis shows the risk residue of DS and the y-axis that of DC. By looking at the figure, it is immediate to see that the optimal solution is that on the bottom left—whose risk residue values are 0.2260 for DS and 0.4168 for DC—as it dominates all other solutions (see the dominance definition described in Section 2.3.2). The tool takes around 2.2 seconds to identify this point as the best one.

After identifying the risk residue levels, one is left with the problem of computing the set of RMPs that generate such values. A method to do this has been illustrated at the end of Section 3.2 and implemented in the tool that takes less than 3 seconds to identify the following tuple

$$\langle x^*_{T_1}, x^*_{T_2}, x^*_{T_3}, x^*_{T_4}, x^*_{T_5}\rangle = \langle 1, 0.05, 0.125, 0.16, 0.16\rangle$$

corresponding to $(0.2260, 0.4168)$ and then to identify all the RMPs associated to the above tuple of x_T values for $T \in \{T_1, T_2, T_3, T_4, T_5\}$. By recalling (3.5) and that $\mathcal{A} = \{0, 0.5, 1\}$, it is not difficult to see that there are $360 = 1 \times 10 \times 4 \times 3 \times 3$ distinct RMPs associated to the tuple of x_T values above since

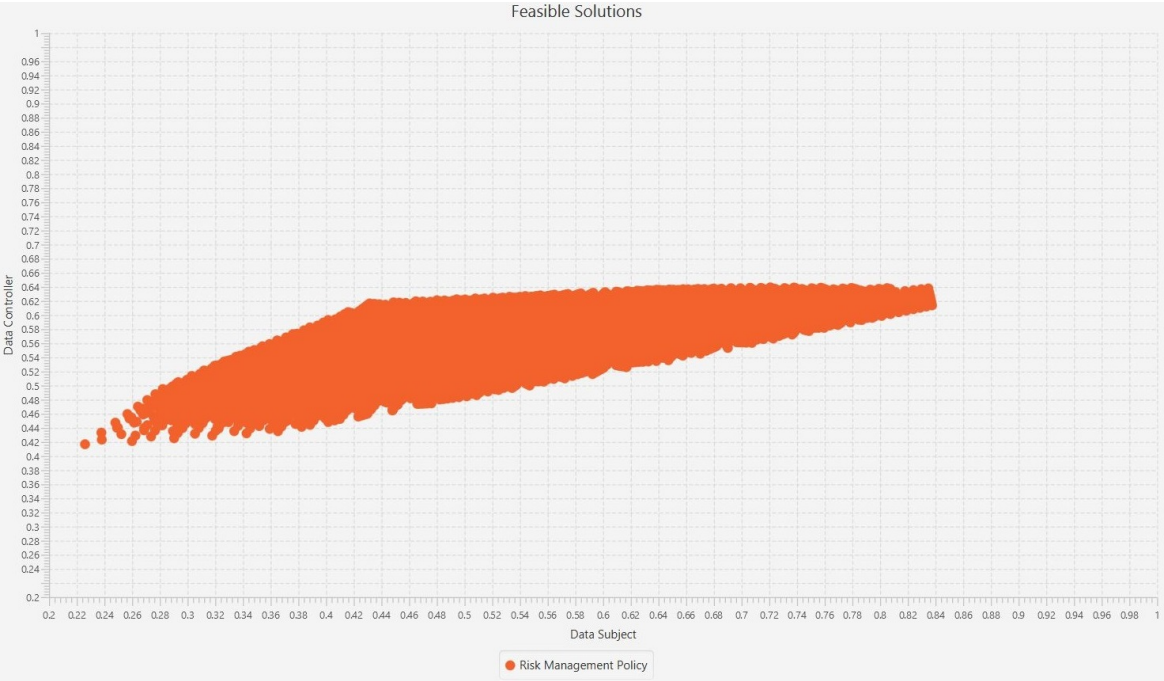

Figure 5.3: All feasible solutions (i.e., the search space) in our scenario.

- there is just one mitigation mapping satisfying

$$\frac{\Sigma_{c\in\{c_1,\ldots,c_5\}}\mu_{T_1}(c)}{5} = 1 - x^*_{T_1} = 0$$

  as the values in $\mathcal{A}$ are non-negative values;

- there are 10 mitigation mappings satisfying

$$\frac{\Sigma_{c\in\{c_6,\ldots,c_{15}\}}\mu_{T_2}(c)}{10} = 1 - x^*_{T_2} = 0.95$$

  as the only way to get 9.5 by adding 10 values from $\mathcal{A}$ is to have nine of them equal to 1 and the remaining one to 0.5;

- there are 4 mitigation mappings satisfying

$$\frac{\Sigma_{c\in\{c_{16},\ldots,c_{19}\}}\mu_{T_3}(c)}{4} = 1 - x^*_{T_3} = 0.875$$

  as the only way to get $3.5 = 4 \times 0.875$ by adding 3 values from $\mathcal{A}$ is to have three of them equal to 1 and the remaining one to 0.5;

- there are 3 mitigation mappings satisfying

$$\frac{\Sigma_{c\in\{c_{20},\ldots,c_{22}\}}\mu_{T_4}(c)}{3} = \frac{\Sigma_{c\in\{c_{23},\ldots,c_{25}\}}\mu_{T_5}(c)}{3} = 1 - x^*_{T_4} = 1 - x^*_{T_5} = 0.84$$

  as the only way to get $3.5 = 3 \times 0.8$ by adding three values from $\mathcal{A}$ is to have two of them equal to 1 and the remaining one to 0.5.

The tool mechanizes the observations above and computes the set of security controls associated to the Pareto optimal solutions by solving a variant of the Sum Subset Problem (SSP) in which multisets are considered instead of sets as explained at the end of Section

Table 5.1: Examples of mitigation mappings associated to the optimal solution in Figure 5.3

Threats ($\mathcal{T}$)	Controls $\{\mathcal{C}_T\}_{T\in\{T_1,T_2,T_3,T_4,T_5\}}$	Possible Mitigation Combinations	x^*_T
T_1	c_1) Purpose specification	○ ○ ○	1
	c_2) Ensuring limited data processing	○ ○ ○	
	c_3) Ensuring purpose related processing	○ ○ ○	
	c_4) Ensuring data minimization	○ ○ ○	
	c_5) Enabling data deletion	○ ○ ○	
T_2	c_6) Ensuring data subject authentication	◐ ● ●	0.05
	c_7) Ensuring staff authentication	● ◐ ●	
	c_8) Ensuring device authentication	● ● ◐	
	c_9) Logging access to personal data	● ● ●	
	c_{10}) Performing regular privacy audits	● ● ●	
	c_{11}) Ensuring data anonymization	● ● ●	
	c_{12}) Providing confidential communication	● ● ●	
	c_{13}) Providing usable access control	● ● ●	
	c_{14}) Ensuring secure storage	● ● ●	
	c_{15}) Ensuring physical security	● ● ●	
T_3	c_{16}) Providing confidential communication	◐ ● ●	0.125
	c_{17}) Logging access to personal data	● ◐ ●	
	c_{18}) Ensuring data subject authentication	● ● ◐	
	c_{19}) Ensuring data anonymization	● ● ●	
T_4	c_{20}) Enabling offline authentication	◐ ● ●	0.16
	c_{21}) Network monitoring	● ◐ ●	
	c_{22}) Prevention mechanisms for DoS attacks like firewalls, etc.	● ● ◐	
T_5	c_{23}) Informing data subjects about data processing	◐ ● ●	0.16
	c_{24}) Handling data subject's change requests	● ◐ ●	
	c_{25}) Providing data export functionality	● ● ◐	

3.2.2. Indeed, this is so because a mitigation mapping μ_T associates a control of $\mathcal{C}_T$ with a value in $\mathcal{A} = \{0, 0.5, 1\}$ for each $T \in \mathcal{T}$ and nothing prevents two or more controls to be mapped to the same value in $\mathcal{A}$. Since all solutions to the SSP should be identified to be able to enumerate all possible mitigation mappings, the algorithm is exponential in the number of security controls associated to each threat, i.e. in the cardinality of $\mathcal{C}_T$ for $T \in \mathcal{T}$. Since such a number is typically low (on average around 5 and at most 10 in our experience), the time consumption is quite reasonable in practice, being around half a second at most for a single threat $T \in \mathcal{T}$. To conclude the discussion, the last column of Table 5.1 reports three mitigation mappings associated to the optimal solution considered above. Notice that all mitigation mappings associated to the optimal solution above suggest avoiding implementing any security control for threat T_1. This is a consequence of the impact defined in Table 4.4 that makes T_1 relevant only for the social situation of the DS, while it is negligible for all other aspects. Given this remark, one may decide to modify the values to increase the impact of T_1 for the DS and then re-run the analysis. This is a clear advantage of having a high level of mechanization of our methodology.

Indeed, the running example is simple and poses no challenges to our prototype implementation. To understand the scalability of the proposed approach, we have designed a set of synthetic optimization problems whose sets of potential solutions is increasingly large and then experiment with two different strategies to generate and visit such a set in the process of identifying Pareto optimal solutions. This is reported in Sections 5.2 below.

Preliminarily, we discuss a variant of the MSRMP that, with little effort, can be solved by a minor modification to our approach. Such a variant is a constrained version of the MSRMP whereby it is possible to identify lower bounds for risk residue levels of the DS and DC, i.e. the stakeholders may be willing to accept a risk residue above a certain threshold according to their risk appetite, i.e. the amount of risk the stakeholder is willing to take in pursuit of objectives it considers valuable. In other words, the set of possible solutions is reduced to consider those that are above certain values for the DC and the DS. To illustrate, we consider the situation in which such lower bounds are set to 0.45 and 0.55 for the DS and the DC, respectively. In this case, the prototype tool is able to identify a set of 6 Pareto optimal solutions (in Table 5.2, below) by taking around 1.4 seconds and then consumes around 5 milliseconds to compute the associated values $x_{T_1}, ..., x_{T_5}$ as reported in Table 5.3. In Table 5.3, the retrieved x_T values are listed for the Pareto optimal solutions of Table 5.2. For example, in this table, the first row (i.e., $\mathbb{RS}_1$) reports the x_T values for $\mathbb{S}_1$ where $oir(DS)$ and $oir(DC)$ are 0.4514 and 0.5518, respectively.

Finally, at this point, we are left with the problem of configuring the security controls associated to each threat $T \in \{T_1, ..., T_5\}$ to obtain the required value x_T. The tool computes the set of security controls associated to the 6 Pareto optimal solutions in around a second by solving (a variant of) the SSP as explained above for the single Pareto optimal solution.

As we mentioned earlier, the tool works with JSON files both input and output. For example, Figure 5.4 shows an example of an input artifact for the tool in JSON document. As an example of the output, Figure 5.5 shows an excerpt JSON document generated by the tool, where it lists the overall risk residues for all possible solutions of the running example scenario.

Table 5.2: Pareto's solutions under the defined risk exposure boundary.

Pareto Solutions	$oir(DS)$	$oir(DC)$
$\mathbb{S}_1$	0.4514	0.5518
$\mathbb{S}_2$	0.4503	0.5536
$\mathbb{S}_3$	0.4504	0.5530
$\mathbb{S}_4$	0.4520	0.5501
$\mathbb{S}_5$	0.4516	0.5502
$\mathbb{S}_6$	0.4505	0.5525

Table 5.3: Retrieved residual risk values for the identified Pareto solutions.

	$x^*_{T_1}$	$x^*_{T_2}$	$x^*_{T_3}$	$x^*_{T_4}$	$x^*_{T_5}$
$\mathbb{RS}_1$	0.9	0.4	0.75	1	0.25
$\mathbb{RS}_2$	0.8	0.4	0.875	0.66	0.33
$\mathbb{RS}_3$	0.6	0.3	0.875	0.33	0.33
$\mathbb{RS}_4$	0.6	0.25	0.75	0.25	0.25
$\mathbb{RS}_5$	0.6	0.2	0.25	1	0.25
$\mathbb{RS}_6$	0.5	0.25	0.875	0.16	0.33

```
{
  "Stakeholders": [
    {
      "name": "data subject"
    },
    {
      "name": "data controller"
    }
  ],
  "Threats": [
    {
      "name": "T1",
      "malicious action": "Personal data is kept stored more than the time necessary for
        the purposes by ACME.",
      "controls": [
        {
          "control_name": "Purpose specification",
          "implementation_status": 1.0
        },
        {
          "control_name": "Ensuring limited data processing",
          "implementation_status": 1.0
        },
        {
          "control_name": "Ensuring purpose related processing",
          "implementation_status": 0.5
        },
        {
          "control_name": "Ensuring data minimization",
          "implementation_status": 0.5
        },
        {
          "control_name": "Enabling data deletion",
          "implementation_status": 0.0
        }
      ]
    },
    {...},
    {...},
    {...},
    {
      "name": "T5",
      "malicious action": "ACME does not implement a procedure (technical and/or processes
        ) that allows the patients to rectify, erase or block individual data.",
      "controls": [
        {
          "control_name": "Informing data subjects about data processing",
          "implementation_status": 0.5
        },
        {
          "control_name": "Handling data subjects change requests",
          "implementation_status": 0.5
        },
        {
          "control_name": "Providing data export functionality",
          "implementation_status": 0.0
        }
      ]
    }
  ]
}
```


Figure 5.4: An example of an input (in ACME scenario) for the tool in the format of JSON document.

Figure 5.5: An excerpt of JSON document generated by the tool, listing all overall risk residues.

5.2 Experimental Results

This section undertakes some experimental evaluations to examine the scalability of proposed methodology through the implemented tool. Hence, we present two test cases to assess the computational time and resources in the following. Since the instances of the variant of the SSP required to enumerate all possible mitigation mappings corresponding to each Pareto optimal solution of the form $\langle x_T \rangle_{T \in \mathcal{T}}$ are typically small, their solution does not consume a relevant amount of resources (both time and memory) and thus we disregard this activity in the discussion below.

5.2.1 Test 1: Upfront Computation of Feasible Solutions

The goal of the first set of tests is to evaluate the strategy of computing the set of feasible solutions upfront and then identify those that are Pareto optimal. The idea is to understand the time and memory occupation required to do this while increasing the number of threats and the number of security controls per threat. We consider two stakeholders (i.e. $|\mathcal{S}| = 2$), the protection criteria $\mathcal{P}$ are the same as those in Table 4.4, the number of protection goals are 6 as those introduced in Section 4.2.2, an increasing number $|\mathcal{T}| = 5, 6, 7, 8$ of

threats, and a number $q = 4, 5$ of security control associated with each threat so that $|\mathcal{C}_T| = q * 5, q * 6, q * 7, q * 8$. For each one of these configurations, we measure the time (in seconds) and the memory occupation (in GB of heap) taken to compute the entire set of feasible solutions when running our prototype on a cluster with a CPU of 3.2 GHz and 500 GB of RAM. We do not include the time to identify the Pareto Optimal solutions as the resource consumption for computing the feasible set of solutions (see the last two columns of Table 5.4) clearly shows the exponential behavior for both computation time and memory occupation despite the dramatic reduction in the search space (consider the values in the column Reduction Factor) obtained by using the approach of solving with respect to risk residues in place of mitigation mappings discussed at the end of Section 3.2.2.

5.2.2 Test 2: Interleaving the Computation of Feasible and Optimal Solutions

The first test set clearly shows that the upfront computation of the whole set of feasible solutions does not scale. For this reason, we designed a different approach whereby the two activities are interleaved by computing non-overlapping sub-sets of the feasible solutions and then identify those that are Pareto Optimal. As already observed, this can be done in different ways, and we propose two strategies, both parameterized by the size d of the sub-set of feasible solutions that are being considered.

- In the first strategy, we collect the Pareto Optimal solutions identified in each sub-set with cardinality d of the set of feasible solutions in a list ℓ and once the entire set of

Table 5.4: Experimental results of Test 1. Legend: Reduction Factor, Computation Time is in Seconds (S), and the maximum Heap Size is in Gigabyte (GB).

$\lvert\mathcal{T}\rvert$	$\lvert\mathcal{C}_T\rvert$	Solution Set Size		Reduction Factor	Computation Time (S)	Heap Size (GB)
		$\Pi_{T\in\mathcal{T}}(k^{\lvert\mathcal{C}_T\rvert}-1)$	$\Pi_{T\in\mathcal{T}}\lvert X_T\rvert$			
5	20	$32,768 \cdot 10^5$	32,768	10^5	0.312	∼0.25
6	24	$262,144 \cdot 10^6$	262,144	10^6	1.2	∼1.5
7	28	$2,097,152 \cdot 10^7$	2,097,152	10^7	9.7	∼12
8	32	$16,777,216 \cdot 10^8$	16,777,216	10^8	237	∼29
5	25	$\sim 8.29 \cdot 10^{11}$	100,000	$\sim 8.29 \cdot 10^6$	0.626	∼0.5
6	30	$\sim 2.01 \cdot 10^{14}$	1,000,000	$\sim 2.01 \cdot 10^8$	3.7	∼9
7	35	$\sim 4.86 \cdot 10^{16}$	10,000,000	$\sim 4.86 \cdot 10^9$	105	∼28
8	40	$\sim 1.17 \cdot 10^{19}$	100,000,000	$\sim 1.17 \cdot 10^{11}$	2,787	∼416

```
/**
Assumptions:
1) A Solution object constituted of two double values (x,y) where represent the value of
    "oir" each stakeholder.
2) The size of feasible set is computed.
**/
List<Solution> final_Solutions = new ArrayList<>(); // Define a final list of solutions
List<Solution> temp = new ArrayList<>(); // Define a temporary list
int feasible_Set_Size; // Compute the size of feasible set
int counter=0; // Define a variable for counting the number of generated solutions
int d; // Define a variable to generate solution list for each iteration
boolean check = true;
while(check)
   {
      //Generate a list of solutions with size of d and adding to the temp list
      temp.addAll(final_Solutions); // Adding also all discovered Pareto Solutions up to
          now to temp
      final_Solutions.clear(); // Clearing final_Solutions to add Pareto Solutions based on
          the new iteration
           for (Solution a : temp) { // Apply Pareto-Optimality definition on the temp list
               boolean flag = false;
               for (Solution b : temp) {
                   if (a != b) {
                       if (a.getX() >= b.getX() && a.getY() >= b.getY()) {
                           flag = true;
                       }
                   }
                   if (flag)
                       break;
               }
               if (!flag)
                   final_Solutions.add(a); // Adding a solution as Pareto solution
           }
           counter+= d;
           temp.clear();
           if (counter >= feasible_Set_Size) { // Checking size of the generated solutions
               check = false;
           }
    }
return final_Solutions; // Return the final Pareto solution list
```

Figure 5.6: Pseudocode of the second strategy of finding Pareto optimal solutions.

feasible solutions has been covered, the list ℓ is processed to extract the final set of Pareto Optimal solutions.

