THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Model-based Approaches to Privacy Compliance

Hanaa Alshareef

Department of Computer Science & Engineering
Chalmers University of Technology
Gothenburg, Sweden, 2022

Model-based Approaches to Privacy Compliance

Hanaa Alshareef

ISBN 978-91-7905-684-1
Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5150
ISSN 0346-718X

Technical report no 221D
Department of Computer Science & Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone +46 (0)31-772 1000

Gothenburg, Sweden, 2022

Model-based Approaches to Privacy Compliance
Hanaa Alshareef
Department of Computer Science & Engineering
Chalmers University of Technology

Abstract

In the last decade, information technologies have been developing dramatically, and therefore data harvested via the Internet is growing rapidly. This technological change has a negative impact on privacy due to the sensitivity of the data collected and shared without convenient control or monitoring. The General Data Protection Regulation (GDPR) of the European Union has been in effect for more than three years, limiting how organizations collect, manage, and handle personal data. The GDPR poses both new challenges and opportunities for technological institutions. In this work, we address various aspects of privacy and propose approaches that can overcome some challenges of the GDPR. We focus on improving two currently adopted approaches to leverage them to enforce some of the GDPR's requirements by design.

The first part of this work is devoted to developing an *access control model* to effectively capture the nature of information accessed and shared in online social networks (OSNs). They might raise serious problems in what concerns users' privacy. One privacy risk is caused by accessing and sharing co-owned data items, i.e., when a user posts a data item that involves other users, some users' privacy might be disclosed. Another risk is caused by the privacy settings offered by OSNs that do not, in general, allow fine-grained enforcement. We propose a collaborative access control framework to deal with such privacy issues. We also present a proof-of-concept implementation of our approach.

In the second part of the thesis, we adopt *Data Flow Diagrams* (DFDs) as a convenient representation to integrate privacy engineering activities into software design. DFDs are inadequate as a modeling tool for privacy, and there is a need to evolve them to be a privacy-aware approach. The first privacy-related lack that we solve is automatically inserting privacy requirements during design. Secondly, since DFDs have a hierarchical structure, we propose a refinement framework for DFDs that preserves structural and functional properties and the underlying privacy concepts. Finally, we take a step towards modeling privacy properties, and in particular purpose limitation, in DFDs, by defining a mathematical framework that elaborates how the purpose of a DFD should be specified, verified, or inferred. We provide proof-of-concept tools for all the proposed frameworks and evaluate them through case studies.

Keywords: GDPR, privacy by design, collaborative access control, social networks, data flow diagram, refinement, purpose limitation

List of publications

This thesis is based on the following publications, each presented in a separate chapter.

Paper A "A Collaborative Access Control Framework for Online Social Networks"
Hanaa Alshareef, Raúl Pardo, Gerardo Schneider, Pablo Picazo-Sanchez.
Journal of Logical and Algebraic Methods in Programming (JLAMP'20), 114, May 2020.

Paper B "Transforming Data Flow Diagrams for Privacy Compliance"
Hanaa Alshareef, Sandro Stucki, Gerardo Schneider.
the 9th International Conference on Model-Driven Engineering and Software Development (MODELSWARD'21), Vienna, Austria, volume 21, pages 207-215. February 2021.

Paper C "Systematic Enhancement of Data Flow Diagrams with Privacy Checks"
Hanaa Alshareef, Sandro Stucki, Gerardo Schneider.
Manuscript (under submission).
This paper is an extended version of Paper B.

Paper D "Refining Privacy-Aware Data Flow Diagrams"
Hanaa Alshareef, Sandro Stucki, Gerardo Schneider.
In 19th International Conference Software Engineering and Formal Methods (SEFM'21), Virtual Event, December 6-10, 2021, volume 13085 of *LNCS*, pages 121-140. Springer, 2021.

Paper E "Precise Analysis of Purpose Limitation in Data Flow Diagrams"
Hanaa Alshareef, Katja Tuma, Sandro Stucki, Gerardo Schneider, Riccardo Scandariato.
The 17th International Conference on Availability, Reliability and Security (ARES'22), Vienna, Austria, August 2022.

“It used to be expensive to make things public and cheap to make them private. Now it is expensive to make things private and cheap to make them public.”
- Clay Shirky

Acknowledgments

In the words of James Allen, “No duty is more urgent than giving thanks”. Many people helped me along my Ph.D. journey and provided tremendous support, advice, time, and energy to help me reach its end.

Firstly, I would like to express my profound gratitude to my supervisor Gerardo Schneider for his support, encouragement, guidance, thoughtful advice, and insightful comments that have been helpful in improving and advancing my knowledge. I could not have completed this journey without my co-supervisor, Sandro Stucki. I owe you my gratitude for your wise guidance, words of encouragement, and priceless advice. There are so many things I have and could have learned from you still. Your support as a co-supervisor is invaluable. For your kind and dedicated mentorship, I remain deeply indebted to you.

I am very grateful to my co-authors Raúl Pardo, Pablo Picazo-Sanchez, Katja Tuma, and Riccardo Scandariato for the fruitful discussions and pleasant collaborations.

I have been fortunate to many whom by being around, have helped me and made my time at Chalmers more pleasant. To all the past and present colleagues in the Computer Science and Engineering department, Thank you all! Iulia for supporting and standing as an inspiring model for women in computer science; Irene for your warm heart and kind words, but also for our long walks, brunches and Fika; Fatima and Monica for your smiles and warm hugs. Wolfgang for the exciting discussions about science, history, and food.

My deepest gratitude is directed toward my parents for their unconditional love and countless support. My warm and heartfelt thanks go to my siblings, their families, and my nephews and nieces, who keep me grounded, remind me of what is important in life, and are always supportive of my adventures. Tagreed, you are a sister of my heart. I am so grateful to you. Thank you for your love, support, and wisdom. I am thankful to my family for always being there.

To my special one, Yassine, thank you for always being there to encourage me and lift my spirits. I am grateful for your care and patience during my irregular working hours and the lack of availability. Thank you for being my partner, lover, muse, and best friend. I am looking forward to all our upcoming adventures.

A big thank goes to my friends, Tagreed (again!), Azhar, and Mohammad, who live far but were the closest to me and always available whenever I needed you. I thank all my friends, especially Mawadah, Rosa, Amal, Esraa, Zeinab, Moufida, and

Somayaa, for remaining engaged and supportive and for giving me many things to enjoy outside the Ph.D.

I am thankful to my opponent Thomas Troels Hildebrandt and the grading committee members Barbara Carminati, Jaap-Henk Hoepman, and Nataliia Bielova for reviewing this thesis.

There are so many people I sincerely wanted to thank for their support in various capacities during my Ph.D., but writing this thesis is limited by time and space. Thank you to all of you who contributed to making my journey as smooth as possible!

شكرَا أمي و أبي كنتم خير دَاعم لي و عذرَا علَى كل تقصير مني.

July 25th, 2022

Contents

Overview

Papers

Overview

1

Introduction

With cookie warnings popping up on every website, it is easy to get the impression that privacy is quite a new concept developed with the capability of new web-service technologies; this view is not accurate. During the early time of the industrial revolution, officials perceived privacy as a fundamental human right. The right of privacy that emerged during the Gilded Age (1840-1950), was formed into a constitutional creed by 1965, which is considered one of the oldest constitutional rights. Warren and Brandeis define privacy as the “right to be let alone” [108]. Decades later, Westin referred to it as “the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others.” [109]. Because of the importance of privacy for an individual’s autonomy, identity, and integrity, many have attempted to define privacy [5, 109]. However, this turns out to be difficult, and some have argued that there is no single definition of privacy universally applicable [97]. Privacy is recognized as a human right by many international and regional agreements, such as the Universal Declaration of Human Rights [11, 64] and the European Convention on Human Rights [80]. Privacy is closely related to other fundamental human rights such as dignity, personal autonomy or self-determination, freedom, individuality, respect, etc. This gives privacy paramount importance.

In the twenty-first century, technological change has apparently shaken up society and privacy along with it; thus, theories and models have been developed to achieve and meet the demands of this change. Information privacy was clearly recognized as an issue when the Internet was commercialized in the United States [35]. As a result, the Data Protection Directive of the European Union has defined information privacy explicitly as a basic human right [39].