- The second strategy is similar to the previous one except for the fact that the content of the list ℓ of Pareto Optimal solutions for a given sub-set of the set of feasible solutions is added to the next sub-set of feasible solutions to be considered so that, when considering the last sub-set, we identify the final set of Pareto Optimal solutions.

The pseudocode of the second strategy (i.e., in the syntax of Java) is shown in Figure 5.6. To study the scalability in terms of resource consumption of these two strategies, we define a second test set with the same parameters of the previous one except for $|\mathcal{T}| = 6, 7, 8, 9$

and the number of security controls associated to each threat is 4. We consider increasing values of $d = 8^h$ for $h = 1, 2, 3, 4, 5, 6$ to understand how the cardinality of the sub-set of the feasible solutions affect performances. As for the previous test set, we measure the timing (in seconds) and the heap occupation (in MB) with a time-out (T/O) of 3 hours. As the results (obtained on a personal computer with a CPU of 1.90 GHz and 16 GB of RAM) in Table 5.5 shows, the scalability is much improved with respect to the results of the first test above, regardless of the strategy adopted to identify the Pareto Optimal solutions. It is worth noticing that for this test set we consider a less powerful computer, and we include the effort of identifying the Pareto Optimal solutions. Although there is no clear winner between the two strategies described above, a closer analysis of the results in Table 5.5 shows that the second strategy is better than the first one in most cases and in particular for larger instances of the MSRMP; for instance, consider the test case with 8 threats and $d = 8$, the computation time and the maximum heap space used for the first strategy are $2,098.4$ seconds and $1,282$ MB, whereas for the second strategy are 68.7 seconds and 256 MB. We observe that setting an appropriate value for the parameter d (neither too small nor too large) seems to be important for the timing behavior of the first strategy while the second strategy seems to be much less independent; unsurprisingly, for the memory occupation, larger values of d corresponds to larger heap sizes but much less than those of the first test set (notice that the numbers in Table 5.4 are in GB whereas those in Table 5.5 are in MB).

Table 5.5: Experimental results based on the two defined strategies of Test 2.

Test Case	$\lvert\mathcal{T}\rvert$	$\lvert\mathcal{C}_{\mathcal{T}}\rvert$	Computation Time (Second) and RAM Heap Size (Megabyte)					
			d=8	d=64	d=512	d=4,096	d=32,768	d=262,144
Strategy 1	6	24	4.7(S) , 308(MB)	2.5(S), 256(MB)	2.7(S), 256(MB)	3.7(S), 320(MB)	11.3(S), 499(MB)	6.8(S), 2,422(MB)
	7	28	95.5(S), 986(MB)	8.1(S), 382(MB)	8.7(S), 256(MB)	10.7(S), 459(MB)	15.8(S), 900(MB)	357(S), 2,509(MB)
	8	32	2,098.4(S), 1,282(MB)	71(S), 308(MB)	52(S), 308(MB)	62.5(S), 497(MB)	159.5(S), 1,004(MB)	317.5(S), 3,500(MB)
	9	36	T/O	7,346.7(S), 533(MB)	541.3(S), 308(MB)	560(S), 522(MB)	575.9(S), 1575(MB)	5,124.9(S), 3,812(MB)
Strategy 2	6	24	2.5(S), 256(MB)	4.5(S), 256(MB)	2.7(S), 256(MB)	3.9(S), 308(MB)	5.6(S), 826(MB)	7.2(S), 2,405(MB)
	7	28	10.7(S), 256(MB)	10.7(S), 256(MB)	12.7(S), 256(MB)	8.9(S), 525(MB)	10.9(S), 1,037(MB)	60.9(S), 2,471(MB)
	8	32	68.7(S), 256(MB)	83.3(S), 256(MB)	70.9(S), 256(MB)	58.3(S), 256(MB)	64.7(S), 1,186(MB)	108.9(S), 4,066(MB)
	9	36	567.3(S), 256(MB)	557.2(S), 308(MB)	507.6(S), 308(MB)	553.7(S), 408(MB)	555.8(S), 1,513(MB)	934(S), 4,066(MB)

5.2.3 Discussion on Experiments

There are two main lessons learned from the experiments discussed above.

- First, the transformation of the original MSRMP (3.3) over $\langle \mu_T \rangle_{T \in \mathcal{T}}$ into the one (3.4) over the $\langle x_T \rangle_{T \in \mathcal{T}}$ allows for a substantial reduction of the search space. To see this, consider the Reduction Factor in Table 5.4.
- Second, considering the family $\mathcal{C}_{T \in \mathcal{T}}$ of controls associated to each threat $T \in \mathcal{T}$ is crucial, in practice, to reduce the search space of the problem of transforming back a solution $\langle x_T^* \rangle_{T \in \mathcal{T}}$ of (3.4) into the set $\{\langle \mu_T \rangle_{T \in \mathcal{T}}\}$ of associated mitigation mappings of the original MSRMP (3.3). This is so because the cardinality of $\mathcal{C}_T$ is usually low for each $T \in \mathcal{T}$ so that, despite the exponential complexity as discussed at the end of Section 3.2.2, the time and memory consumption are reasonable in practice.

Chapter 6

Application in the Trace4Safe Project

We began validating our proposed methodology in the previous chapter, where we demonstrated its applicability by developing a tool for defining and solving an instance of the MSRMP in Section 5.1, and then we conducted several experiments (in Section 5.2). In this chapter, we apply our methodology in the Trace4Safe project, whose activities were included in the context of a European initiative to develop technological solutions to help handle the COVID pandemics.

Due to the global out-breaking of COVID-19, applying contact tracing solutions is getting the attention of several countries such as China, Italy, Singapore, Germany, etc., and some big IT enterprises like Microsoft, Google, and Apple. Contact tracing systems work based on individuals' location and data is collected automatically from personal devices to

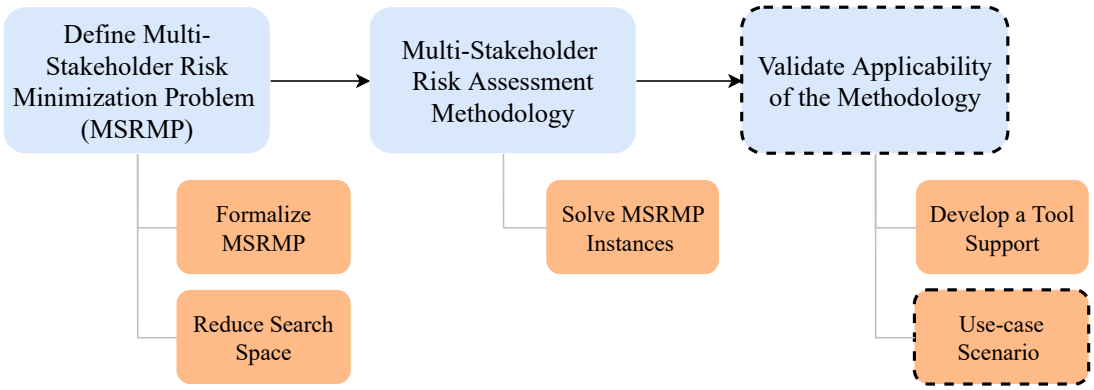

Figure 6.1: The highlighted part (dashed lines) illustrates the contribution of Chapter 6 in accordance with the contribution flow outlined in Section 1.2.

identify and, in the case of being in contact with anyone who tests positive for COVID-19, inform people by sending a notification such as a buzzer, vibration, email, SMS, etc. The "Trace4Safe" project is a solution to monitor and control social distancing in the case of COVID-19 by using contact tracing systems. Trace4Safe proposal was selected among several that were submitted to the EIT[1] call to develop a technological solution for better handling COVID-19 pandemics. In this project, the main objective of the Security & Trust Research Unit of the Center for Cybersecurity in FBK[2] was to assess the security and privacy of the designed solutions and provide suggestions to integrate mitigations to security threats and privacy-enhancing technologies. We performed our assessment in the framework of the Data Protection Impact Assessment (DPIA) of the GDPR and focused on consent management together with data retention and sharing as the cornerstones of transparency and user trust.

In Section 6.1, we introduce the Trace4Safe project and then briefly overview the contact tracing solution offered by Trace4Safe, the anticipated product, and the objective services. We helped to conduct a DPIA within the Trace4Safe as described in Section 6.2, where for that, we initially investigated general security and privacy requirements in contact tracing systems that were then refined to the Trace4Safe project (in Section 6.2.2). In Section 6.3, we deep dive into the processes and procedures to comprehend how the data is acquired and transferred inside the system. Lastly, in Section 6.4, we evaluate risks within Trace4Safe in the context of the multi-stakeholder risk assessment by applying our methodology.

6.1 "Trace4Safe" a Hybrid Contact Tracing Monitoring and Detection for a Safe Workplace

The sanitary and socio-economic crisis produced by the COVID-19 outbreak has significantly influenced many different aspects of our lives, changing both personal and business interests. Workplaces, in particular (industrial sites, offices, stores, and so on), must ensure that a one-meter physical separation and rigorous adherence to hygiene requirements are adhered to in order to protect the health of employees and clients alike. Production lines at manufacturing enterprises were among the first activities permitted to recommence operations. Since the beginning of the lockdown, they have been actively exploring new solutions to put in place in order to be prepared for a safe reopening.

Trace4Safe, as an EIT activity, focused on the design, implementation, and validation of a token-based contact tracing solution that helps businesses in facing the challenge posed by the spread of the COVID-19 virus by letting them trace contacts within their premises

[1]The European Institute of Innovation and Technology (EIT), `https://eit.europa.eu/`

[2]Fondazione Bruno Kessler (FBK), `https://www.fbk.eu/`

and preventively analyze risks, thus providing the necessary information to best manage a potential outbreak. The activity was participated by different groups namely Thinkinside[3], the University of Helsinki[4], Fondazione Bruno Kessler[5] (FBK), and Telefonica[6].

This project aims to design, implement, and market a physical token-based contact tracing solution specifically tailored to large complex industrial sites and office spaces. In the end, the application supports:

- monitoring and enforcing social (physical) distancing in real-time;
- evaluating the risks associated to a COVID-19 outbreak, identifying the areas to intervene in order to mitigate risks;
- detecting and managing possible COVID-19 outbreaks in workplaces, back tracing and isolating contacts, and confining the impact on production and site operations.

The enabling infrastructure for detecting and tracing contacts is a combination of the following systems:

- a Real Time Locating System (RTLS) which detects the location of every single worker in the monitored area. Location information is sent to a remote server where a dedicated service is able to detect when physical distancing is not respected, and (i) provide a feedback (e.g., vibration or buzzer) to the involved parties, and (ii) register the contact for a later processing;
- a P2P contact tracing system whereby, the tokens, without the need of any external infrastructure, are able to detect the proximity and to register the contact. Contacts are then offloaded to a remote server when employees are in the proximity of a dedicated gateway.

The solution is designed and developed to fully address the many aspects involved, including the privacy of users (GDPR compliance) and epidemiological effectiveness (identification of contacts at risk and simulation of isolation/quarantine strategies of contacts) with the ultimate goal to become an additional Individual Protection System to ensure a safe workplace.

[3] `Https://thinkin.io/`, with the role of the coordination of the project, including technical activities.

[4] The contribution of the University of Helsinki has been on the research and development support on (i) User experience of the final product and (ii) the design and development of gamification techniques for promoting the proper utilisation of the system.

[5] The role of the Fondazione Bruno Kessler has been on the research and development support on (i) outbreak modelling and risk management simulation and (ii) system management support and (iii) security and privacy assessments.

[6] Technology supported the integration of Blockchain in the product

Table 6.1: Terms and acronyms used in Trace4Safe project.

Term	Description
BLE	Bluetooth Low Energy
Contact	A contact is defined by the proximity of 2 tokens for a sufficiently long time. It is defined based on the distance between 2 tokens and the time duration they will stay in proximity.
P2P	P2P refers to the monitoring of social distancing and tracing of contacts relying on the tokens only, without the need for an external infrastructure.
RTLS	Real Time Locating System: this a technology utilized to locate a device (e.g., a TAG) within a monitored area.
TAG	An active, battery-powered device used to monitor social distancing and trace contacts. A TAG can be used in conjunction with an RTLS or P2P. The term TAG can be used interchangeably with Token.
Token	An active, battery-powered device used to monitor social distancing and trace contacts. A Token can be used in conjunction with an RTLS or P2P. The term Token can be used interchangeably with TAG.
UWB	Ultra Wide Band – A high accuracy localization technology utilized to locate TAGs

6.1.1 Overview on the Trace4Safe Solution

In this section, we briefly overview the product design and specification, including the service description and the supporting system architecture. The solution is based on a token-based approach, which means a device is provided to the workers and should be carried all time during working hours. The token is a battery-powered device (e.g., a TAG) and will be meant to monitor social distancing and trace contacts. In Table 6.1, we have provided the terms and acronyms used in the project along with a short description.

6.1.1.1 Product Design

One of the main features of the product is the possibility to support different methods to monitor social distancing and trace contacts, which can be done through the following two approaches:

- RTLS: the Real Time Locating System is used to locate with high precision (less than 1 meter) the location of every single Token, and therefore every single worker. Starting from such information, the system measures the distance of every single

Table 6.2: Difference between the RTLS and P2P contact tracing approach.

	RTLS	P2P
Precision	< 1 meter	Depends on the technology used in the Token: (BLE: 1-3 meters, UWB: 0,5 meters)
Availability of data (contacts, presence)	Real-time	Offline
Support for other services (occupancy, desk reservation, asset tracking, etc.)	Yes	No
Management (e.g., reconfiguration, etc.)	Low	Medium/low depending on the technology used in the Token.
Infrastructure cost	High (requires a full RTLS to be deployed in the area)	Low

worker with each other, and monitors if social distancing is respected and in case traces the contacts of workers in close proximity. When social distancing is not respected, the system can command the Token to vibrate or buzz so to notify the involved workers of the infringement of the social distancing measures. This approach requires that all the areas covered by the service to be infrastructured with RTLS antennas.

- P2P: each token is able to measure the distance from other tokens close by. When the distance is below a recommended threshold (e.g., 2 meters) the token can provide a feedback (e.g., vibration or buzzer) and save the contact in an internal memory. A gateway placed in some high traffic location (e.g., entrance/exit from the production site) is then used to download the contacts from each Token. This approach does not require any infrastructure beyond a few gateways to download the contacts.

Trace4Safe seeks to combine both approaches into a unique solution so to combine the benefits of both: in the case of dense environments, with employees working in close proximity, an RTLS approach is to be preferred, as it provides the most accurate and reliable approach to monitor social distancing and trace contacts. Conversely, in the case of more sparse environments where workers are isolated most of the time, a P2P approach should be preferred due to its limited costs and complexity. These approaches are then integrated into one unified application that seamlessly manages the data originating from the two sources and provides the services and applications to the end-users. The key differences between the 2 approaches are reported in Table 6.2.

6.1.1.2 Service Objectives

The aim of the service is to support businesses to best manage the current situation determined by the COVID-19 pandemic. In particular, the service will target the following objectives:

1. **To monitor and enforce physical distancing**: this service should focus on the education towards a proper and safe behavior at work, based on the current physical distancing regulations. The system will then provide immediate feedback to the involved parties when such distancing is not respected. The feedback could come directly from the Token (e.g., buzzer, vibration or visual feedback) or from some dedicated device (e.g., light or screen).

2. **To analyze the risks associated with a COVID-19 outbreak**: analyze the risks related to a possible COVID-19 outbreak under the current working environment and business processes. This will be based on the analysis of the current contact network and will identify critical aspects to be addressed and reviewed.

3. **To best manage a possible COVID-19 outbreak**: preserve business productivity, confining the outbreak to the involved parties only and to the production areas affected. The system will then allow businesses to immediately react to the outbreak, identifying all parties involved and supporting them in managing the case.

All services will be based on the initial definition of a contact, which will be characterized by the distance of the parties involved and the time the parties are within such distance. As an example, a contact could be defined as "workers in less than 2 meters for more than 3 minutes". It shall be possible to define and evaluate different strategies, based on different definitions of contact.

6.2 DPIA within Trace4Safe

Given the nature of the system, with a Token handed to workers to track their contacts over time, it is extremely important to thoroughly study and understand the implications on privacy and how the system could comply with the current regulations. In accordance with Articles 35 and 36 of GDPR, before implementation, a DPIA (see Section 2.2.2) is required if the processing of personal data poses a significant risk to the freedom and rights of natural persons. In [Boa20], the EDPB ruled that a DPIA must be conducted prior to deployment of a contact tracing system, *"as the processing is considered likely high risk (health data, anticipated large-scale adoption, systematic monitoring, use of new technological solution).*

In the context of our contact tracing scenario, we can consider the occurred contacts as personally identifiable information, which contains some critical information such as Token IDs, the timestamp, as well as the location information in the case of RTLS approach. This information may lead indirectly to identifying the involved individuals. Our contribution to the project is to conduct a DPIA. By conducting a deep investigation of the security and privacy threats in the Trace4Safe contact tracing scenario. For this assessment, we first seek to identify possible issues in the system by pointing out threat scenarios together with potential mitigation solutions that need to be considered to reduce them. After that, we evaluate the security and privacy risks associated with the identified threats in the Trace4Safe contact tracing scenario.

6.2.1 Security, Privacy, and the GDPR

In general, the information security aspects of a system are dealing with the concepts of confidentiality, integrity, and availability, which are commonly addressed by means of security mechanisms as encryption, authentication, secure storage, and access control. Privacy, according to the GDPR, stands for the respect of fundamental rights and freedoms of individuals concerning the processing of personal data. It overlaps with security, particularly in regard to confidentiality, but many other privacy principles should be addressed (e.g., purpose limitation, transparency, data minimization, etc.). It implies that although privacy-preserving systems require strong security, security by itself is not enough. There exist several reasons to enforce privacy in the context of proximity contact tracing applications where the sensitive data are collected from users. Therefore, it is crucial to follow privacy principles during the design of such systems. Privacy principles have been widely discussed in the scientific literature and incorporated into legal frameworks in various jurisdictions; for instance, GDPR has laid down a general regulation for protecting personal data under organizational accountability. GDPR is a standard set of guidelines to control and protect Personally Identifiable Information (PII). It has brought significant changes to how companies and organizations should manage and process personal data, the privacy risk assessments they conduct, and the privacy compliance programs they develop to mitigate the identified risks to the privacy of the data subjects. From the GDPR perspective, a data subject is any natural person (user) in the system that his/her data (e.g., personal data, movement info, occurred contacts, etc.) are being collected, and the data controller (who applied the contact tracing system) has control over these data in the system and decides how to process the collected data. Sometimes, the processing of personal data has been done by a third party who is doing it on behalf of the data controller. It is a duty of the data controller to set up the proper technical measures for ensuring that the rights of users are respected. Given that, our main research question is *"How to design a privacy-aware and secure proximity contact tracing system?"* To answer this question, in this work, we focused on identifying security and privacy requirements, as well as potential

threats which may exist in such systems.