Given the dramatic improvements in information technologies (e.g., big data, digital identity, biometrics, internet of things and online social networks) in the last few decades, along with the increasing processing and storing capacity of computer devices, technology has become pervasive in our daily activities. An extensive amount of information is thus available over those technologies, making privacy particularly important in the socio-technical landscape. The proliferation of online data collected in everyday life has a destructive effect on privacy due to the sensitivity of the data collected and shared without convenient control or monitoring. This data covers more than just our names, phone numbers, emails, or addresses. It extends much further beyond that, consisting of information about our preferences for food, clothes, places, interests, daily routines, health information, and so much more. Consequently, companies are presented with the opportunity to use this data outside of its original context. For example, Facebook handed over personally identifiable information of more than 87 million users to the company Cambridge Analytica who had used personal data to create algorithms to affect the US Presidential Election [90]. Sharing our personal information with others and how it can be used should be managed by the owners of this information. We do not want to be watched by everyone. End-user research indicates that people become increasingly uneasy and fearful of losing control over their personal data. The majority of users in the United States and Europe believe they have lost control of their personal information and are anxious about third-party corporations or the government gaining access to it [75, 47].

As information technology is evolving, it is essential that the legislation follows the same development. For over three years now, the European Union (EU) General Data Protection Regulation (GDPR) has been in effect, restricting organizations' personal data collecting, managing, and processing actions [1]. Organizations that process personal data relating to EU residents will be held liable for violations of the GDPR's provisions.

The GDPR presents both a new challenge and an opportunity for technology institutions. In this work, we address different aspects of privacy and propose approaches that can overcome some challenges of the GDPR. Primarily, we focus on developing technical solutions to protect "personal data by design" and achieve legal compliance with certain GDPR obligations.

I.1 General Data Protection Regulation

The GDPR is a European Union legislation that has been enforced since May 25, 2018. It regulates the way organizations manage, collect, and process personally identifiable information for EU citizens (regardless of the location of their data). This regulation constitutes the most crucial change to the EU's data protection rules in the last 20 years since it aims to transform the handling of personal data and the attitude towards it [48, 4, 72]. According to the regulations, EU courts allow penalizing any business that mistreats its citizens' data.

The EU seeks to grant individuals significant control and power regarding their data with this regulation. The GDPR has an extensive interpretation of *personal data* that covers any information related to an identifiable natural person, namely, the *data subject* (e.g., a name, an identification number, location data, email addresses, telephone numbers, online identifiers, etc.). The legislation appoints data subjects' rights over the information that data *controllers* (e.g., web services) collect and data *processors* (e.g., cloud providers) store and process. For example, the GDPR requires corporations to promptly provide customers with electronic copies of their personal data ("right of access", Article 15), and completely remove the user's personal data from their databases on request ("right to be forgotten", Article 17 and "withdraw consent", Article 7)

The GDPR stipulates tough financial penalties for violators and sets stringent rules on an organization's data collection and process practices. Therefore, many organizations have adopted the GDPR as their default privacy standard. However, the GDPR is a complex and extensive regulation, which is a challenge in itself [48, 55]. Moreover, the GDPR is primarily a legal document, delivering little if any guideline regarding technologies that should be used to comply with its provisions [105]. This was a conscious choice, as the EU did not want to link the GDPR to explicit technologies that would tend to specific platforms and solutions. This approach, however, causes unexpected complications to many organizations endeavoring to adapt their internal processes to the GDPR's requirements [10].

To achieve compliance with the GDPR, among a list of technical and non-technical challenges to be tackled, translating the GDPR's provisions into software requirements and technical solutions is a significant challenge. The compliance process may be expensive and time-consuming as it demands substantial financial and human resources, especially in case of the lack or insufficiency of legal and privacy knowledge and expertise [72, 95]. These shortages in legal and privacy knowledge translate into a lack of awareness or difficulty in understanding the regulation and may require an extra budget to recruit privacy experts. A software engineer is not likely to be a legal expert nor trained in privacy law [78]. A study conducted across the EU, UK, and US reported that the primary dilemma mentioned was the complexity and expansion of the GDPR [53].

The costs of the GDPR compliance ranged from $\$500,000$ to over a million dollars. A study with Norwegian companies reported that 23% of the respondents stated that a budget's insufficiency is one of the main challenges in complying with the GDPR [84]. A large organization may be able to allocate a large amount of money in both technologies and legal consulting to work predominantly on the GDPR compliance work. Unfortunately, many Small and Medium-sized Enterprises (SMEs), representing 99% of all businesses in the EU, are restricted by the number of resources they can allocate towards the GDPR compliance [72, 95, 19, 104].

Successful compliance reduces the risk of hefty fines of 4% of the business's global revenue or up to 20 million euros [105, 19, 1]. Moreover, applying and demonstrating the GDPR compliance brings a new opportunity to companies [103]. First, they can gain control over their personal data, which contributes to preventing the misuse of personal data and ensuring data consistency across the organization [84, 96]. Another potential benefit of being GDPR compliant is generating a reputation as a trustworthy company owing to the ability to guarantee data security governance, which will drive them to gain customers' trust [71]. In recent years, personal data privacy breach scandals and reports of how companies inappropriately use and sell data they collect from their customers have evoked general concern that negatively impacted customer trust [25]. Capgemini's report reveals that customers' spending increases when they are convinced that a company protects their personal data [62].

The scientific communities and private companies are actively working to provide theoretical and practical solutions at different levels of the development life cycle for different sectors to ensure the GDPR compliance (e.g., [3, 13, 14, 19, 83, 87, 81]). The GDPR has introduced critical changes to privacy and data protection regulation, thus significantly influencing those who need to design a technical system. One of the main stipulations in the GDPR is *Privacy by Design* (PbD), which indicates that data privacy is not an addition to the process but an integral part of it [23]. In this thesis, our approaches to enhancing privacy follow the by-design and by default paradigms. The PbD and its effectiveness as a technological concept are briefly discussed in the following section.

I.2 Privacy by Design

Privacy by Design advocates for the proactive consideration and incorporation of privacy protection requirements and measures during the design stage of technological systems, making privacy the default setting and ensuring transparency regarding collecting, processing, disseminating, and storing personal data throughout the data lifecycle [22, 98, 51]. For example, PbD principles require that systems be designed with minimum data collecting methods and with appropriate notice and consent interactions. PbD was advocated mainly by Cavoukian, in 1995, after her long experience in the Office of the Information and Privacy Commissioner [91]. PbD has been incorporated into the GDPR in Article 25. Similarly, it is included in some recommendations issued by the U.S. Federal Trade Commission (FTC) [99, p.2]. It is accepted by data-protection commissioners worldwide as a concept that will assure sustainable compliance with most privacy protection principles in a world of constantly evolving IT systems.

Despite the apparent simplicity and generality of the concept of PbD, it is unclear how it should be translated into concrete guidelines. With time, various studies have proposed PbD schemes (e.g., principles, strategies, guidelines, patterns) to guide and encourage software designers and developers to produce privacy-aware systems (e.g., [22, 54, 82, 86]). Due to the variations in both socio-cultural and technical aspects of privacy in each of these schemes, they stand isolated without proper connection. Such a disconnected view makes the schemes challenging and confusing as means to guide software designers or developers toward particular practices of privacy. Recently, aware of the rather theoretical character of this principle, the European Data Protection Board (EDPB) has released an official document for providing PbD guidelines [41]. However, with the intent to be more concrete and tangible, this guideline is still at a high level and offers few practical indications [83, 65].

It is challenging for engineers (i.e., software designers, software architects, information architects, interaction designers, product designers, and related specialties) to extract, translate, integrate and encode the PbD principles. By being aware of their essential roles in the process of implementing PbD, several studies have recently emerged on the perceptions and interpretations of software development professionals regarding privacy regulations generally and PbD especially ([15, 52, 68, 18]). Similar to the challenges of implementing the GDPR's provisions, software designers struggle to identify, extract, translate, integrate and encode the PbD principles into engineering activities [15, 52].

What is still required is a well-defined structured method that enables organizations to apply PbD and a set of tools to automate such a method. PbD is gaining momentum in part due to its inclusion in the GDPR and policy suggestions from the United States. Thus, its disregard is now a potential foundation for a hefty sanction in the EU. For example, the French Data Protection Authority recently fined a French-based small translation company, Uniontrad Company, 20,000 EUR for failing to observe PbD [65, 59]. The supervisory authority of Berlin also issued a 14.5 million EUR fine, Germany's largest GDPR fine, to Deutsche Wohnen for violating Article 5, which covers principles relating to the processing of personal data, and Article 25 (PbD) [28, 70].