6.2.2 Security and Privacy Requirements

The significant risks associated with contact tracing applications are related to the possible abuse of the information these applications collect. Generally, contact tracing applications gather an outstandingly massive amount of sensitive data about the contacts and relationships between users as well as connections among users and objects such as places, devices, etc. It is noteworthy to mention, apart from those sensitive data inherently collected by contact tracing applications, individuals' health information is somehow associated in the proposed contact tracing system. In fact, in such contact tracing systems, people (especially who got infected) are more concerned about their health data in the case of being disclosed to anyone they do not trust. For this reason, the European Data Protection Board (EDPB) has specifically issued guidelines[7] on the use of location data and contact tracing tools in the context of the COVID-19 outbreak, in which the EDPB firmly believes that, when the processing of personal data is necessary for managing the COVID-19 pandemic, data protection is indispensable to build trust.

To design a proximity contact tracing system, the designer (system developer) must consider the privacy and security requirements in the early stages of design of the system. In general, the proposed system must provide the following privacy and security requirements:

1. **Fair and lawful data processing through transparency**[8]**.** The data processing activities related to the system must be transparent to data subjects.

2. **Processing only for legitimate purposes**[9]**.** The purposes of data processing must be legitimate and related to the primary purpose, which is controlling social distancing between users (workers and employees) and mitigate the risk of contagion.

3. **Data minimization**[10]**.** The central server which is controlled and monitored by the data controller must only observe anonymous identifiers of COVID-19 positive users without any further info about the users (who were close to the positive case or the places visited by positive person). Thus, the proposed system must collect and use as little personal information as possible.

[7]Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak, `https://edpb.europa.eu/our-work-tools/our-documents/linee-guida/guidelines-042020-use-location-data-and-contact-tracing_en`

[8]The lawful bases for processing are set out in article 6 of the GDPR

[9]According to Article 5(1)(b) of the GDPR

[10]According to Article 5(1)(c) of the GDPR

4. **Limited storage**[11]**.** The system must store data no longer than necessary need it, and it must explicitly define the expiration time[12] for the collected data (any data related to COVID-19 positive cases and their contacts) because users, especially those who got infected, are concerned about how long their information will be kept in the database. In other words, storage limitation should consider the true needs and the medical relevance (this may include epidemiology-motivated considerations like the incubation period, etc.) and personal data should be kept only for the duration of the COVID-19 crisis. Afterwards, as a general rule, all personal data should be erased or anonymized.

5. **Data accuracy**[13]**.** The system must take reasonable steps to update or delete data that is inaccurate or incomplete. Individuals have the right to request to erase or rectify the inaccurate data that is related to them, e.g., in the case of the wrong contact recorded by the system.

6. **The confidentiality, integrity, and availability of personal data storage, processing and transmission**[14]**.** Privacy depends on robust security, which means security measures (technical and organizational) must be used to protect data in terms of confidentiality, integrity, and availability. Any accidental or deliberate attacks to the system may cause data breaches, loss, abuse, etc. and consequently results in compromising individuals' privacy and safety.

7. **No abuse of data (ensures purpose limitation).** The central server must receive the minimum amount of information tailored to its requirements, to be prevented from misusing the collected data for other purposes. In other words, with regard to the principle of purpose limitation, the purposes must be specific enough to exclude further processing for purposes unrelated to the management of the COVID-19 health crisis (e.g., commercial or law enforcement purposes).

8. **No tracking of users.** No one can track users' location and contacts, it means these data must be kept anonymous in the system and the system must only collect these data for the primary purpose, which is creating the network of contacts to be kept social distancing by warning users. Depending on the chosen implementation, users can only receive alert notifications relating to the happened contacts.

9. **Anonymity of users.** Always, the identity of users must be held anonymous for all involved individuals in such contact tracing systems.

[11] According to Article 5(1)(e) of the GDPR

[12] According to the requirements set out by the EDPB in the analysis guide section which says "Any identifier included in the local history must be deleted after X days from its collection (the X value being defined by the health authorities)."

[13] According to Article 5(1)(d) of the GDPR

[14] According to Article 5(1)(f) of the GDPR

Furthermore, we have pointed out some other privacy concerns that may induce users to distrust the system, for instance, in the case of any unauthorized access, abuse of data, disclosure, and so on. If these privacy concerns are not considered in the system, users may refrain to using the service (e.g., wearing the device) and/or hide their positive COVID-19 test infection. The additional privacy concerns are the following:

(i) **Tracking users.** Users are concerned about being tracked in the system, e.g., by their bosses. On the other hand, tracking users may result in profiling users' behaviors (e.g., how many times an individual goes out of the place that he must be in, or how often he goes to smoke, etc.) so that users may feel that they lost their individual freedom and rights. Although contact tracing systems do not explicitly collect or record the true identities of individuals, movement profiles based on pseudonymous tracking data make it possible to identify a large fraction of users with high probability.

(ii) **Identifying positive cases.** Fear of being identified (or revealed) the identity of the user (especially among their friends) who got infected makes users distrust the system. So, users may try to hide their illness or prefer not to use the device (token). The system must ensure the infected cases' identities and contacts remain anonymous forever, not only during their treatment, but also after the cases heal.

(iii) **Identifying users' interactions and relations.** The users may want nobody knows about the usual interactions they have in the organization. The central system must not reveal any information about close-range physical interactions between users to any entity which can be used to analyze users' communications.

(iv) **Identifying exposed places.** The system shall not reveal any information about exposed places where the positive cases have passed, or any data about the number of positive cases in a particular area. Proximity tracing must be done without any entity finding out about these sorts of information.

As a preparation for data protection impact assessment, we provided a questionnaire (see Appendix C.1, Table C.1) and shared it with other partners involved in the project. This list of questions helped us to understand the privacy issues in each stage of the scenario; We mapped each question to a privacy principle (e.g., Data quality) and a privacy target (e.g., ensuring data minimization). This list of privacy targets and principles (see Appendix C.2, Table C.2) are derived from existing literatures in the context of privacy [OS14, OSG+11] and the legal principles put forward by the GDPR. Indeed, these principles align with the data protection goals introduced in [fD17], and the privacy targets can assist in identifying mitigation controls. In the following sections, we will first review the processes and procedures of the contact tracing system (in Section 6.3), and then we will describe how and when the data is collected and for what purposes they are used. This step helps us

in identifying potential security and privacy risks within the system that might have deployed the Trace4Safe solution. Following that, we evaluate the risks posed by security and privacy threats in the system (by conducting a risk assessment process in Section 6.4) and the potential consequences of those risks.

6.3 Processes and Procedures

This section focuses on analyzing the various phases in which data is processed in the system and how this activity contributes to the production of input artifacts for the risk assessment process. We identify four phases, pointing out the data exchanged and the relevant aspects concerning privacy. For each event, we have provided a table in which the first column represents the exchanged data, where the data owner is the worker. In column 2, we report who provides the data: it can be either a person or a device who stored (e.g., system) or collected (i.e. Token) the data. In column 3, we report the entity receiving the data. Other relevant information is the purpose (column 4) and the communication method (column 5) used to transfer the data, given that the system must ensure that sensitive data are kept confidential during the transmission.

6.3.1 Registration

During this phase, the worker (employee or individual who will be wearing the wearable device) must provide some basic information (such as his or her name, surname, email address, and role in the organization) to the COVID safety person who is responsible for registering users in the system and assigning them a token (rows 1 and 2 in Table 6.3). Apart from this personal information, users may be asked to supply information regarding their health condition, such as if they use a pacemaker. With this information, the system will be able to update risk analysis parameters for those who are considered to be at high risk of developing a serious illness.

Table 6.3: Data exchange in registration phase.

Data	Who provides data	To	Purpose	How
Personal info + worker's health condition (in the case of having a pacemaker)	Worker	COVID safety responsible	To register the user and assign a token to him/her.	In-person
	COVID safety responsible	System		system interface

6.3.2 During Working Shifts

As we mentioned in Section 6.1.1, the system supports two approaches for monitoring social distancing and tracing contacts. Given that they collect different data about the employees, they have a different impact concerning privacy.

6.3.2.1 RTLS approach

The tokens (the wearable device) are collecting data consists of workers' locations to measure physical distancing and these data will be sent to the system in order to store for a later process, for example, for creating the network of contacts (row 1 in Table 6.4).

6.3.2.2 P2P approach

In this approach, the tokens themselves measure the physical distance between each other (between two users) and in the case of two tokens being closer together than the defined distance, the tokens consider it as a contact and store it in their memory storage (row 2 in Table 6.4). Once the tokens get close to a gateway, they upload the data on the gateway (row 3 in Table 6.4), and after receiving data from the tokens, the gateway will forward the data on the edge server or directly to the cloud in order to store for later processing (row 4 in Table 6.4).

Table 6.4: Data exchange during working shifts.

Data	Who provides data	To	Purpose	How
Users' Location and Contacts(RTLS approach)	Token	System	To measure physical distancing and to store for a later process (creating network of contacts)	BLE/UWB
User's Contacts (P2P approach)	Token	Token	To store for a later process (creating network of contacts)	BLE
	Token	Gateway	To forward the data to the edge server or directly to the cloud	Upload on the gateway through BLE channel
	Gateway	Edge server/ Cloud/System	To store for a later process (creating network of contacts)	Ethernet network

6.3.3 Reporting by Users

This phase concerns the data exchanges if users report any illness, symptoms, or positive cases. Workers can provide information regarding their health condition to the team leader in the case of feeling sick, having symptoms, or reporting the positive test (see row 1 in Table 6.5) through the defined communication methods (e.g., email, call, messaging, or other possible methods). After the team leader gets informed about a possible COVID-19 outbreak, he/she activates a COVID-19 management procedure in the system. The system processes the info to create a network of contacts to inform the COVID safety responsible person (see rows 2 and 3 in Table 6.5). Then this person prepares information for the team leader by considering different aspects of the situation (such as the severity of the event, the number of people involved in the contact network, etc.), and this helps the team leader to develop an adequate plan for isolating the ill person from others in the workplace, limiting the number of workers who have contact with the sick person, and contacting the local health authorities for the treatment. In light of the received feedback, the team leader can take appropriate decisions, for example, redesigning the work shifts or enforcing the isolation of exposed team members (see rows 4 in Table 6.5).

6.3.4 Sending Report by the System

Sometimes, the system may provide information (e.g., name and surname) about the workers who have infringed safe work guidelines to the team leader, or the system may directly send information to the worker about his/her behavior (e.g., gamification approach).

Table 6.5: Data exchange by users in the case of illness, symptoms, or positive case.

Data	Who provides data	To	Purpose	How
User's Health condition	Worker	Team leader	To inform the company that a possible COVID-19 outbreak could be in place.	Email/Call/messaging or other possible methods
	Team leader	System	To activate a COVID-19 case management procedure.	system interface
Network of contacts	System	COVID safety responsible		email/messaging
Health Risks	COVID safety responsible	Team leader, workers	To notify the redesign of the work shifts or to enforce the isolation of exposed team members.	Feedback/Notification: tag/public screen/email/messaging

6.4 Risk Assessment

In the previous sections, we reviewed the Trace4Safe solution and its data processing and procedures with the aim of acquiring a comprehensive understanding of the solution by highlighting the security and privacy requirements and potential threats. Up to this point, the ultimate goal was to provide feedback about the design and the ongoing implementation of the Trace4Safe solution. In this section, we go one step further by analyzing security and privacy risks associated with the Trace4Safe contact tracing scenario. We will assess the risks associated with the Trace4Safe scenario in the following sections by adopting the processes outlined in Chapter 4 (see Figure 4.2). We assess risk levels for all stakeholders in the scenario and also explore alternative risk management policies to help designers conduct a DPIA.

6.4.1 Risk Identification Process

Contact tracing is a privacy-invasive procedure as it collects quite sensitive personal information. Without assuring the security and privacy of the gathered data, users' privacy can be seriously threatened. Accordingly, it requires a comprehensive investigation of any contact tracing protocol's security and privacy implications. The reference scenario in which the Trace4Safe solution will operate is shown in Figure 6.2, which depicts the main actors and their interactions with the system's components. In this figure, the system's main actors are the COVID safety responsible person as an authorized person for the service provider (i.e., the organization that uses the Trace4Safe product) and the end-users (i.e., employees). Also, the team leader can access the contact tracing data as an authorized person. For ease of reference, we have assigned a number to each communication channel among components and between each actor and a component where data are collected and transferred (e.g., number 4 is the channel between gateways and the edge server). As depicted earlier in Figure 4.2, we called `Artifact Preparation` the *Risk Identification* step, where various tasks must be carried out in order to prepare artifact inputs for evaluating risk levels.

Table 6.6: Sending report by system to manager/employee.

Data	Who provides data	To	Purpose	How
Personal info (name and surname)	System	Team leader	To take proper initiatives against the worker	Email/messaging
Individual behaviour	System	Worker	Change behaviour/ gamification	Email/messaging

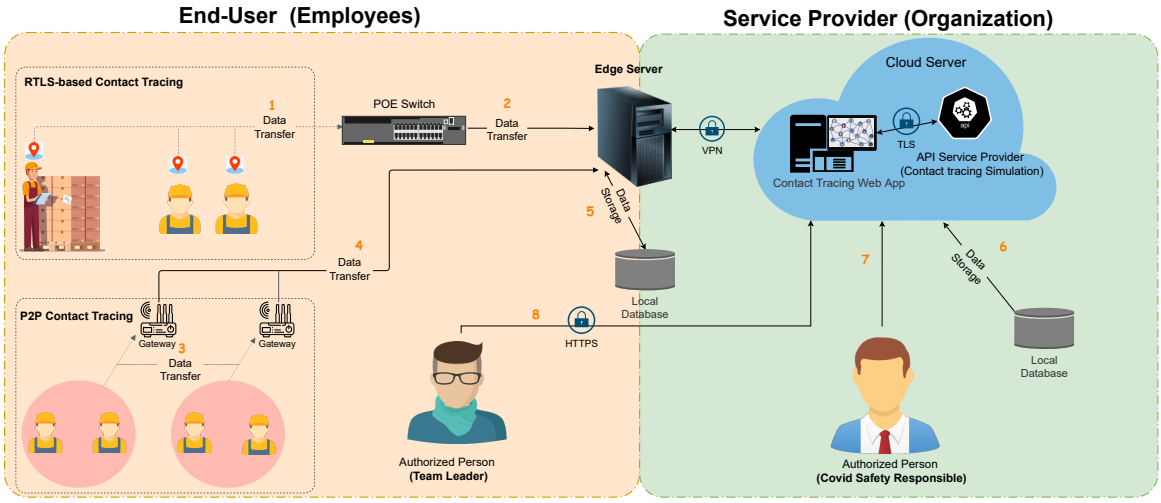

Figure 6.2: System's components and the main actors in Trace4Safe.

6.4.1.1 Threat Identification

For this task, we identified a list of security and privacy threats related to our contact tracing scenario along with their consequences (See Table C.3 in Appendix C). We also mapped each of these threats to the corresponding affected components/channels reported in Figure 6.2. It is worth mentioning that, for security threats, we adopted Microsoft STRIDE threats [Sho14], and we used LINDDUN [WJ15] for privacy threats. Although Table C.3 listed fifteen security and privacy threats, for this assessment, we assumed that the local databases on the edge server and cloud keep data securely and hence neglected the threats associated with database breaches caused by unauthorized access. Also, some privacy threats, such as purpose limitation and transparency, do not affect personal data. In Table 6.7, five threats (out of the fifteen threats considered in Table C.3) are chosen for the risk assessment because they have a more adverse impact on either the contact tracing network (e.g., manipulating the contact network) or have some consequences on users, like tracking and identification.

6.4.1.2 Mitigation Controls Identification

We identified mitigation controls for each threat and reported them in the second column of Table 6.7. For instance, *Regularly checking Token's battery* is considered as a security mitigation control for the *Battery drain attacks* (i.e., T3), whereas *Facilitating the report by workers to the system* is considered as a privacy mitigation control for the *intervenability threat* (i.e., T5) by exploiting the "privacy target" concept defined in [OS14].

Table 6.7: Selected threats and their mitigation controls.

Threats ($\mathcal{T}$)	Mitigation Controls $\{\mathcal{C}_T\}_{T\in\{T1,T2,T3,T4,T5\}}$
T1- Tracking and Eavesdropping	c_1) Using pseudo-random identifiers and changing over time c_2) Using Encryption methods
T2- Data Tampering	c_3) Encrypting data-at-rest and data-in-transit by using encrypted connections such as SSL, TSL, HTTPS, etc. c_4) Assign role-based controls to restrict access to encrypted data c_5) Implement multifactor authentication
T3- Unlimited Data Storage	c_6) Implementing deletion mechanism c_7) Ensuring data minimization c_8) Using anonymization techniques in data-at-rest
T4- Battery drain Attack (Dos Attack)	c_9) Network monitoring c_{10}) Regularly checking Token's battery c_{11}) Validating contacts on Token's side c_{12}) Define a rate-limitation (e.g., control the rate of requests sent or received by a Token).
T5- Intervenability Threat	c_{13}) Facilitating the report by workers to the system c_{14}) Handling the workers' change requests c_{15}) Informing the workers about data processing (e.g., providing information about their daily activities by mentioning the contacts, locations, etc.)

6.4.1.3 Stakeholders & Protection Criteria Determination

As can be seen in Figure 6.2, we have two main stakeholders: (i) the organization as the service provider who plans to deploy the Trace4Safe solution in its organization, and (ii) the end-users, who are the organization's employees, and their contact activity will be collected by carrying a wearable device. From the GDPR perspective, the organization is the data controller and end-users are the data subjects.

These stakeholders have different protection criteria when they want to evaluate the risk of potential threats. For instance, organizations as the data controller typically adopt business impact criteria, such as financial or reputation impact, whereas data subjects evaluate risk on the basis of their impact on their personal sphere. We consider that the *reputational* and *financial situation* are the impact criteria for the organization, which are linked to indirect and direct pecuniary loss or damage deriving from privacy violations, and by taking inspiration from [OS14, MSR20], we have specified the following impact criteria for the employees in the system who are concerned about:

- **Health condition**: Since the contact tracing scenario has a direct relation with the health condition of employees, they are concerned about their health, any attack to disrupt contact data may result in wrongly tracing the contacts. Consequently, it impacts on the health condition of employees.
- **Individual freedom**: Employees are worried about being tracked, as we enumerated some privacy issues in Section 6.2.2. This could lead to profiling of user behaviors, such as relationships and interactions, and these may result in losing their individual freedom.
- **Social situation**: This criterion may be compromised when an information disclosure happens as a consequence of a data breach. This can result in more severe health concerns and make disease outbreaks more difficult to control. As a consequence, employees may face discrimination at the moment of the COVID-19 pandemic or in the future.

6.4.1.4 Protection Goals Specification

We introduced the data protection goals in Section 2.2.3, and for this assessment, we have chosen five protection goals that can be compromised by the identified threats in the scenario, namely: **C**onfidentiality, **I**ntegrity, **A**vailability, **U**nlinkability, and **In**tervenability.

6.4.2 Risk Analysis Process

This process mainly focuses on establishing relations between the prepared artifacts derived in the previous process. As we assigned an abstract definition for this process (see Figure 4.2)—i.e., `Association Processes`—three sub-process tasks must be done in this step: Threat-Controls association, Threat-Protection Goals association, and Threat-Protection Criteria association. Below, we briefly describe these three tasks:

-**Threat-Controls association.** In this task, we must identify mitigation controls for each threat. As we reported earlier in Table 6.7, fifteen security and privacy mitigation controls have been identified for the selected threats. For instance, the controls *Using pseudo-random identifiers and changing over time* and *using Encryption methods* can mitigate the risk of *Tracking and Eavesdropping* threat (T1).

-**Threat-Protection Goals association.** This task specifies the relation between selected protection goals and threats. As described in Chapter 4 (see Section 4.2.2), the purpose of this task is to identify how many goals each threat is impacting and then measure the amplitude of the impact on each goal of a given threat. In Table 6.8, we specified the association between the identified threats and the five selected protection goals from the

Table 6.8: Association between threats and protection goals.

Threat	Protection Goals					Observation Weight
	C	I	A	U	In	
T1	×			×		2/8
T2		×				1/8
T3	×			×		2/8
T4		×	×			2/8
T5					×	1/8

previous process. The observation weight value reported in the third column is computed by using Formula 4.2 at page 59.

-Threat-Protection Criteria association. This association defines impact levels according to stakeholders. For each identified threat, we must estimate the adverse impact level on each protection criterion of the stakeholders. However, for the sake of brevity, we reported the result of this association containing the aversion levels for each threat on the stakeholders' criteria in Table 6.9. The whole analysis is reported in Appendix C (see Section C.4).

At the end of this process, as shown in Table 6.9, the output is expressed in a quantitative form where we need the aversion levels and the weight assigned to each protection criterion for the risk evaluation process.

Table 6.9: The estimated aversion levels for each selected threat on the stakeholders' protection criteria, and the assigned weight to each stakeholder's preference.

Stakeholder	Protection Criteria	Weights	Aversion Level				
			T1	T2	T3	T4	T5
Organization	Financial situation	0.7	3	3	3	3	1
	Reputational situation	0.3	3	2	1	2	1
Employee	Health condition	0.5	0	4	0	3	3
	Individual freedom	0.2	4	0	0	0	2
	Social situation	0.3	3	2	1	0	2

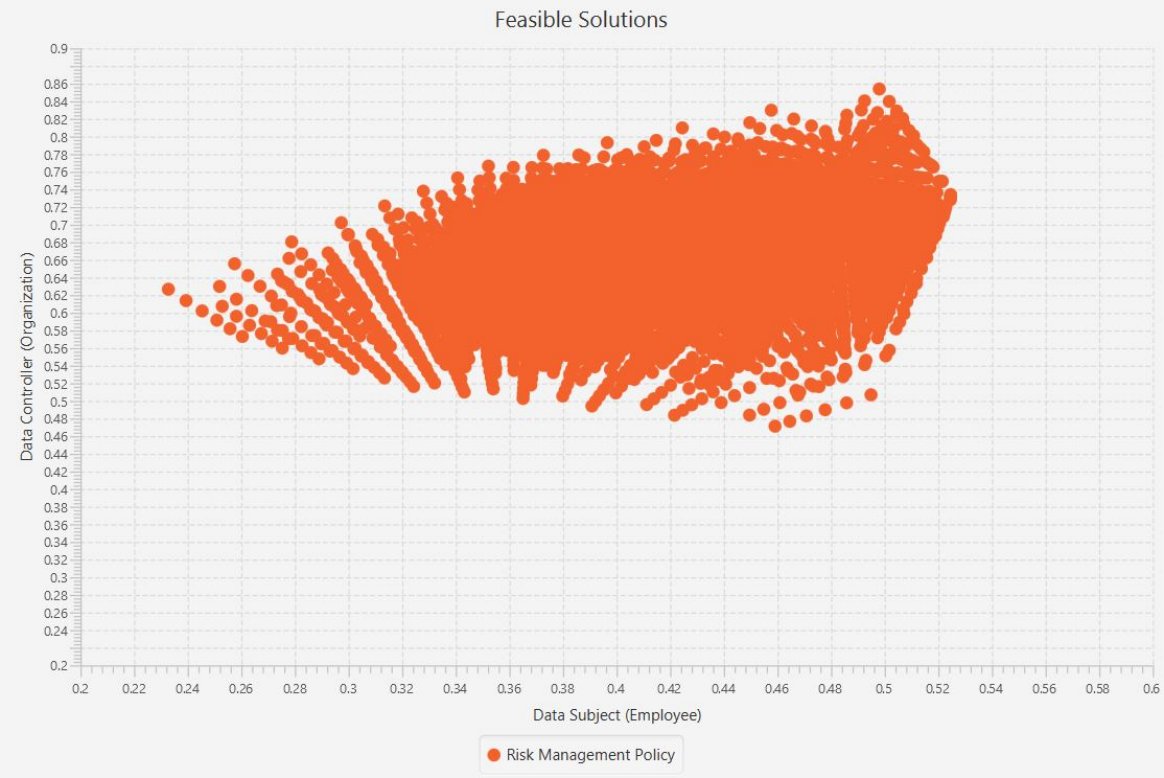

Figure 6.3: All feasible solutions in Trace4Safe scenario.

6.4.3 Risk Evaluation Process

Referring back to Figure 4.2, this process is composed of several computation tasks (sub-processes) that evaluate risk levels by the inputs provided from the previous process. We followed all these processes as described in Chapter 4 (see Section 4.2.2). To avoid re-describing each task, we briefly report only the results in the following section.

6.4.3.1 Applying the Prototype Tool on Trace4Safe Scenario

This section presents some experimental results from our tool for the Trace4Safe scenario (as described in Chapter 5). The input data for the tool are the list of threats and their mitigation controls (see Table 6.7), the stakeholders and their protection criteria introduced in Section 6.4.1 along with the aversion levels (see Table 6.9), and the association between threats and protection goals reported in Table 6.8. By applying the tool on Trace4Safe Scenario, the obtained feasible set contains $6,912$ solutions. This means the search space around 1.6K times is reduced through the approach we discussed at the end of Section 3.2.2 and also reported a comparison in Table 5.4. In Figure 6.3, we plotted the all feasible

solutions and each point represents a solution under an RMP. The x-axis depicts the risk exposure level of the data subject (employee), while the y-axis depicts the risk exposure level of the data controller (organization).

After generating the feasible set of solutions, we need to apply the Pareto optimality algorithm to find the optimal solutions. Table 6.10 lists twenty-two found Pareto solutions, where the overall risk residue for the employee and the organization are reported in the second and third columns. The retrieved risk residuals (i.e., x_T) for each threat are reported in Table 6.11, which shows the residual risks under the identified Pareto solutions.

Table 6.10: Pareto's solutions for Trace4Safe scenario.

Pareto Solutions	$oir(Data\ Subject)$	$oir(Data\ Controller)$
$\mathbb{S}_1$	0.3243	0.5164
$\mathbb{S}_2$	0.3134	0.5261
$\mathbb{S}_3$	0.3017	0.5366
$\mathbb{S}_4$	0.289	0.5478
$\mathbb{S}_5$	0.2753	0.5599
$\mathbb{S}_6$	0.2604	0.5731
$\mathbb{S}_7$	0.3432	0.5101
$\mathbb{S}_8$	0.3651	0.5028
$\mathbb{S}_9$	0.3908	0.4942
$\mathbb{S}_{10}$	0.4216	0.4838
$\mathbb{S}_{11}$	0.4591	0.4713
$\mathbb{S}_{12}$	0.323	0.5216
$\mathbb{S}_{13}$	0.3116	0.5318
$\mathbb{S}_{14}$	0.2993	0.5429
$\mathbb{S}_{15}$	0.286	0.5548
$\mathbb{S}_{16}$	0.2715	0.5678
$\mathbb{S}_{17}$	0.2558	0.5819
$\mathbb{S}_{18}$	0.3424	0.5154
$\mathbb{S}_{19}$	0.2509	0.5916
$\mathbb{S}_{20}$	0.2454	0.6021
$\mathbb{S}_{21}$	0.2394	0.6138
$\mathbb{S}_{22}$	0.2328	0.6267

6.4.3.2 Discussion on the Results

At this point, we are in the position of selecting the optimal solution out of all the identified Pareto solutions. To this end, we define two criteria that can help in the selection process as follows:

(i) Look for a solution with fewer residual risks equal to 1, which means the solution has a major contribution in mitigating more threats.

Table 6.11: Retrieved risk residual values for the identified Pareto solutions.

	$x^*_{T_1}$	$x^*_{T_2}$	$x^*_{T_3}$	$x^*_{T_4}$	$x^*_{T_5}$
$\mathbb{RS}_1$	0.25	1	1	0.125	1
$\mathbb{RS}_2$	0.25	1	1	0.125	0.83
$\mathbb{RS}_3$	0.25	1	1	0.125	0.66
$\mathbb{RS}_4$	0.25	1	1	0.125	0.5
$\mathbb{RS}_5$	0.25	1	1	0.125	0.33
$\mathbb{RS}_6$	0.25	1	1	0.125	0.16
$\mathbb{RS}_7$	0.25	1	0.83	0.125	1
$\mathbb{RS}_8$	0.25	1	0.66	0.125	1
$\mathbb{RS}_9$	0.25	1	0.5	0.125	1
$\mathbb{RS}_{10}$	0.25	1	0.33	0.125	1
$\mathbb{RS}_{11}$	0.25	1	0.16	0.125	1
$\mathbb{RS}_{12}$	0.25	0.83	1	0.125	1
$\mathbb{RS}_{13}$	0.25	0.83	1	0.125	0.83
$\mathbb{RS}_{14}$	0.25	0.83	1	0.125	0.66
$\mathbb{RS}_{15}$	0.25	0.83	1	0.125	0.5
$\mathbb{RS}_{16}$	0.25	0.83	1	0.125	0.33
$\mathbb{RS}_{17}$	0.25	0.83	1	0.125	0.16
$\mathbb{RS}_{18}$	0.25	0.83	0.83	0.125	1
$\mathbb{RS}_{19}$	0.25	0.66	1	0.125	0.16
$\mathbb{RS}_{20}$	0.25	0.5	1	0.125	0.16
$\mathbb{RS}_{21}$	0.25	0.33	1	0.125	0.16
$\mathbb{RS}_{22}$	0.25	0.16	1	0.125	0.16

(ii) In the case where more than one solution satisfies the first defined criterion above, look at the observation weight value for the threats. As we described in Section 4.2.2, the more a threat impacts multiple goals, the more it is considered pervasive. Therefore, a solution must be selected where the threat with the highest observation weight—that means the threat is more critical—has the lowest residual risk value.

By looking at Table 6.11 and considering the first criterion above, we can select the retrieved risk residual values from $\mathbb{RS}_{13}$ to $\mathbb{RS}_{22}$ where there is only one residual risk is equal to 1. Among these selected residual risks, only for $\mathbb{RS}_{18}$, the residual risk of threat $T5$ is 1, whereas for other cases, the residual risk of $T3$ is equal to 1. By considering the second criterion, among $T3$ (i.e., *unlimited data storage* threat) and $T5$ (i.e., *intervenability* threat), as reported in Table 6.8, $T3$ impacts two protection goals, the *confidentiality* and *Unlinkability* whereas $T5$ impacts only on *intervenability*. Therefore, $\mathbb{RS}_{18}$ can be selected as the optimal solution, where the overall risk for the employee and organization are 0.3424 and 0.5154, respectively, and the residual risk value for each threat is as below

$$\langle x^*_{T1}, x^*_{T2}, x^*_{T3}, x^*_{T4}, x^*_{T5} \rangle = \langle 0.25, 0.83, 0.83, 0.125, 1 \rangle$$

To conclude the discussion, the third column of Table 6.12 illustrates five possible mitigation mappings associated to the optimal solution selected above. Notice that all mitigation

Table 6.12: Five possible mitigation mappings associated to the optimal solution

Threats ($\mathcal{T}$)	Controls $\{\mathcal{C}_T\}_{T\in\{T1,T2,T3,T4,T5\}}$	Possible Mitigation Combinations					x^*_T
T1	c_1	●	◐	◐	●	●	0.25
	c_2	◐	●	●	◐	◐	
T2	c_3	◐	○	○	◐	○	0.83
	c_4	○	◐	○	○	◐	
	c_5	○	○	◐	○	○	
T3	c_6	○	○	◐	○	○	0.83
	c_7	○	◐	○	◐	○	
	c_8	◐	○	○	○	◐	
T4	c_9	●	◐	●	●	●	0.16
	c_{10}	●	●	◐	◐	●	
	c_{11}	●	●	●	●	◐	
	c_{12}	◐	●	●	●	●	
T5	c_{13}	○	○	○	○	○	1
	c_{14}	○	○	○	○	○	
	c_{15}	○	○	○	○	○	

mappings associated to the optimal solution above suggest avoiding implementing any mitigation control for the threat $T5$.

Additional to this assessment we have performed one further assessment, in Section C.5 of Appendix C, where we have assessed the Trace4Safe contact tracing solution with the requirements that European Data Protection Board (EDPB) is designed on the use of location data and contact tracing tools in the context of the COVID-19 outbreak.

Chapter 7

Related work

This chapter discusses the related work. First, we look into several techniques for assessing cybersecurity risks that are widely used in literature in Section 7.1. Then, in Section 7.2, we review some privacy impact assessment techniques along with covering some methodologies, standards, GDPR guidelines, and some existing tools for assessing privacy impacts. We examine different methods for selecting controls and multi-criteria risk assessment techniques in Section 7.3. Finally, in Section 7.4, we provide a discussion of our lesson learned from the state of the art, the limitations of existing approaches, and highlight our contributions.

7.1 Cybersecurity Risk Assessment Methodologies

In the scope of information security, a wide range of risk assessment approaches have been proposed by standard institutes and organizations like NIST SP 800-30, ISO/IEC 27005, etc. Each of these methodologies has its own specific scope, procedures, and assessment technique [MKW+16]. Despite this, they all follow the same path: plan, execute, and report on the results [QJD+19]. The process of preparing for an assessment begins with a thorough inventory of the facility's hardware and software, followed by a review of all applicable regulations, policies, procedures, and controls. In the second phase, the assessment is put into action, which entails looking for potential security flaws and software pitfalls. It is then documented and coordinated that the reported flaws have been remedied. The National Institute of Standards and Technology (NIST) [GAA02] published a special publication in 2002 that reflects the guideline for the process of organizational risk management, and it was revised for the first time in 2012 [NIS12]. Framing, assessing, responding, and monitoring risks have all been included in the risk management processes outlined in this guideline. The first process demonstrates how security researchers are

framing or constructing a risk context in order to develop a risk management strategy for further assessing, responding to, and observing risks. Following that, the risk is assessed based on the frame of risk to identify external and internal vulnerabilities, as well as threats to the investigated system, thereby preventing potential harm. Meanwhile, responding to hazards demonstrates how security researchers should respond when a risk is identified during the evaluation. As the last process, monitoring risks relates to how organizations keep track of risk throughout time, notably in terms of verifying the effectiveness of reactions to risks and determining shifts in operating systems that are caused by risk.

ISO/IEC 27005 [2713] is the risk analysis standard for the ISO 27000-series. The ISO/IEC 27000 standards are designed to assist businesses in maintaining the security of their information assets. The standard establishes principles for information security risk management and applies to any type of organization that wishes to manage risks effectively in order to prevent jeopardizing the business's information security. ISO/IEC 27001, the information security management system (ISMS) standard, is the most well-known of the ISO 27000 family. Its purpose was to establish requirements for the execution of information security in accordance with a risk management framework.

Octave Allegro is the next generation of the Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE) methodology for identifying and evaluating information security risks. The OCTAVE risk assessment process is designed to be as efficient as possible when working with limited resources. Generally speaking, it is best suited for smaller to mid-size businesses. This methodology is primarily concerned with information assets in terms of their use, storage, transportation, and processing, as well as their exposure to threats, vulnerabilities, and disruptions as a result [BFH+16].

Factor Analysis of Information Risk (FAIR) [FJ14b] is a pragmatic risk management methodology that identifies and quantifies threats to a business's operational and cybersecurity frameworks. The FAIR model, which is compliant with international standards, was developed in 2005 and is widely regarded as the leading Value at Risk (VaR) framework for operational risk and cybersecurity. The FAIR model discovers and aggregates many aspects that may pose a risk to an organization, and it then thoroughly examines how these factors relate to or trigger another potential concern for the organization.

Regardless of the particular processes each of these security risk assessment approaches above have, they all point out to the risk as an unexpected incident that would damage business assets, either tangible (e.g., organization's hardware infrastructures) or intangible (e.g., organization's services). The ultimate goal of an information security program based on risk management is to augment the organization's output (product and service) while simultaneously limiting the unexpected adverse outcomes generated by potential risks. However, these risk assessment methodologies are restricted in terms of what risks are related to data subjects and how to evaluate them, which is requested by more and more legal frameworks around the world, especially concerning privacy and other fundamental

rights. The GDPR and other applicable regulations mandate a risk-based approach and expressly recommend the execution of Data Protection Impact Assessments (DPIA), which is based on an assessment of the privacy impact against the privacy rights of data subjects. Such assessments are discussed in the following.