Translating the privacy requirements (i.e., GDPR's provisions) into privacy engineering activities and then embedding them within software and systems engineering methods would support these approaches to address privacy concerns during software development activities.

This thesis focuses on architectural approaches to implementing PbD for several reasons [7, 29]. To begin, architectures carry the first and hence most fundamental design decisions; consequently, disregarding architectural choices can significantly impair the integration of privacy considerations into a system's design. Second, architectural approaches do not mention implementation details while at the same time describing the relevant aspects: problem, solution, and consequences. Thus, this elegant side of the architectural approach directs the inventiveness of developers, reducing design and system complexity. It enables the abstraction of superfluous details and the focus on critical issues, assisting software designers in reasoning about privacy provisions. Thirdly, the architecture makes it possible to create a transferable, reusable model. As a result, it can play a critical role in increasing the reusability of privacy-friendly technologies, potentially resulting in significant cost savings. Finally, the privacy-aware system design model helps acquire a privacy mindset from engineers responsible for the system across the organization [52].

The following sections introduce the two architectural approaches that we adopt to specify and represent privacy engineering activities in a way that is amenable to privacy compliance checking and assurance. First, we employ access control models, the de facto procedures for restricting data access, as a technical solution for protecting personal data by design and meeting some the GDPR standards. The access control models can be described as design patterns (i.e., software patterns)[85, 67]. Section I.4 discusses a design model called Data Flow Diagrams (DFDs) that we adopt as a convenient representation to integrate privacy engineering activities into software design.

I.3 Access Control

Access control mechanisms regulate how a subject may access an object (resource) and is one of the essential features of today's systems to protect access to data items [50, 16]. It has three main concepts: setting the *policies* that authorize certain individuals to access certain data items; *authenticating* evidence associated with an access request; *assessing* the access request based on the given policies [50]. Generally, access control focuses on addressing three main issues: confidentiality, integrity, and availability [89, 88].

Tolone et al. indicate access control mechanisms are designed to meet specific organizational structures, scenarios, and requirements; in other words, they commonly vary from domain to domain [106]. The focus of this research is the online social network domain.

After discussing access control in the GDPR, we briefly explain the structure of online social networks (OSNs) and their possible privacy breaches. Additionally, we provide background on the access control model OSNs implement.

I.3.1 Access Control and GDPR

41.4% of the 99 obligatory articles of the GDPR have been assessed as being related to access control [13]. Particularly, 33 related articles concern access control models, and 38 articles refer to access control policies. Enforcing a solid access control mechanism and policy are significant steps towards compliance with the GDPR. For example, integrating appropriate access control mechanisms and policies ensures by design the right of access to personal data (Article 15) and the right to data portability (Article 20).

Access control mechanisms rely on rules (i.e., policies) that determine who has access to which resources and under which events. However there is a gap between current access control mechanisms as a technical solution for protecting personal data and the requirements of the GDPR. For example, access control mechanisms embedded in online social network systems do not facilitate, as we will see in Chapter A, access for all involved users (data subjects) when dealing with a personal data co-owned by multiple users. Recent work has attempted to minimize the gap between current access control systems and its policies and the requirements of the GDPR [e.g. 30, 31, 32, 33, 12]. These works focus on augmenting access control systems with policies elicited from the GDPR's provisions [e.g. 33, 43, 13]. This results in a technical solution for protecting personal data that reaches legal compliance with the GDPR. Moreover, improving the currently adopted access control models based on the GDPR will leverage them to fulfill the principle of privacy by design. The proposal of our first paper attempts to improve the current OSNs' privacy protection mechanisms.

I.3.2 Privacy Policies and Access Control Models in Online Social Networks

Online services, such as online social networks, provide immense benefits for the society. However, they have also created unanticipated privacy breaches that compromise individual privacy. In this section, we introduce the online social network structure and its possible privacy violations.

Online social networks (OSNs) promote online social interactions between individuals [57]. Given their inherent structure, the most common way to represent OSNs is as *graphs* (usually called a *social graph* in this context) [20]. Vertices in the graph represent users and resources (e.g., pictures, posts, etc.) and edges of the graph are utilized to model the relationships among users and resources.