7.2 Privacy Impact Assessment

Privacy Impact Assessment (PIA) is a risk management technique that entails assessing the possible impact of systems on privacy as a result of processing operations on personal data [Cla09]. Organizations should anticipate risks associated with their efforts throughout their life-cycle, beginning with the design phase but also during their operational life-cycle through iterative evaluation. Security by Design (SbD) and Privacy by Design (PbD) are principles increasingly identified as necessary for dealing with design faults that may jeopardize security or privacy in the system [AS17]. Furthermore, their application is also envisaged by the GDPR, as it demands Data Protection by Design (DPbD) for any systems that involve personal data in their processing. Conducting PIA is mandated by data protection authorities (DPAs) and standardization bodies, who have developed legislative frameworks and guidelines. The General Data Protection Regulation (GDPR) asserts that the data controller must do an impact assessment and document it before starting to process the data. This is done to make European citizens more trustful of digital services (Art. 35). The International Organization for Standardization (ISO) released ISO/IEC 29134:2017 gives guidelines for (i) a process on privacy impact assessments, and (ii) a structure and content of a PIA report [ISO17a]. Numerous methodologies and frameworks in the context of privacy impact assessment have been proposed, we overview some methodologies, standards, and GDPR guidelines in the following section, and then discuss tool support for PIAs in Section 7.2.2.

7.2.1 Methodologies, Standards and GDPR Guidelines

Several privacy data protection standards, including BS 10012:2017 [BSI17], ISO/IEC 29151:2017 [ISO17b], and ISO/IEC 27018:2014 [ISO14], incorporate the PIA as a required step in performing cyber risk assessments. In the absence of a clear methodology, it is impossible to conduct a PIA in conjunction with a risk assessment procedure. Even though, according to the NIST privacy framework [BL$^+$20], data protection lies at the convergence of cyber security and privacy, the great majority of organizations treat the PIA separately from the cyber risk assessment [OS14, WWLC20]. Although ISO/IEC 29134:2017 provides detailed guidelines for conducting a PIA, it only outlines the fundamental concepts of impact analysis and provides insufficient information for the risk assessor [WWLC20]. The

literature has documented countless privacy metrics, although these often employ criteria of privacy-enhancing technologies (PETs), such as the quantification of leaked information or the number of indistinguishable users, rather than the impact on privacy [BG19].

Prior to the GDPR, PIA was not a legally required assessment. When data processing is likely to result in a high risk to the rights and freedoms of natural people, controllers are mandated to undergo a Data Protection Impact Assessment (DPIA), as provided by article 35 of the GDPR. On the other hand, the GDPR does not prescribe a specific assessment process [VDGR16, DLM16, BFH+16, DLM17, vPH17, MH15] while at the same time mandating a clear understanding of the Personal identifiable information (PII), because any improper management of PIIs may be considered a violation of the GDPR.

The French National Institute of Data Protection (CNIL [CNI18]), the British Information Commissioner's Office (ICO [ICOI18]), and the Canadian Privacy Act [oCS10] are just a few of the national authorities who have issued guidance for DPIA. Such instructions have been updated to better serve DPIAs and to provide thorough guidelines on the regulatory standards and processes that they must follow. These guidelines include a variety of techniques and provide a variety of processes for performing a PIA, but they are abstract or vague, making it extremely difficult to conduct such methodologies [ASRJ18]. As a result, organizations have difficulty adopting a single methodology, which leads to a lack of support for PIAs [VK18]. To demonstrate the lack of completeness among the most well-known DPIA approaches, a more recent study examined seventeen questions culled from the literature [VK18].

Along with the regulatory steps described above, academics have recommended improvements to the DPIA processes, which are currently under consideration. To this end, formal modeling techniques for privacy threats are being used to make the DPIA process more systematic and structured. For instance, LINDDUN [WJ15] is a threat modeling framework to identify privacy threats, and it comprises of six main processes that can assist analysts in systematically eliciting and mitigating privacy threats. It is an acronym for *Linkability, Identifiability, Non-repudiation, Unawareness, Detectability, Disclosure of information,* and *Non-compliance*. Both LINDDUN and the CNIL methodology are based on the same principles. In comparison to the CNIL, however, LINDDUN includes the capability to visualize data flow diagrams and privacy threat tree patterns. From a legal perspective, however, LINDDUN lacks assessment steps and is not integrated into a risk assessment process [PMGG21].

The Standard Data Protection Model [fD17] (SDM) provides appropriate measures to transform the regulatory requirements of the GDPR to qualified technical and organizational measures. For this purpose, the SDM first records the legal requirements of the GDPR and then assigns them to the protection goals *Data Minimization, Availability, Integrity, Confidentiality, Transparency, Unlinkability* and *Intervenability*. The SDM thus transposes the legal requirements of the GDPR on protection goals into the technical and

organizational measures required by the Regulation, which are described in detail in the SDM's catalog of reference measures. However, this methodology can not be used alone for conducting a risk assessment; indeed, it is a valuable supplement to performing a DPIA where it can help data controllers to specify which GDPR requirements may be at risk in a system.

7.2.2 PIA Tools

Existing PIA tools can be broadly classified as products of the following standardization efforts and their resulting schemes: ENISA Tool [ENI20], GS1 EPC/RFID PIA Tool [GS115], CNIL tool [CNI], SPIA Tool [SPI16], and ASPIA Tool [PMGG21]. The majority of tools on the market have a limited application scope, with typically a single use case. Existing solutions facilitate the documentation of data processing procedures, the formation of consent templates, and the documentation of privacy and data protection policies in significant numbers. Nonetheless, the cybersecurity posture of the organization performing the impact analysis is largely disregarded [PMGG21].

ENISA Tool. ENISA has provided an online tool for assessing the amount of risk associated with the processing of personal data [ENI20]. This tool is intended to provide direction to small and medium-sized enterprises (SMEs) and help data controllers and processors. The adopted approach includes six steps that give a streamlined approach, steer SMEs toward a data processing operation, and enable them to assess privacy-related security risks. The assessor establishes the context of the processing operation by following the processes that have been suggested, and then manually analyzes how the fundamental rights and freedoms of individuals may be compromised as a result of the potential breach of the security of the personal data. Four levels of impact are supported, ranging from Low to Very High. Furthermore, the assessor manually documents both external and internal threats to the system and assesses the likelihood of their occurrence. The final risk assessment is provided following an analysis of the personal data processing operation's impact and the associated threat probability. The tool facilitates the process of adopting new security and privacy measures based on the outcome.

GS1 EPC/RFID PIA Tool. The tool [GS115] aids in the assessment of privacy issues associated with RFID implementations and assists in the selection of privacy safeguards to be addressed during application development. The tool is an MS Excel file that facilitates in the calculation of risk level scores using the formula

$$Risk = Impact \times Likelihood - Controls.$$

To evaluate the residual risk, the score takes control efficacy into account. The assessor answers specific questions/considerations during the procedure and can establish arbitrary controls and their effectiveness on a scale of 1 to 5. When it comes to privacy issues

that can be triggered due to actual attack vectors targeting the deployment, the tool does not focus on identifying technical aspects of the implementations. Furthermore, the score criteria are fairly vague and unspecific for privacy threats [Aga15], and the assessment is limited to the technology sector of EPC/RFID applications.

CNIL Tool. The CNIL tool [CNI] is designed to help data controllers perform DPIAs using the methodology released by CNIL in [CNI18, CNI12]. According to CNIL's methodology, a PIA is based on two main aspects: (i) fundamental rights and principles, which are "non-negotiable", mandated by law and which must be respected, regardless of the risk nature, and (ii) management of data subjects' privacy risks, which determines the appropriate technical and organizational controls to protect personal data. The following steps must be followed by PIA practitioners:

- Define and document the context of the data processing action under consideration
- Analysis of controls that can protect fundamental principles
- Assessment and management of privacy risks related to data
- Formal documentation and validation of the PIA

The PIA tool assists practitioners in carrying out the actions that were indicated earlier in this paragraph. The evaluation outcome is depicted as a heat map, in which the risks are arranged in a manner that considers both their criticality and likelihood. The CNIL tool supports four levels of severity scales; Negligible, Limited, Significant, Maximum.

SPIA Tool. The Security and Privacy Impact Assessment (SPIA) is a tool developed by the University of Pennsylvania [SPI16] intended to assist organizations in conducting PIA by identifying risk-prone regions and choosing the most appropriate tactics and timetables for risk reduction. This tool concentrates on safeguards while focusing on both security and privacy for the protection of data. The tool has two versions; the first version is an MS Excel file, whereas the second version (SPIA 2.0) is a web-based application. The tool allows organizations to take probability rankings and threat consequences and automatically score risk into categories of High, Significant, Moderate and Low. Additionally, the SPIA Tool is a flexible and adaptable tool that supports various security and privacy threats.

ASPIA Tool. The Automated Privacy and Security Impact Assessment (APSIA) is powered by the use of interdependency graph models and data processing flows used to create a digital reflection of the cyber-physical environment of an organization. The methodology presents an extensible privacy risk scoring system for quantifying the privacy impact triggered by the identified vulnerabilities of the ICT infrastructure of an organization. Indeed, APSIA seeks to bridge the gap between the cyber-risk and privacy risk assessments, which are typically handled as separate management processes. In APSIA, the impact

level is defined as a combination of three components, namely, (a) the level of impact on the fundamental rights and freedoms of the individuals, (b) the scope of impact to the data processing activities, and (c) the type (i.e., sensitivity) of the processed data. However, in ASPIA, the selection of optimal mitigation controls is not considered. Their aim is only to support the decision-makers in making informed decisions during the risk mitigation life cycle.

7.3 Multi-Criteria Risk Assessment & Control Selection Methodologies

In practice, cost and time constraints, feasibility, and other organizational considerations make it impractical to implement all mitigations (also known as security controls) for every threat. Optimizing mitigation selections has been approached by several researchers, who have taken an extensive list of possible mitigations and narrowed it down to just a few that meet specific criteria or goals [LMN19]. The criteria themselves and the analysis methods are the most intriguing dimensions in this area. Based on a grounded theoretical research, authors in [DE16] present a model of information security investment decision-making. Numerous factors, including policy, competitive advantage, financial considerations, quality, compliance, customer expectations, and strategy, have a profound impact on the manner in which organizations take these decisions. Selection of a security control portfolio for a given circumstance requires taking into account several factors, such as an organization's overarching security concerns, the criteria of individual assets in the environment, potential threats, and the quality of controls. Authors in [LMN19], for example, offer a review of the literature that leads to the identification of four criteria: organizational, asset, threat, and control. For instance, in [RDRB11], authors developed a decision support system for assessing the unknown risk an organization faces during a cyber attack as a function of uncertain threat rates, countermeasure costs, and impacts on its assets. The system employs a genetic algorithm to search for the optimal combination of countermeasures, allowing the user to determine the preferred tradeoff between the portfolio cost and the resulting risk. In [GRCC06], a Genetic Algorithm (GA)-based method was developed that enables enterprises to select the lowest-cost security profile with the greatest vulnerability coverage. Authors, in [KSK21] developed a technique that permits the best selection of cybersecurity controls for complex cyber-physical systems (CPSs) that contain other CPSs as components. The technique estimates the overall risk by considering the likelihood and impact values for each of the system's components and analyzing how risk propagates across information and control flow components. Then to discover the optimal set of controls for each component, the approach applies a genetic algorithm workflow.

Multi-criteria decision-making (MCDM) [FGE05], commonly known as multiple-criteria

decision analysis (MCDA), is widely applied in the selection of security portfolios. MCDM is a method for analyzing multiple conflicting criteria and is used to examine problems in which there are several measurements of costs and benefits that may be traded off to arrive at the optimal solution within the limits that have been specified. Fuzzy set theory [Ote14], multi-attribute utility theory (e.g., value functions, knapsack strategy) [SSGG15, PFM+14, FPM+16, SM14], and evolutionary multi-objective optimization (EMO), commonly known as genetic algorithms [KEG+16], are some of the MCDM methodologies being investigated by researchers to address this problem.

There are a lot of risk assessment approaches which consider multi-criteria to calculate risk exposure. In [ZMMM13] risk analysis is modeled as MCDM problem in which experts express their preferences for each risk, over two traditional criteria: probability and impact. A risk-based decision framework [GQP+17] is proposed for cybersecurity strategy prioritization. There are a few approaches that have defined risk impact criteria for different stakeholders. For instance, in the context of cloud computing, in [ASS+14] a security risk assessment framework is proposed that can enable cloud service providers to assess security risks in the cloud computing environment and allow cloud clients with different risk perspectives to contribute to risk assessment. In analyzing the conflict of interest between the risk owner and the risk actors in [RS12] authors proposed conflicting incentives risk analysis (CIRA) method in which risks are modelled in terms of conflicting incentives. The goal of CIRA is to provide an approach in which the input parameters can be audited more easily. In [Wri12], the authors provide a seven-step approach to PIA. They have declared that privacy risk shall be assessed from both data subjects and system perspective. The authors recommend privacy controls that can help to minimize, mitigate, or eliminate the identified risk. Similarly, recently, the authors [IFHÅM19] proposed a privacy risk assessment by considering both perspectives. Their approach is based on the PIA methodology proposed by [Wri12] in the case of mobile health data collection systems, which proposes a systematic identification and evaluation of privacy risks.

7.4 Discussion

In this thesis, we intended to propose a methodology to help controllers in making more informed decision in the control selection process, as they are required in conducting a DPIA. In light of this, we examined the state of the art to learn about cybersecurity risk management methodologies, privacy impact assessment approaches and tools, multi-criteria and control selection approaches. From the studied approaches, we have learned:

(i) As we reviewed (in Section 7.1), classical cybersecurity risk assessment approaches do not consider the risk related to the privacy realm, where a threat or a vulnerability exploit can only affect the three security attributes, namely the Confidentiality (C),

Integrity (I), and Availability (A), of a given asset (also known as the CIA triad). Therefore, these approaches/methodologies disable in defining the impact level of a threat on data subject.

(ii) The GDPR introduces a risk-based approach for determining which technical and organizational measures are appropriate in the given situation [VvdB17]; DPIA is considered to be a risk management tool [VDGR16] which considers risks to the rights and freedoms of data subjects, yet it did not clarify the steps and provided no assistance to the controllers. The UK [ICO18] and French data protection authority [CNI] have also offered tools that show the steps and what to consider, without providing assistance to controllers in which how to evaluate risk-related to data subjects.

(iii) Academics have recommended improvements to the DPIA processes, for instance in [WJ15] used formal modeling techniques for privacy to make DPIA more systematic and structured. This methodology is helpful in detecting threats to privacy, as well as identifying controls that can be implemented to mitigate those concerns. However, it misses assessment steps from a legal perspective and is not integrated in a risk assessment process.

(iv) Lastly, the majority of methodologies at the end of the evaluation leave the risk assessor with a list of control suggestions without any risk trade-off analysis and support in selecting the optimal solution. The existing approaches, regardless of the factors incorporated in the impact assessment that only consider the security risks (do not consider privacy risks and the impact on data subjects), their optimization technique is not the case that we look at in our case. For instance, they [GRCC06, KSK21] used evolutionary algorithms (e.g., genetic algorithms) in finding an optimal solution where according to their inherent features (i.e., using a random generator), they do not guarantee that all optimal solutions (i.e., Pareto optimal) can be enumerated.

There are several aspects with respect to which we can evaluate the tool support for PIA methodologies, including (i) *multi-criteria evaluation*, (ii) *GDPR requirements*, (iii) *cyber risks* in the assessment, (iv) *trade-off analysis*, and (v) identification of *optimal solutions*. These aspects have been identified by reviewing the state-of-the-art, as discussed above. In Table 7.1, we compare these aspects between our approach and existing approaches/tools for performing a privacy risk assessment. Concerning *GDPR requirements* in our methodology, as we discussed earlier in Section 2.2.3, we use the term "data protection goal" (as defined in the SDM model [fD17]) to pave the way to legal compliance considerations based on the outcomes of our proposal.

Our work focuses on the selection of the security and privacy controls in a systematic and principled way while being agnostic with respect to the methodologies used to perform

Table 7.1: Aspects comparison between existing PIA approaches and ours

	Multi-Criteria Evaluation	GDPR Requirements	Cyber Risk	Risk Trade-off Analysis	Optimal Solution
ENISA	-	✓	-	-	-
CNIL	-	✓	-	-	-
GS1	-	-	-	-	-
SPIA	-	-	✓	-	-
ASPIA	-	✓	✓	-	-
[IFHÅM19]	✓	✓	✓	-	-
Ours	✓	✓	✓	✓	✓

PIA. This means that we do not intend to offer yet another methodology or an alternative way to perform the PIAs. Instead, a risk analyst can use whichever PIAs he/she prefers and then integrate our approach to help in the last phase of selecting the security and privacy controls; an activity that the majority of the methodologies neglect.

To achieve this, our main consideration is the identification of the input parameters for our approach and how these are used to identify "optimal" configurations of security and privacy controls to minimize risks in a highly automated way by mapping the problem of identifying security and privacy controls to a multi-objective optimization problem that can be solved by the available solvers off-the-shelf. We also consider the legal aspects and discuss how various input parameters into our approach are related to the legal constraints by using the protection goals in order to make it clear how to use the results of our approach and also for the DPIA mandated by the GDPR or similar legal frameworks.

It is worth to mention that, our proposed methodology considers the availability of a threat list and a control list (as input artifacts), and does not deal with threat identification and mitigation processes. This means our approach is not limited to a defined list of security and privacy threats; on the contrary, it is flexible and can be integrated with other approaches. Below, we enumerate some unique capability of our proposed methodology:

1. Evaluates risk levels from multiple perspectives.
2. Provides a mapping between threats and GDPR's requirements by knowing which data protection goal might affect.
3. Assists data controllers in making informed decisions by facilitating the auditability and traceability of suggested solutions.
4. It is scalable in using large case scenarios.

Chapter 8

Conclusions and Future Work

As technology advances, cyberspace expands, and digitalization increases, it is getting more challenging for businesses to defend themselves against cybersecurity threats. The risks associated with cybersecurity are present in every business, irrespective of its size or sector. It makes businesses consider adopting a cybersecurity risk management strategy that is systematic and disciplined to protect vital information systems and infrastructures. Cybersecurity risk management consists of several steps, including the selection of appropriate mitigation controls to minimize risks. This is a difficult task that requires searching through all possible subsets of a set of available controls and identifying those that minimize the risks of all stakeholders. Conflicting goals may arise due to the fact that different stakeholders may have different perceptions of the risks (especially when considering the impact of threats). This necessitates the need to find the best possible trade-offs among the various needs, such as costs and the expertise that is required to deploy mitigation controls. The ability to tackle this kind of problem is particularly relevant when considering privacy provisions deriving from national or international regulations (such as the General Data Protection Regulation, GDPR) whereby the organization offering a data processing activity should reduce the user's risk to an acceptable level while keeping costs in check and meeting other business goals.

The GDPR requires that data subjects' risks be minimized while also considering other aspects, such as consent best practices and the budget constraints of the involved stakeholders (e.g., the data controller, the data processor, and involved third parties). Additionally, the GDPR requires that a Data Protection Impact Assessment (DPIA) be conducted in order to evaluate the security and privacy measures that have been implemented and to minimize the impact of threats on the rights and freedoms of individuals. This kind of assessment differs from classical risk analyses, in which the actor carrying out the evaluation is also interested in diminishing its risk. Therefore, the data controllers must adhere to an approach, method, or framework that will assist them in making more informed

decisions about which security mechanisms will provide a better trade-off between their requirements and those of the data subjects.

This task—providing a more favorable trade-off between organizations' needs and those of their users—is non-trivial, as it may require to search through a large set of available controls to mitigate the previously identified set of threats. It is important to note that, in an ideal situation, it is not sufficient to identify a solution, i.e., a subset of the available controls that reduces risk to the desired level; instead, it is desirable to identify those subsets that not only minimize risk but also satisfy other criteria, such as cost reduction or the availability of cybersecurity skills to correctly deploy the selected controls. In this context, being able to compute the subsets of controls that minimize the risks of both the organization of the system and its users is a necessary prerequisite to identify the most appropriate configuration of the controls that offer the best possible trade-off among the various objectives.

This thesis provides the following contributions:

1. **Introduced the Multi-Stakeholder Risk Minimization Problem (MSRMP)** to assist in the definition of the best (with respect to all the stakeholders involved in the system) Risk Management Policies (RMPs)—as an appropriate set of security controls to mitigate the identified set of threats—in the fundamental step of selecting mitigations for risk management.

2. **Formalized the MSRMP** as a multi-objective optimization problem that can be solved by using state-of-the-art techniques for Pareto Optimality. On top of such techniques, we proposed a semi-automated approach to define and solve instances of the MSRMP. We also discussed strategies to reduce the large search space resulting from real instances of the MSRMP. We illustrated the main notions of our approach on a simple yet representative running example.

3. **Developed a Tool support** to demonstrate the applicability of the proposed methodology. The implementation of the proposed approach allowed us to perform an experimental evaluation whose results confirmed the practical viability of the proposed approach. For instance, two test cases have been defined to assess the computational time and resources used for increasingly large optimization problem instances.

4. **Validated the proposed methodology** by applying it to assist in performing the DPIA of a contact tracing solution developed in the context of the European project `Trace4Safe`.

8.1 Future Work

New research possibilities and ideas can be investigated through the exploration and development of the contributions of this thesis. Below, we outline some of these research possibilities for future work:

1. First, we plan to further validate the flexibility of our approach by integrating it with a methodology for the risk evaluation of identity proofing solutions introduced in [PSR21]. In that work, the authors present a framework to analyze the risks of enrollment solutions at design time. In particular, they focus on associating security controls with threats deriving from a set of attackers, so to reduce risks at an acceptable level while guaranteeing usability and economy. However, it is left open the problem of determining the optimal set of mitigations, and this is the reason for which the approach presented in this work becomes an interesting complement. As indicated in [GFL+17], performing a usability evaluation on the enrollment and identity proofing process is critical. Due to this, we intend to take another angle into account when selecting the best combination of security controls. Considering the usability factors—e.g., effectiveness, efficiency, and satisfaction—in our proposed approach will open up a new perspective in finding optimal solutions. Hence, in this case, rather than including various stakeholders in the risk assessment process, we will have security and usability dimensions. To make this possible, some tasks must be performed such as (i) defining usability factors, (ii) identifying mitigation controls that may have an impact on usability factors (before and during the enrollment procedures), and (iii) investigating how to quantify the usability risks.

2. The second possibility for future work is to identify a comprehensive baseline of controls (such as the one in the Risk Management Framework of NIST[1]) and provide an approach to tailor it to the use case scenario under consideration in order to lower the barrier of adoption of the approach proposed here by addressing the intricacies of evaluating the trade-offs of security controls including costs and skills required.

3. The third possible line of future work is to investigate how it is possible to smoothly combine the approach proposed in this work with available methodologies for risk management (e.g., STRIDE). Additionally, combining the proposed approach with some existing methodologies that have tried to practically estimate the cost that may be imposed on organizations by law, such as GDPR. For instance, in [SWB21], several types of costs are estimated, such as the costs of implementing privacy measures, the costs of jeopardizing the rights and freedoms of data subjects, the cost of administrative fines, and the cost of compensation granted by courts to data subjects under the GDPR.

[1] `https://csrc.nist.gov/Projects/risk-management/about-rmf/select-step`

Bibliography

[2713] Joint Technical Committee ISO/IEC JTC1. Subcommittee SC 27. Information technology–security techniques–information security management systems–requirements, 2013.

[A-116] Managing information as a strategic resource. `https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/circulars/A130/a130revised.pdf`, 2016.

[Aga15] Sushant Agarwal. Developing a structured metric to measure privacy risk in privacy impact assessments. In *IFIP International Summer School on Privacy and Identity Management*, pages 141–155. Springer, 2015.

[ARP18] Vanessa Ayala-Rivera and Liliana Pasquale. The grace period has ended: An approach to operationalize gdpr requirements. In *2018 IEEE 26th International Requirements Engineering Conference (RE)*, pages 136–146. IEEE, 2018.

[AS17] Majed Alshammari and Andrew Simpson. Towards a principled approach for engineering privacy by design. In *Annual Privacy Forum*, pages 161–177. Springer, 2017.

[ASRJ18] Amir Shayan Ahmadian, Daniel Strüber, Volker Riediger, and Jan Jürjens. Supporting privacy impact assessment by model-based privacy analysis. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, pages 1467–1474, 2018.

[ASS+14] Sameer Hasan Albakri, Bharanidharan Shanmugam, Ganthan Narayana Samy, Norbik Bashah Idris, and Azuan Ahmed. Security risk assessment framework for cloud computing environments. *Security and Communication Networks*, 7(11):2114–2124, 2014.

[BBG+17] Sean Brooks, Sean Brooks, Michael Garcia, Naomi Lefkovitz, Suzanne Lightman, and Ellen Nadeau. *An introduction to privacy engineering and risk management in federal systems*. US Department of Commerce, National Institute of Standards and Technology, 2017.

[BCH15] Nigel Bevan, James Carter, and Susan Harker. Iso 9241-11 revised: What have we learnt about usability since 1998? In *International conference on human-computer interaction*, pages 143–151. Springer, 2015.

[BFH+16] Felix Bieker, Michael Friedewald, Marit Hansen, Hannah Obersteller, and Martin Rost. A process for data protection impact assessment under the European general data protection regulation. In *Annual Privacy Forum*, pages 21–37. Springer, 2016.

[BG19] Tamas Bisztray and Nils Gruschka. Privacy impact assessment: comparing methodologies with a focus on practicality. In *Nordic Conference on Secure IT Systems*, pages 3–19. Springer, 2019.

[BL+20] Kaitlin R Boeckl, Naomi B Lefkovitz, et al. Nist privacy framework: A tool for improving privacy through enterprise risk management, version 1.0, 2020.

[Boa20] European Data Protection Board. Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the covid-19 outbreak, 2020.

[BSI17] Data protection-specification for a personal information management system. Available at: `https://www.bsigroup.com/en-GB/BS-10012-Personal-information-management/`, 2017.

[CF12] Pietro Colombo and Elena Ferrari. Towards a modeling and analysis framework for privacy-aware systems. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, pages 81–90. IEEE, 2012.

[Cla09] Roger Clarke. Privacy impact assessment: Its origins and development. *Computer law & security review*, 25(2):123–135, 2009.

[CLRS01] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. 35.5: The subset-sum problem, 2001.

[CNI] CNIL (Commission Nationale de l'Informatique et des Libertés). The open source pia software helps to carry out data protection impact assesment. `https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assesment`. Accessed: February 2020.

[CNI12] CNIL (Commission Nationale de l'Informatique et des Libertés). Methodology for privacy risk management: How to implement the data protection act. `https://www.cnil.fr/sites/default/files/typo/document/CNIL-ManagingPrivacyRisks-Methodology.pdf`, 2012.

[CNI18] CNIL (Commission Nationale de l'Informatique et des Libertés). Privacy risk assessment (PIA). `https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf`, 2018.

[CNS15] Committee on national security systems instruction no. 4009. `https://www.cnss.gov/CNSS/issuances/Instructions.cfm`, 2015.

[DE16] Daniel Dor and Yuval Elovici. A model of the information security investment decision-making process. *Computers & security*, 63:1–13, 2016.

[Dir95] EU Directive. Directive 95/46/ec of the european parliament and of the council of 24 october 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Communities*, 38(281):31–50, 1995.

[DLM16] Sourya Joyee De and Daniel Le Métayer. Priam: a privacy risk analysis methodology. In *Data Privacy Management and Security Assurance*, pages 221–229. Springer, 2016.

[DLM17] Sourya Joyee De and Daniel Le Métayer. A refinement approach for the reuse of privacy risk analysis results. In *Annual Privacy Forum*, pages 52–83. Springer, 2017.

[DP16] Giuseppe D'Acquisto and Georgia Panagopoulou. Guidelines for smes on the security of personal data processing, 2016.

[ENI20] Evaluating the level of risk for a personal data processing operation. `https://www.enisa.europa.eu/risk-level-tool/risk`, 2020.

[fD17] Unabhängiges Landeszentrum für Datenschutz. The standard data protection model: A concept for inspection and consultation on the basis of unified protection goals, 2017.

[FGE05] J Figueira, S Greco, and M Ehrgott. State of the art surveys. In *Multiple Criteria Decision Analysis*. Springer, 2005.

[FJ14a] Jack Freund and Jack Jones. *Measuring and managing information risk: a FAIR approach*. Butterworth-Heinemann, 2014.

[FJ14b] Jack Freund and Jack Jones. *Measuring and managing information risk: a FAIR approach*. Butterworth-Heinemann, 2014.

[FPM+16] Andrew Fielder, Emmanouil Panaousis, Pasquale Malacaria, Chris Hankin, and Fabrizio Smeraldi. Decision support approaches for cyber security investment. *Decision support systems*, 86:13–23, 2016.

[GAA02] Stoneburner Gary, Goguen Alice, and Feringa Alexis. Risk management guide for information technology systems. *Special Publication*, pages 800–300, 2002.

[GFL+17] Paul Grassi, James Fenton, Naomi Lefkovitz, Jamie Danker, Yee-Yin Choong, Kristen Greene, and Mary Theofanos. Digital identity guidelines: Enrollment and identity proofing, 2017.

[GQP+17] Alexander A Ganin, Phuoc Quach, Mahesh Panwar, Zachary A Collier, Jeffrey M Keisler, Dayton Marchese, and Igor Linkov. Multicriteria decision framework for cybersecurity risk assessment and management. *Risk Analysis*, 2017.

[GRCC06] Mukul Gupta, Jackie Rees, Alok Chaturvedi, and Jie Chi. Matching information security vulnerabilities to organizational security profiles: a genetic algorithm approach. *Decision Support Systems*, 41(3):592–603, 2006.

[GS115] Gs1. epc/rfid privacy impact assessment tool. `https://www.gs1.org/standards/epc-rfid/pia`, 2015. Accessed: January 2021.

[HN+15] Michael D Hogan, Elaine M Newton, et al. Supplemental information for the interagency report on strategic us government engagement in international standardization to achieve us objectives for cybersecurity, 2015.

[ICO18] ICO (Information Commission's Office). Data protection impact assessments. `https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias-1-0.pdf`, 2018. Accessed: June 2019.

[ICOI18] Information Commission's Office (ICO). Data protection impact assessments. `https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias-1-0.pdf`, 2018. (accessed on 6 June 2019.

[IFHÅM19] Leonardo Horn Iwaya, Simone Fischer-Hübner, Rose-Mharie Åhlfeldt, and Leonardo A Martucci. Mobile health systems for community-based primary care: Identifying controls and mitigating privacy threats. *JMIR mHealth and uHealth*, 7(3):e11642, 2019.

[ISO05] Information technology - security techniques - code of practice for information security management., 2005.

[ISO14] International Organization for Standardization (ISO). iso/iec 27018: 2014-information technology–security techniques–code of practice for protection of

personally identifiable information (pii) in public clouds acting as pii processors. Available online: https://www.iso.org/obp/ui/#iso:std:iso-iec:27018:ed-1:v1:en, 2014.

[ISO17a] International Organization for Standardization (ISO). iso/iec 29134: 2017-information technology–security techniques–guidelines for privacy impact assessment. Available online: https://www.iso.org/obp/ui/#iso:std:iso-iec:29134:ed-1:v1:en, 2017.

[ISO17b] International Organization for Standardization (ISO). iso/iec 29151:2017-information technology—security techniques—code of practice for personally identifiable information protection. Available online: https://www.iso.org/obp/ui/#iso:std:iso-iec:29151:ed-1:v1:en, 2017.

[KEG+16] Elmar Kiesling, Andreas Ekelhart, Bernhard Grill, Christine Strauss, and Christian Stummer. Selecting security control portfolios: a multi-objective simulation-optimization approach. *EURO Journal on Decision Processes*, 4(1-2):85–117, 2016.

[Kla09] Kathrin Klamroth. Discrete multiobjective optimization. In *Evolutionary Multi-Criterion Optimization*, pages 4–4. LNCS 5467, Springer, 2009.

[KSK21] Georgios Kavallieratos, Georgios Spathoulas, and Sokratis Katsikas. Cyber risk propagation and optimal selection of cybersecurity controls for complex cyber-physical systems. *Sensors*, 21(5):1691, 2021.

[LMN19] Thomas Llansó, Martha McNeil, and Cherie Noteboom. Multi-criteria selection of capability-based cybersecurity solutions. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

[LSS10] Mass Soldal Lund, Bjørnar Solhaug, and Ketil Stølen. *Model-driven risk analysis: the CORAS approach*. Springer Science & Business Media, 2010.

[MA04] R Timothy Marler and Jasbir S Arora. Survey of multi-objective optimization methods for engineering. *Structural and multidisciplinary optimization*, 26(6):369–395, 2004.

[MH15] Rene Meis and Maritta Heisel. Supporting privacy impact assessments using problem-based privacy analysis. In *ICSOFT*, pages 79–98. Springer, 2015.

[MKW+16] Stephen McLaughlin, Charalambos Konstantinou, Xueyang Wang, Lucas Davi, Ahmad-Reza Sadeghi, Michail Maniatakos, and Ramesh Karri. The cybersecurity landscape in industrial control systems. *Proceedings of the IEEE*, 104(5):1039–1057, 2016.

[MR22] Majid Mollaeefar and Silvio Ranise. Identifying and quantifying trade-offs in multi-stakeholder risk evaluation with applications to the data protection impact assessment of the gdpr, 2022.

[MSR20] Majid Mollaeefar, Alberto Siena, and Silvio Ranise. Multi-stakeholder cybersecurity risk assessment for data protection. In *Proceedings of the 17th International Joint Conference on e-Business and Telecommunications - Volume 3: SECRYPT*, pages 349–356. INSTICC, SciTePress, 2020.

[NIS12] SP NIST. 800-30, revision 1. *Guide for Conducting Risk Assessments*, 2012.

[oCS10] Treasury Board of Canada Secretariat. Directive of privacy impact assessments. `https://www.isc.upenn.edu/sites/default/files/introduction_to_spia_program.pdf`, 2010. (accessed on 29 December 2020).

[OS14] Marie Caroline Oetzel and Sarah Spiekermann. A systematic methodology for privacy impact assessments: a design science approach. *European Journal of Information Systems*, 23(2):126–150, 2014.

[OSG+11] Marie Caroline Oetzel, Sarah Spiekermann, Ingrid Grüning, Harald Kelter, and Sabine Mull. Privacy impact assessment guideline for rfid applications. *German Federal Office for Information Security (BSI)*, 2011.

[Ote14] Angel Rafael Otero. An information security control assessment methodology for organizations, 2014.

[PFM+14] Emmanouil Panaousis, Andrew Fielder, Pasquale Malacaria, Chris Hankin, and Fabrizio Smeraldi. Cybersecurity games and investments: A decision support approach. In *International Conference on Decision and Game Theory for Security*, pages 266–286. Springer, 2014.

[PMGG21] Dimitrios Papamartzivanos, Sofia Anna Menesidou, Panagiotis Gouvas, and Thanassis Giannetsos. A perfect match: Converging and automating privacy and security impact assessment on-the-fly. *Future Internet*, 13(2):30, 2021.

[PSR21] M. Pernpruner, G. Sciarretta, and S. Ranise. A framework for security and risk analysis of enrollment procedures: Application to fully-remote solutions based on edocuments. In *Proceedings of the 18th International Joint Conference on e-Business and Telecommunications - Volume 3: SECRYPT*, pages 222–233. INSTICC, SciTePress, 2021.

[Pur10] Grant Purdy. Iso 31000: 2009—setting a new standard for risk management. *Risk Analysis: An International Journal*, 30(6):881–886, 2010.

[QJD+19] Qais Saif Qassim, Norziana Jamil, Maslina Daud, Ahmed Patel, and Norhamadi Ja'affar. A review of security assessment methodologies in industrial control systems. *Information & Computer Security*, 2019.

[R+11] Ronald S Ross et al. Managing information security risk: Organization, mission, and information system view. `https://doi.org/10.6028/NIST.SP.800-39`, 2011.

[RDRB11] Loren Paul Rees, Jason K Deane, Terry R Rakes, and Wade H Baker. Decision support for cybersecurity risk planning. *Decision Support Systems*, 51(3):493–505, 2011.

[reg16] Regulation (eu) 2016/679 of the EUROPEAN parliament and of the council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN`, 2016.

[RMO16] Ron Ross, Michael McEvilley, and Janet Oren. Systems security engineering: Considerations for a multidisciplinary approach in the engineering of trustworthy secure systems, 2016.