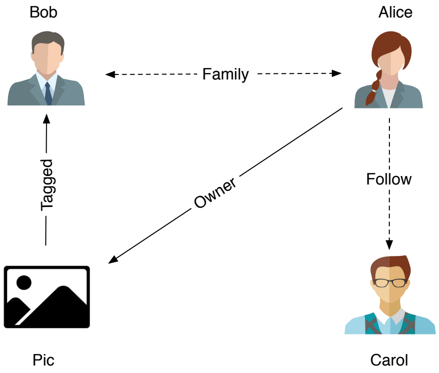

Figure I.1: Social Graph Example

Figure I.1 shows an example of such a graph. In this example, there are three users: Alice, Bob and Carol and one resource, a picture, indicated as Pic. Dotted arrows represent social relationships between users, while plain arrows are relationships between users and resources. The `family` relationship between Alice and Bob is a bidirectional relation (*symmetric*), as in Facebook. On the other hand, `follow` is a unidirectional relation (*asymmetric*) which means that a user can follow others without being followed. Additionally, the plain arrows indicate the connections between users and the resource Pic. The connection between Alice and Pic denotes that Alice is the owner of the picture, while Bob is tagged in it.

Privacy settings, from which hereafter we will refer to as *privacy policies*, in today OSNs allow users to set *who* can access *what* information. In most existing OSNs, users are provided with a large variety of relationships to create their own social circles (such as family, friends, colleagues, hiking group and acquaintances). The current privacy policies in OSNs are specified in terms of relationships. Typically, granting access to a data item is subject to the type of the relationship or its composition (e.g., friends, friends-of-friends). For instance, in the privacy policies of Facebook, users can determine who can access their posts or friends lists and also specify the actions that other users can perform. Although privacy policies enable users to protect their personal information from other users within the network, they are notoriously difficult to configure correctly and do not easily match the OSN's users sharing intentions [77, 76, 73, 63, 2, 6, 34]. Madden found that 48% of OSN users report struggles in managing their OSN privacy settings [74]. Privacy policies should be understandable and user-friendly, sticky, fine-grained, relationship-based, multiparty, specific to the type of content and trust-based [49, 117].

Concerning privacy policies in OSNs, we focus only on *Multiparty* and *Fine-grained* privacy policy features. In what follows, we discuss and give examples of problems arising from the lack of having these features.

Multiparty privacy policy Hundreds of billions of data items that are uploaded and shared in OSNs are co-owned by more than one user [61, 116, 79]. Nowadays, it is possible only for the space owner to specify the privacy policies of co-owned data items regardless of the privacy preferences of other users who are identified in these data items. Current OSNs offer limited support for managing co-owned data items where users can only use strategies like untagging or reporting inappropriate content. Such reporting mechanism does not solve the privacy issue due to co-ownership.

For instance, when tagging users upload a picture and name other users in it with a link to their profile. In some OSNs, the tagged users receive notifications about being tagged so they can approve it. If the tagged users in a picture do not want to share it with their list of connections, they can untag themselves from it. However, this strategy does not fulfill the users' desires for different reasons. The focal reason, when tagged users untag themselves from a picture, it does not mean that the actual picture is removed or that they block the possibility of the picture being accessed by undesired users.

The bottom line is that existing solutions are, in general, not enough [101, 100]. Privacy policies need to be designed to enable all users who are related to a given piece of information to involve in deciding who should access that co-owned item.

Coarse-grained privacy policy Privacy settings in most OSNs are not fine-grained adequately. For instance, on Facebook, users can not state policies like "I do not want to be tagged in pictures by anyone other than the members of my close friends' group" or "My post can be seen by my friends and friends of friends, but nobody apart from my family group can share it".

Also, OSNs' users cannot specify their privacy policies according to the type of data item. For instance, in Facebook users cannot choose a policy like "Only my friends can see a post having my location". Furthermore, the privacy protection mechanisms presently do not equip the users with features to identify the level of privacy concerns regarding their data item. For example, users cannot express policies like "I have high sensitivity level for all posts containing location".

OSNs' privacy policies determine who can access which data items. However, looking more in-depth at what users can do with someone else' data items, there are two main actions: accessing and disseminating (sharing). Since most privacy policies are centered around the data items' accessibility, it can be inferred that privacy policies are indirectly in control of sharing. However, the two actions are functionally different. This lack of policy options about sharing might lead to undesirable results and privacy breaches.

Given this, privacy policies need to be flexible to accommodate users' needs and intentions, and more fine-grained settings are needed. However, an equilibrium between too little flexibility and an excessively complicated privacy policy management is required.

Access Control Models in OSNs

In this section, we offer a brief background on the access control model implemented by OSNs. Several access control models that have been developed in recent years are aimed at effectively capturing the nature of information accessed and shared in OSNs [46, 26, 24, 45, 27, 21]. Numerous studies revealed ample evidence that users' relationships should be considered a central concept in modeling the privacy protection mechanism of OSN [24, 56, 110, 49, 45].

Relationship-Based Access Control (ReBAC) is a paradigm that captures the character of information accessing in OSNs by considering users' relationships as a core concept [46]. However, the ongoing privacy violations in OSNs indicate that ReBAC, as applied on OSNs, has limitations which means this model might need to be retrofitted. Fogues et al. discuss a few open challenges in ReBAC for OSNs [44]. They allude, among others, the following issues:

- *A privacy protection mechanism is needed to enforce the privacy preferences of all involved users when dealing with a data item that is related to other users.* This issue relates to the aforementioned problem of lack of having *Multiparty* policies.

- *Privacy policies that OSNs provide do not capture how data items should be disseminated.* This issue is related to the problem arising from lack of *Fine-grained* policy.

In conclusion, OSNs' privacy protection mechanisms should be supported by multiparty privacy management and their privacy policies should be more expressive to mitigate undesired disclosure of sensitive data. In the first part of this thesis, we provide a framework that empowers OSNs' users to collectively manage viewing and sharing their co-owned data items.

I.4 Data Flow Diagram

Data Flow Diagrams (DFDs) are used as activity-oriented models for a structured analysis technique. This software design model is a graphical approach that is easy to understand and helps to depict logic models and express data transformation from input to output in a system [69, 66]. DFDs have been widely applied in requirement analysis and structured analysis in software development. They are often used during the early phases of software design. DFDs may be used also to evaluate security and privacy issues (e.g., threats and vulnerabilities) of software systems [92, 113].

DFDs include four main elements. The first component is data flows ("arrows" movement of data in the system). The other three components are activators: external entities ("boxes" representing sources or destinations outside the system boundary), processes (computation applied to the data in the system and transformation of incoming data flows into outgoing data flows) and data stores. A process may symbolize detailed low-level operations or complex high-level functionality that could be illustrated in more detailed processes; a double-lined circle or ellipse represents such composite processes. At the bottom of Figure I.2 (standard notation), all DFD elements are shown.

All the elements in DFDs should be labeled. Moreover, the composition of these elements must adhere to well-formedness criteria in order to maintain diagram consistency. First, a data flow coming from (going to) a data store or external entity must go to (come from) a process (composite process). Second, processes must have

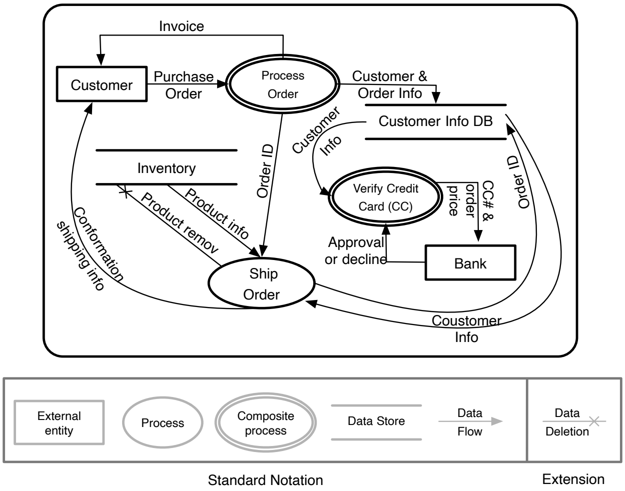

Figure I.2: Example of a Data Flow Diagram (DFD). The 'data deletion' element is an extension to the standard DFD notation (Business-oriented DFD (B-DFD).

at least one incoming and outgoing flow [see e.g. 42, 60, 38]. Antignac et al. extended the standard notation of DFDs with *data deletion* type of flow [8, 9]. It is an incoming flow for data stores, reflecting the deletion of previously stored data whose reference is indicated in the flow. This extension is referred to as a *Business-oriented DFD* (B-DFD), which we adopt in this work. Note that we use the terms DFD and B-DFD interchangeably in this thesis.