[RS12] Lisa Rajbhandari and Einar Snekkenes. Intended actions: Risk is conflicting incentives. In *International Conference on Information Security*, pages 370–386. Springer, 2012.

[SBP+10] M Swanson, P Bowen, AW Phillips, D Gallup, and D Lynes. Nist special publication 800-34, rev. 1, contingency planning guide for federal information systems, 2010.

[Sch03] Alexander Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer Science & Business Media, 2003.

[SCO+18] Nataliya Shevchenko, Timothy A Chick, Paige O'Riordan, Thomas P Scanlon, and Carol Woody. Threat modeling: a summary of available methods. Technical report, Carnegie Mellon University Software Engineering Institute Pittsburgh United,, 2018.

[SD94] Nidamarthi Srinivas and Kalyanmoy Deb. Muiltiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary computation*, 2(3):221–248, 1994.

[Sho14] Adam Shostack. *Threat modeling: Designing for security.* John Wiley & Sons, 2014.

[SLE05] Paul Saitta, B Larcom, and M Eddington. Trike v1 methodology document. *Draft, work in progress*, 2005.

[SM14] Fabrizio Smeraldi and Pasquale Malacaria. How to spend it: optimal investment for cyber security. In *Proceedings of the 1st International Workshop on Agents and CyberSecurity*, pages 1–4, 2014.

[SP812] Joint task force transformation initiative guide for conducting risk assessments. `https://doi.org/10.6028/NIST.SP.800-30r1`, 2012.

[SPI16] Introduction to the spia program. `https://www.isc.upenn.edu/sites/default/files/introduction_to_spia_program.pdf`, 2016. Accessed: December 2020.

[SSGG15] Maryam Shahpasand, Mehdi Shajari, Seyed Alireza Hashemi Golpaygani, and Hoda Ghavamipoor. A comprehensive security control selection model for interdependent organizational assets structure. *Information & Computer Security*, 2015.

[SWB21] Annika Selzer, Daniel Woods, and Rainer Bohme. An economic analysis of appropriateness under article 32 gdpr. *Eur. Data Prot. L. Rev.*, 7:456, 2021.

[T+05] Nancy R Tague et al. *The quality toolbox*, volume 600. ASQ Quality Press Milwaukee, 2005.

[UM15] Tony UcedaVelez and Marco M Morana. Risk centric threat modeling: Process for attack simulation and threat analysis, 2015.

[VDGR16] Niels Van Dijk, Raphaël Gellert, and Kjetil Rommetveit. A risk to a right? Beyond data protection risk assessments. *Computer Law & Security Review*, 32(2):286–306, 2016.

[VK18] Konstantina Vemou and Maria Karyda. An evaluation framework for privacy impact assessment methods. *MCIS*, page 5, 2018.

[vPH17] JPM van Puijenbroek and J-H Hoepman. Privacy impact assessments in practice: Outcome of a descriptive field research in the Netherlands, 2017.

[VvdB17] Paul Voigt and Axel von dem Bussche. *Practical Implementation of the Requirements Under the GDPR. In: The EU General Data Protection Regulation (GDPR)*. Springer, 2017.

[WB18] Isabel Wagner and Eerke Boiten. Privacy risk assessment: from art to science, by metrics. In *Data Privacy Management, Cryptocurrencies and Blockchain Technology*, pages 225–241. Springer, 2018.

[WJ15] Kim Wuyts and Wouter Joosen. Linddun privacy threat modeling: a tutorial. `https://www.linddun.org/linddun`, 2015.

[Wri12] David Wright. The state of the art in privacy impact assessment. *Computer law & security review*, 28(1):54–61, 2012.

[WWLC20] Yu-Chih Wei, Wei-Chen Wu, Gu-Hsin Lai, and Ya-Chi Chu. pisra: privacy considered information security risk assessment model. *The Journal of Supercomputing*, 76(3):1468–1481, 2020.

[ZH11] Harald Zwingelberg and Marit Hansen. Privacy protection goals and their implications for eid systems. In *IFIP PrimeLife International Summer School on Privacy and Identity Management for Life*, pages 245–260. Springer, 2011.

[ZMMM13] Yeleny Zulueta, Vladimir Martell, Juan Martínez, and Luis Martínez. A dynamic multi-expert multi-criteria decision making model for risk analysis. In *Mexican International Conference on Artificial Intelligence*, pages 132–143. Springer, 2013.

Appendix A

Glossary

Risk [A-116]: A measure of the extent to which an entity is threatened by a potential circumstance or event, and typically is a function of: (i) the adverse impact, or magnitude of harm, that would arise if the circumstance or event occurs; and (ii) the likelihood of occurrence.

Risk assessment [R+11]: The process of identifying risks to organizational operations (including mission, functions, image, reputation), organizational assets, individuals, other organizations, and the Nation, resulting from the operation of a system.

Risk management [A-116]: The program and supporting processes to manage risk to agency operations (including mission, functions, image, reputation), agency assets, individuals, other organizations, and the Nation, and includes: establishing the context for risk-related activities; assessing risk; responding to risk once determined; and monitoring risk over time.

Risk mitigation [CNS15]: Prioritizing, evaluating, and implementing the appropriate risk-reducing controls/countermeasures recommended from the risk management process.

Security [CNS15]: A condition that results from the establishment and maintenance of protective measures that enable an organization to perform its mission or critical functions despite risks posed by threats to its use of systems. Protective measures may involve a combination of deterrence, avoidance, prevention, detection, recovery, and correction that should form part of the organization's risk management approach.

Security control [A-116]: The safeguards or countermeasures prescribed for an information system or an organization to protect the confidentiality, integrity, and availability of the system and its information.

Threat [SP812]: Any circumstance or event with the potential to adversely impact orga-

nizational operations, organizational assets, individuals, other organizations, or the Nation through a system via unauthorized access, destruction, disclosure, modification of information, and/or denial of service.

Threat assessment [A-116]: Formal description and evaluation of threat to an information system.

Threat Event [ISO05]: It is a potential cause of an unwanted incident, which may result in harm to a system or organization.

Vulnerability [SP812]: Weakness in an information system, system security procedures, internal controls, or implementation that could be exploited or triggered by a threat source.

Personally identifiable information(PII) [ISO05]: Any information that (a) can be used to establish a link between the information and the natural person to whom such information relates, or (b) is or can be directly or indirectly linked to a natural person.

Impact [SBP+10]: The magnitude of harm that can be expected to result from the consequences of unauthorized disclosure of information, unauthorized modification of information, unauthorized destruction of information, or loss of information or information system availability.

Likelihood of occurrence [SP812]: A weighted factor based on a subjective analysis of the probability that a given threat is capable of exploiting a given vulnerability or a set of vulnerabilities.

Usability [BCH15]: Extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

Appendix B

The Standard Data Protection Model

The GDPR lays down rules on the protection of natural persons with regard to the processing of personal data and protects the fundamental rights and freedoms of natural persons, in particular their right of protection of personal data. Fundamental requirements on the security of processing personal data are provided in Articles 5, 12, 25 and 32 GDPR. The GDPR calls for appropriate technical and organizational measures to adequately reduce the risks to the rights and freedoms of natural persons. This concerns both measures to safeguard the rights of data subjects (Chapter III GDPR) and measures to implement data protection principles (Art. 25 para. 1 GDPR), including Data Minimization (Art. 25 para. 2 GDPR) and ensuring the security of processing (Art. 32 para. 1). The principle of data protection by design and by default (Art. 25 GDPR) calls for the controller to address data protection requirements at a very early stage in the planning of processing operations. The GDPR requires a process for regular testing, assessment and evaluation of the effectiveness of technical and organizational measures (Art. 24 para. 1 sentence 2, Art. 32 para. 1 sentence 1 lit. d GDPR). Finally, the GDPR provides a consistency mechanism that integrates the independent supervisory bodies in a complex consultation procedure (Chapter VII GDPR – Cooperation and Consistency). Especially this process requires a coordinated, transparent and verifiable system to assess the processing of personal data with regard to data protection.

The Standard Data Protection Model (SDM) [fD17] offers suitable measures to understand and interpret GDPR legal requirements into qualified organizational and technical measures. To this end, the SDM first documents the legal requirements of the GDPR and then assigns them to the protection goals *Data Minimization, Availability, Integrity, Confidentiality, Transparency, Unlinkability* and *Intervenability*. The SDM thus transposes the legal requirements of the GDPR on protection goals into the technical and organizational measures required by the Regulation, which are described in detail in the SDM's catalog of reference measures. It thus supports the transformation of abstract legal requirements

into concrete technical and organizational measures. The authors in [fD17] have provided the following table (Table B.1), in which all data protection requirements of the GDPR (in column 2) are mapped to at least one protection goal (in column 3).

Table B.1: Mappings between protection goals and the GDPR's requirements.

ID	Requirement of the GDPR	Protection Goal
R_1	Transparency for data subjects (Art. 5 para. 1 lit a, Art. 12 para. 1 and 3 to Art. 15, Art. 34)	Transparency
R_2	Purpose limitation (Art. 5 para. 1 lit. c)	Unlinkability
R_3	Data minimization (Art. 5 para. 1 lit. c)	Data Minimization
R_4	Accuracy (Art. 5 para. 1 lit. d)	Integrity
R_5	Storage limitation (Art. 5 para. 1 lit. e)	Data Minimization
R_6	Integrity (Art. 5 para. 1 lit. f GDPR, Art. 32 para. 1 lit. b)	Integrity
R_7	Confidentiality (Art. 5 para. 1 lit. f, Art. 28 para. 3 lit. b, Art. 29, Art. 32 para. 1 lit. b, Art. 32 para. 4, Art. 38 para. 5)	Confidentiality
R_8	Accountability and Verifiability (Art. 5 para. 2, Art. 7 para. 1, Art. 24 para. 1, Art. 28 para. 3 lit. a, Art. 30, Art. 33 para. 5, Art. 35, Art. 58 par. 1 lit. a and lit. e)	Transparency
R_9	Support in exercising data subjects' rights (Art. 12 para. 2)	Intervenability
R_{10}	Identification and Authentication (Art. 12 para. 6)	Intervenability
R_{11}	Rectification of data (Art. 5 lit. d, Art. 16)	Intervenability
R_{12}	Erasure of Data (Art. 17 para. 1)	Intervenability
R_{13}	Restriction of data processing (Art. 18)	Intervenability
R_{14}	Data portability (Art. 20, para 1)	Intervenability
R_{15}	Possibility to intervene in processes of automated decisions (Art. 22 para 3)	Intervenability
R_{16}	Freedom from error and discrimination in profiling (Art. 22 para 3, 4 in connection with recital 71)	Integrity
R_{17}	Data protection-friendly default settings (Art. 25 para 2)	Data Minimization, Intervenability
R_{18}	Availability (Art. 32 para 1 lit. b)	Availability
R_{19}	Availability (Art. 32 para 1 lit. b)	Availability, Integrity, Confidentiality
R_{20}	Restorability (Art. 32 para 1 lit. b, lit. c)	Availability
R_{21}	Remedy and mitigation of data protection breaches (Art. 33, para 3 lit. d, Art. 34 para 2)	Integrity, Intervenability, Availability, Confidentiality
R_{22}	Adequate monitoring of the processing (Art. 32, 33, 34)	Transparency, Integrity
R_{23}	Consent management (Art. 4 No. 11, Art. 7 and 4)	Transparency, Intervenability
R_{24}	Implementation of supervisory orders (Art. 58 para 2 lit. f and lit. j)	Intervenability

Appendix C

Trace4Safe Project

C.1 The Questionnaire

Table C.1 provides a list of relevant privacy questions (column 2) associated with the Trace4Safe system, and we linked these questions to some privacy principles and targets (column 5). The privacy principles and targets derived from the literature are listed in the following (Table C.2). The questions are assigned to the relevant partner (column 3), and their responses are reported in column 4.

Table C.1: Trac4Safe Scenario: Questionnaire shared with other involved partners.

Stages	Questions	Asked from?	Answer	Privacy Principles & Targets
Registration	**Which data about the worker are collected in the registration phase?**	Thinkinside	- name, surname of the worker - role/group within the organization - health info (in the case of having a pacemaker)	Data quality (Transparency, data minimisation)
	Who can access these data?	Thinkinside	Only pre-defined users from the end customer with a dedicated account can access the data. This in addition to Thinkinside (or other) for system management.	Security of data
	How long these data will be kept?	Thinkinside	Untill the service is dismissed or the employee leaves the end customer company.	Data quality (Limited storage, data minimisation)
	Which are the security properties guaranted by the channel between the user interface and system? HTTPS?	ThinkInside	HTTPS	Security of data
	How does COVID safety responsible person authenticate in the system? What are his/her authorization level?	ThinkInside	Through a login/password. He has a "user" authorization level.	Security of data

Stages	Questions	Asked from?	Answer	Privacy Principles & Targets
Working Shifts	**Which data are collected by the Token?**	Thinkinside	- location (when contact tracing is through RTLS) - contacts among workers	Data quality (Transparency, data minimisation)
	How are these data stored in the system? (plain, encrypted, anonymized)	ThinkInside	Plain and anonymised	Security of data
	How long these data will be kept?	ThinkInside	Can be configured, but the idea is that (unless used for other purposes) the data (location and contacts) is deleted after a pre-defined amount of time.	Data quality (Limited storage, data minimisation)
	What is a contact? TokenId? Name?	Thinkinside	Timestamp, Token ID 1, Token ID2, location (if RTLS)	Security of data
	Is the tokenID (pseudo-)randomly generated? How?	Thinkinside	Incremental number, it is not randomly generated.	Security of data
	Does the tokenID change over the time or not? If yes, When? How?	ThinkInside	No	Security of data
	What kind of information is stored in the token?	Thinkinside	Only in the case of P2P approach, contacts will be stored on the Token	Data quality (data minimisation)
	Which are the security properties guaranted by the channel between tokens and the system? (in the case of RTLS)	Thinkinside	The channel communication can be secured (encrypted) if needed (at least when using Quuppa technology)	Security of data
	Which are the security properties guaranted by the channel between tokens and gateways? (in the case of P2P)	Thinkinside	Asking to vendor.	Security of data
	How can workers inform the system (or COVID safety responsible person/ team leader) in the case of wrongly happened contacts? for example, in the case of losing/dropping the token somewhere	Thinkinside	This should be part of the internal procedures for managing the Tokens. Many exceptions can indeed happen (token lost, forgotten at home, etc.). All this cases should properly documented and adressed systematically by the end customer.	Intervenability
	Do tokens perform mutual authentication among themselves and gateways / central server? (P2P approach)	Thinkinside	To be investigated from the vendors.	Security of data
Reporting by users	**What kind of data is going to collect in this phase?**	Thinkinside	- confirmation of a positive test - date when the symptoms started	Data quality (data minimisation)
	What is the communication channel in order to report?	Thinkinside	Depends on the specific internal procedures of the end customer. This is something we do not cover.	Security of data
	How do workers at risk get inform in the system?	Thinkinside	Depends on the specific internal procedures of the end customer. This is something we do not cover.	Data quality (Transparency), Security of data
	What is the communication channel to inform the workers at risk in the case of a reported positive case?	Thinkinside	Depends on the specific internal procedures of the end customer. This is something we do not cover.	Security of data
	Does the mathematical algorithm (contact tracing simulation algorithm) run on an FBK machine? or it would be part of the product?	FBK-MobS	Yes, for this moment it will be run on the FBK machine.	Data quality (Transparency), Intervenability, Security of data
	What are the input and output of the contact tracing simulation algorithm? Is it possible to provide different inputs so as to change the privacy level? (e.g., do not require the role)	FBK-MobS	Data as input for each data record contain timestamps, token IDs (the two token IDs involved in contact) for a contact, and some optional info about the rooms. Also, the role could be useful to reschedule the job turns but it is not essential.	Security of data
System feedback	**How do workers get informed by the system in the case of any feedback concerning individual behaviour?**	Univ. of Helsinki	Sending messages through email, or sms.	Data quality (Transparency), Security of data

C.2 Privacy Principles and Targets

The following table presents a list of privacy principles and targets derived from the GDPR and originally conceived by Oetzel and Spiekermann [OS14, OSG+11].

Table C.2: Privacy principles and targets derived from [OS14, OSG+11].

Privacy Principles	Privacy Targets
1-Data quality	Ensuring processing in a lawful, fair, and transparent manner
	Ensuring processing only for legitimate purposes
	Providing purpose specification
	Ensuring limited processing for specified purpose
	Ensuring data avoidance
	Ensuring data minimisation
	Ensuring data quality, accuracy and integrity
	Ensuring fair and lawful processing through transparency
2 - Processing legitimacy	Ensuring legitimacy of personal data processing
	Ensuring legitimacy of sensitive personal data processing
3 - Information right of data subject (ex ante transparency)	Providing adequate information in cases of direct collection of data from the data subject
	Providing adequate information where data has not been obtained directly from the data subject (e.g., from third parties)
4 - Access right of data subject (ex post transparency)	Facilitating the provision of information about processed data and purpose
	Facilitating the provision of an (electronic) copy of data
5 - Intervenability	Facilitating the rectification, erasure or blocking of data
	Facilitating the portability of data
	Facilitating the notification to third parties about rectification, erasure and blocking of data
	Providing the ability to withdraw consent
6 - Data subject's right to object	Facilitating the objection to the processing of personal data
	Facilitating the objection to direct marketing activities
	Facilitating the objection to disclosure of data to third parties
	Facilitating the objection to decisions that are solely based on automated processing of data
	Facilitating the data subject's right to dispute the correctness of machine conclusions
7 - Security of data	Ensuring the confidentiality, integrity and availability of personal data storage, processing and transmission
	Ensuring the detection of personal data breaches and their communication to data subjects
8 - Accountability	Ensuring the accountability of personal data storage, processing and transmission

C.3 Security and Privacy Threats

The identified threat scenarios along with their type (security or privacy) and the consequences are reported in Table C.3. Each of these threats are mapped to the corresponding affected components/channels reported in Figure 6.2.

Table C.3: Trac4Safe Scenario: The identified threats along with their consequences.

Threat Scenarios	Channels & Components	Type (SEC/PR)	Consequences
T1- An adversary equipped with a Bluetooth Beacon Tracker can observe tokens, and in the case of token IDs do not change over the time, the attacker can re-identify token holders.	1, 2, 3, 4	PR-Identifiability, Detectability	Tracking users, identifying users, profiling users' behavior, learning about places
T2- An attacker can eavesdrop the network traffic by setting up her device close to the gateways when data is uploading on gateways.	3	SEC-Confidentiality PR-Identifiability, Detectability	Identifying users
T3- Tampering data may happen in different phases of data exchanging in the system.	2, 3, 4 Communication channels between edge server and cloud server	SEC-Tampering data	Loss of data integrity
T4- An unauthorized access to the local stores.	5, 6	SEC-Unauthorized access	Disclosure information
T5- An unauthorized access to the data/application.	7,8	SEC-Unauthorized access	Disclosure information, Impact on contact tracing network
T6- A malicious user tries to submit the same data more than once to maliciously impact the protocol execution.	3, 2	SEC-Replay attack	Impact on contact tracing network
T7- An adversary can set up his/her proximity tracking device, which is equipped with a sensitive antenna and powerful transmitter in a crowded space to increase the range of his/her Bluetooth contacts artificially. Consequently, other tokens consider it as a real contact due to feeling the proximity is in the defined range.	1, 3	SEC- False-positive contacts	Impact on contact tracing network, receiving wrongly notifications (RTLS)
T8- An attacker can make a denial of service to the gateways by sending massive contact messages or sending fake contact messages to impact on constructing the network of contacts which result in the wrong tracing contacts.	Gateways	SEC- Denial of service	Impact on contact tracing network
T9- The data is stored for longer and it increases the chance of data abuse and decreases its security.	5, 6	PR- Data longevity (unlimited data storage)	User identification

T10- Identifying an entity from a set of collected data, e.g., in our case, identifying positive cases.	5, 6	PR- Identifiability, Linkability	Re-identify users
T11- An adversary can drain users' device battery by sending fake contacts messages which the victim device assumes as real contacts. The P2P approach may be more vulnerable to this type of attack.	-	SEC-Denial of service	Impact on contact tracing network
T12- Users' information is shared with a third party or submitted to the health authority without their explicit consent.	-	PR- Policy and consent non-compliance	Non-compliance with the law
T13- Lack of sufficient and complete description of the service and the operation details (such as data flows, data storage location, transmission methods, etc.) and their impacts on users' data.	-	PR- Lack of transparency	Non-compliance with the law
T14- Users do not have the possibility to submit correction requests (in the case of wrongly recorded contacts) that need to be evaluated by the system administrator. There is no implemented procedure in the system to allow the users to notify the system administrator to rectify, erase, or block the wrong registered contacts. For instance, in undefined events, if wrong contacts are uploaded (registered) in the system, it causes the contact tracing network to be created wrongly and result in incorrect notifications.	-	PR- lack of control and inability to rectify or erase the wrong registered contacts	Loss of trust in the system, impact on contact tracing network
T15-An attacker tries to steal or make a copy of a worker's token. In this scenario, the attacker spoofs the token's identity in order to impact the contact's network.	-	SEC- Spoofing identity	Impact on contact tracing network

C.4 Risk Analysis (Aversion Level Estimation)

This section evaluates each of the selected threats (see Table 6.7) in order to determine their aversion levels. Thus, each threat will be challenged by its potential damage to stakeholders' protection criteria, and for that, we ask the question "what would the impact if threat...?". As we described in Chapter 6 (see Section 6.4.1) and also can be seen in the following table, we considered two protection criteria for the organization (i.e., the data controller), namely: financial and reputational situations, and three protection criteria for the employees (i.e., the data subjects), namely: Health condition, Individual freedom, and Social situation. Table C.4 outlines the approach used to evaluate the aversion level of each threat on the protection criteria.

Table C.4: Stakeholders' protection criteria and impact levels.

What could be impacted on the protection criteria (for each perspective) if the threat happens?				
Organization		Employee		
Financial situation	Reputational situation	Health condition	Individual freedom	Social situation
0= No impact.				
1= Low. The impact of any loss or damage is limited and calculable.				
2= Medium. The impact of any loss or damage is considerable.				
3= High. The impact of any loss or damage is significant.				
4= Catastrophic. The impact of any loss or damage is devastating.				

We give an in-depth analysis of the potential impact associated with the identified threats in Table 6.7 in the following. To accomplish so, we expand our explanation on each threat regarding:

- "What if...?" – The potential damages to the organization and employees can be anticipated.
- Impact Level – The estimated aversion level (e.g., no, low, medium, high, and catastrophic) for the different perspectives.

T1: Tracking and Eavesdropping

If encryption and anonymization measures are not implemented, or if random IDs are not generated and do not change over the time, the likelihood of tracking and eavesdropping threats increases, and the associated parties may face the following consequences:

- The organization's financial situation could be severely harmed, and the Trace4safe deployment effort could fail because employees may lose trust in the system.

- The organization's reputation will severely damage and may lead to losing its employees' trust, and as a consequence, the employee may not use/wear the Token device.

- The health condition of employees is not affected by the lack of mitigation controls for this threat.

- The individual freedom of the employees can be catastrophically affected in the case of tracking their contacts which may lead to losing their freedom.

- The social situation of the employees can be significantly affected due to tracking and identifying users (in particular positive cases) that may lead to discrimination or social pressure.

Aversion-Level of T1: High, High, No, Catastrophic, High

T2: Data Tampering

Security measures such as encrypting data-at-rest and data-in-transit by using encrypted connections such as SSL, TSL, HTTPS, etc., will protect data integrity. Accidental or deliberate attacks on the system can cause to impact data integrity, and the associated parties may face the following consequences:

- The organization's financial situation can be severely affected depending on the consequences, in particular, the rate of infection (the number of positive cases in the system) due to data tampering which the organization may even temporarily lose some employees due to quarantine regulations or in worse case the organization may decide to close some businesses or closing down some premises.

- The organization's reputation can be considerably affected depending on the consequences and how serious the data tampering is.

- The health condition of employees can be seriously affected since data tampering directly impacts the employees' health condition.

- The individual freedom of the employees will not be impacted due to data tampering attacks.

- The social situation of the employees can be considerably affected if their data is distorted (accidentally or deliberately), for instance, being discriminated against by an unfavorable health condition.

Aversion-Level of T2: High, Medium, Catastrophic, No, Medium

T3: Unlimited Data Storage

The system must store data no longer than necessary need it, and it must explicitly define the expiration time for the collected data (any data related to COVID-19 positive cases and their contacts) because users—especially those who got infected—are concerned about how long their information will be kept in the database. Therefore, if data is stored longer than necessary and no clear rules are implemented to limit data storage, the associated parties may face the following consequences:

- The organization's financial situation can be severely affected by the increasing cost of storage and management of the IT infrastructure.
- It is very unlikely that the organization's reputation would be damaged by excessive storage of employees' contacts.
- The health condition of employees is not affected by excessive collecting of their contacts.
- The individual freedom of the employees is not affected by excessive storage of employees' contacts.
- The social situation of the employees is not affected by excessive storage of employees' contacts. However, as the volume of data stored grows, so does the possibility of data breaches and leaks.

Aversion-Level of T3: High, Low, No, No, Low

T4: Battery drain Attack (DoS Attack)

An adversary can drain employees' Token batteries by sending fake contact messages that the victim's device assumes as real contacts. The P2P approach may be more prone to this type of attack. Therefore, this kind of attack can have a significant impact on the contact tracing network, and the associated parties may face the following consequences:

- The organization's financial situation can be severely affected. Because these kinds of attacks may impose some cost on the organization, such as the cost of recovery procedures or the cost of energy might need to recharge the Tokens.
- The organization's reputation can be considerably affected due to this Dos attack where the service might be temporarily unavailable and as a consequence employees may lose trust in the system.
- The health condition of employees can be severely affected due to unavailable contact tracing service or the fake messages can create wrongly the network of contacts.
- The individual freedom of the employees is not affected by this type of attack.
- The social situation of the employees is not affected by this type of attack.

Aversion-Level of T4: High, Medium, High, No, No

T5: Intervenability Threat

Suppose there is no implemented procedure in the system to allow the employees to notify the system administrator to rectify, erase, or block the wrong registered contacts. In that case, it causes the contact tracing network to be created incorrectly and results in incorrect notifications as well as losing trust in the system. Therefore under this threat, the associated parties may face the following consequences:

- The organization's financial condition may suffer if employees are unable to rectify their wrong contacts, for example, inaccurate data may result in employees being forced to take unneeded leave or being in quarantine.
- The organization's reputation may be affected if employees cannot rectify their wrong contacts. This results in low-quality data and thus, imperfect contact tracing service.
- The health condition of employees can be severely affected due to not being able to rectify wrong recorded contacts or not being able to report to the system for the case of being somewhere without carrying out the Token.
- The individual freedom of the employees may be slightly affected due to wrong recorded contacts. For instance, they may be forced to stay away from others for a while.
- The social situation of the employees may be slightly affected in case of wrong recorded contacts that may lead to discrimination or social pressure.

Aversion-Level of T5: Low, Low, High, Medium, Medium

C.5 Assessing Trace4Safe solution according to the EDPB's requirements

Table outlines the requirements set out by the EDPB in the "analysis guide" appended to its Guidelines 04/2020 [1] and describes, for each of them, whether the Trace4Safe solution is compliant. The compliance assessment is done based on the following scales and colours:

Table C.5: Compliance scale.

1 =	Fully-Compliance: The solution fully complies with the requirement.
2 =	Partially-Compliance: Compliance with the requirement is possible but requires an adaption to the Trace4Safe solution.
3 =	Non-Compliance: The Trace4Safe solution cannot comply with the requirement.
4 =	Not applicable: The requirement is not applicable in Trace4safe solution.

In Table C.6, we have assessed the Trace4Safe solution with the requested requirements by EDPB; as can be seen, the rows with red color mean non-compliance with the requirements. The first non-compliance requirement (DATA-4) is related to using Puedo-random identifiers that are not used in the case of Trace4Safe. The second one (SEC-6) is associated with the lack of use of authentication mechanisms to prevent impersonation attacks. There are other requirements that we point out in yellow color which are not fully compliant in our scenario; for example, the DATA-2 states that the broadcast data must be encrypted and use strong Puedo-identifiers, while, in the case of Trace4Safe, only anonymized and fixed identifiers are used. The rows with the gray color show the requirements that are not applicable in our scenario.

[1] Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak, `https://edpb.europa.eu/our-work-tools/our-documents/linee-guida/guidelines-042020-use-location-data-and-contacttracing_en`

Table C.6: Assessing the Trace4Safe solution according to the EDPB's requirements.

Ref.	Requirement	Assessment	Comments
PUR-1	The application must pursue the sole purpose of contact tracing so that people potentially exposed to the SARS-Cov-2 virus can be alerted and taken care of. It must not be used for another purpose.	1	
PUR-2	The application must not be diverted from its primary use for the purpose of monitoring compliance with quarantine or confinement measures and/or social distancing.	1	Trace4Safe cannot be used for this purpose.
PUR-3	The application must not be used to draw conclusions on the location of the users based on their interaction and/or any other means.		
FUN-1	The application must provide a functionality enabling users to be informed that they have been potentially exposed to the virus, this information being based on proximity to an infected user within a window of X days prior to the positive screening test (the X value being defined by the health authorities).	1	
FUN-2	The application should provide recommendations to users identified as having being potentially exposed to the virus. It should relay instructions regarding the measures they should follow, and they should allow the user to request advises. In such cases, a human intervention would be mandatory.	1	
FUN-3	The algorithm measuring the risk of infection by taking into account factors of distance and time and thus determining when a contact has to be recorded in the contact tracing list, must be securely tuneable to take into account the most recent knowledge on the spread of the virus.	1	
FUN-4	Users must be informed in case they have been exposed to the virus, or must regularly obtain information on whether or not they have been exposed to the virus, within the incubation period of the virus.	1	
FUN-5	The application should be interoperable with other applications developed across EU Member States, so that users traveling across different Member States can be efficiently notified.	4	
DATA-1	The application must be able to broadcast and receive data via proximity communication technologies like Bluetooth Low Energy so that contact tracing can be carried out.	1	

DATA-2	This broadcast data must include cryptographically strong pseudo-random identifiers, generated by and specific to the application.	2	The broadcasting messages contain Token IDs (users' identifiers), Timestamp, and Location (in the case of RTLS).
DATA-3	The risk of collision between pseudo-random identifiers should be sufficiently low.	2	
DATA-4	Pseudo-random identifiers must be renewed regularly, at a frequency sufficient to limit the risk of re-identification, physical tracking or linkage of individuals, by anyone including central server operators, other application users or malicious third parties. These identifiers must be generated by the user's application, possibly based on a seed provided by the central server.	3	
DATA-5	According to the data minimisation principle, the application must not collect data other than what is strictly necessary for the purpose of contact tracing.	1	
DATA-6	The application must not collect location data for the purpose of contact tracing. Location data can be processed for the sole purpose of allowing the application to interact with similar applications in other countries and should be limited in precision to what is strictly necessary for this sole purpose.	4	In Trace4Safe, the indoor location will collect only in the case of RTLS approach.
DATA-7	The application should not collect health data in addition to those that are strictly necessary for the purposes of the app, except on an optional basis and for the sole purpose of assisting in the decision making process of informing the user.	1	
DATA-8	Users must be informed of all personal data that will be collected. This data should be collected only with the user authorization.	1	
TECH-1	The application should available technologies such as use proximity communication technology (e.g. Bluetooth Low Energy) to detect users in the vicinity of the device running the application.	1	
TECH-2	The application should keep the history of a user's contacts in the equipment, for a predefined limited period of time.	1	
TECH-3	The application may rely on a central server to implement some of its functionalities.	1	
TECH-4	The application must be based on an architecture relying as much as possible on users' devices.	4	
TECH-5	At the initiative of users reported as infected by the virus and after confirmation of their status by an appropriately certified health professional, their contact history or their own identifiers should be transmitted to the central server.	1	
SEC-1	A mechanism must verify the status of users who report as SARS-CoV-2 positive in the application, for example by	4	Depends on the specific internal procedures of the end customer.

	providing a single-use code linked to a test station or health care professional. If confirmation cannot be obtained in a secure manner, data must not be processed.		
SEC-2	The data sent to the central server must be transmitted over a secure channel. The use of notification services provided by OS platform providers should be carefully assessed, and should not lead to disclosing any data to third parties.	1	For securing the channel between the edges and the cloud server data is transmitted via an VPN connection. For securing the channel between the cloud server and API service provider, data is transmitted through an encrypted TLS connection.
SEC-3	Requests must not be vulnerable to tampering by a malicious user.	1	
SEC-4	State-of-the-art cryptographic techniques must be implemented to secure exchanges between the application and the server and between applications and as a general rule to protect the information stored in the applications and on the server. Examples of techniques that can be used include for example : symmetric and asymmetric encryption, hash functions, private membership test, private set intersection, Bloom filters, private information retrieval, homomorphic encryption, etc.	1	
SEC-5	The central server must not keep network connection identifiers (e.g., IP addresses) of any users including those who have been positively diagnosed and who transmitted their contacts history or their own identifiers.	4	
SEC-6	In order to avoid impersonation or the creation of fake users, the server must authenticate the application.	3	There is no authentication phase to authenticate tokens.
SEC-7	The application must authenticate the central server.	4	
SEC-8	The server functionalities should be protected from replay attacks.	2	
SEC-9	The information transmitted by the central server must be signed in order to authenticate its origin and integrity.	2	The information is not signed but it transfers via an VPN connection.
SEC-10	Access to all data stored in the central server and not publicly available must be restricted to authorised persons only.	1	
SEC-11	The device's permission manager at the operating system level must only request the permissions necessary to access and use when necessary the communication modules, to store the data in the terminal, and to exchange information with the central server.	4	
PRIV-1	Data exchanges must be respectful of the users' privacy (and notably respect the principle of data minimisation).	1	
PRIV-2	The application must not allow users to be directly identified when using the application.	2	Trace4Safe uses the fixed token ID (identifier) which increases the chance of being de-identified, as well as users' movements.
PRIV-3	The application must not allow users' movements to be traced.	2	

PRIV-4	The use of the application should not allow users to learn anything about other users (and notably whether they are virus carriers or not).	1	Only the Covid safety responsible person knows about the virus carriers.
PRIV-5	Trust in the central server must be limited. The management of the central server must follow clearly defined governance rules and include all necessary measures to ensure its security. The localization of the central server should allow an effective supervision by the competent supervisory authority.	1	
PRIV-6	A Data Protection Impact Assessment must be carried out and should be made public.	1	
PRIV-7	The application should only reveal to the user whether they have been exposed to the virus, and, if possible without revealing information about other users, the number of times and dates of exposure.	1	
PRIV-8	The information conveyed by the application must not allow users to identify users carrying the virus, nor their movements.	1	
PRIV-9	The information conveyed by the application must not allow health authorities to identify potentially exposed users without their agreement.	1	Health authorities have no access to such information.
PRIV-10	Requests made by the applications to the central server must not reveal anything about the virus carrier.	1	Only the COVID safety responsible person in the company knows about the virus carrier where he has to submit the positive report.
PRIV-11	Requests made by the applications to the central server must not reveal any unnecessary information about the user, except, possibly, and only when necessary, for their pseudonymous identifiers and their contact list.	2	Partially compliance (in the case of collecting name and surname by the central server)
PRIV-12	Linkage attacks must not be possible.	2	
PRIV-13	Users must be able to exercise their rights via the application.	4	
PRIV-14	Deletion of the application must result in the deletion of all locally collected data.	1	
PRIV-15	The application should only collect data transmitted by instances of the application or interoperable equivalent applications. No data relating to other applications and/or proximity communication devices shall be collected.	4	
PRIV-16	In order to avoid re-identification by the central server, proxy servers should be implemented. The purpose of these non-colluding servers is to mix the identifiers of several users (both those of virus carriers and those sent by requesters) before sharing them with the central server, so as to prevent the central server from knowing the identifiers (such as IP addresses) of users.	4	

PRIV-17	The application and the server must be carefully developed and configured in order not to collect any unnecessary data (e.g., no identifiers should be included in the server logs, etc.) and in order to avoid the use of any third party SDK collecting data for other purposes.	1	
ID-1	The central server must collect the identifiers broadcast by the application of users reported as positive to COVID-19, as a result of voluntary action on their part.	1	
ID-2	The central server must not maintain nor circulate the contact history of users carrying the virus.	1	
ID-3	Identifiers stored on the central server must be deleted once they were distributed to the other applications.	4	
ID-4	Except when the user detected as positive shares his identifiers with the central server, no data must leave the user's equipment or when the user makes a request to the server to find out his potential exposure to the virus, no data must leave the user's equipment.	4	
ID-5	Data in server logs must be minimised and must comply with data protection requirements.	1	Data after a specified period of time will be deleted, and a blockchain mechanism just on some anonymized data will be performed.