The example in Figure I.2 shows a part of the e-store ordering system that allows customers (external entity) to order products. A customer's order has to take place in the "Customer Info" database via "Process Order" subsystems in order to ship the requested products. Then, the shipping subsystem acquires the necessary information for dispatching the order from the "Customer Info" database by using the order ID.

Although DFDs have been widely used in both requirement and structured analysis in software development, they focus mainly on functional aspects. With the requirement from the GDPR to comply from the design stage, there is a demand to evolve DFDs to be a privacy-aware approach. DFDs are lacking as a modeling tool for privacy in general. In what follows, we discuss the related privacy problems that DFDs have.

Incorporating privacy requirements DFDs highlight the processing and transformation of data as they move through various processes. When such operations are performed on sensitive (private) data, there should be some kind of control on that the operations respect privacy principles (i.e., GDPR provisions).Thus, there is a need to embed privacy concepts into DFDs.

Threat modeling is an important activity for eliciting potential security and privacy flaws in a software system. This activity is regarded as one of the keys to developing a secure and privacy-friendly system [92]. For example, the STRIDE threat modeling methodology for security [58], and LINDDUN, which focuses solely on privacy [37, 112], both begin with a Data Flow Diagram (DFD). LINDDUN leverages a DFD as a representation of a system to be analyzed and examined for privacy threats. It was developed by a research group at KU Leuven in Belgium [115]. LINDDUN acronym stands for the types of privacy threats that the methodology helps to specify: Linkability, Identifiability, Non-repudiation, Detectability, Disclosure of information, Unawareness and Non-compliance. It has three initial steps. First, a system is modeled as a DFD to capture the main data-centric activities (e.g., processing, storage, collecting and disclosing) in a system for conducting a privacy assessment. The second step of the framework is eliciting threats. Each element in the DFD is mapped with some privacy threat types. LINDDUN supports the privacy analyst with a mapping table template and threat trees that help identify potential privacy threats and select corresponding privacy controls. After eliciting all the relevant threats, it is the phase of manually prioritizing and assessing them with respect to their importance. To facilitate this managing threats process, LINDDUN provides a list of mitigation strategies, which helps to find countermeasures, and lists of privacy enhancing technologies (PETs).

LINDDUN is a well-established risk analysis methodology that is gaining popularity in academics and industry [29, 114]. Its primary strength is the integration of a methodological approach with a comprehensive privacy knowledge base consisting of the mapping table, threat trees, the taxonomy of mitigation strategies and the classification of privacy solutions.

However, even with the provided systematic support to elicit and mitigate privacy threats, performing this analysis requires significant effort. LINDDUN demands that the analyst conducting the analysis has considerable privacy expertise and adequate experience with the threat modeling technique [114, 94]. Instead of eliciting privacy threats incorporating privacy provisions into software architectures (e.g., DFDs) can alleviate some of the efforts.

Antignac *et al.* enhanced DFD models by embedding privacy requirements in spots where potential privacy violations might occur. The enhanced diagrams were named *Privacy-Aware Data Flow Diagram*, or PA-DFD, for short. PA-DFDs contain checks for retention time and purpose limitation privacy concepts. The approach also explores the feasibility of logging all information (for accountability) and managing privacy policies. This work aims to achieve privacy by design in software systems, and the transformation from DFDs to PA-DFDs is described using high-level graphical "rules". However, neither a comprehensive algorithm nor a reference implementation is provided. Using Antignac et al.'s approach means software engineers should apply the rules to DFDs informally and manually during the design phase. This is an error-prone task and can become a challenge and place a burden on software engineers. Thus, assisting them by systematically and automatically inserting privacy checks during design is required.

Refinement framework DFDs have a hierarchical structure, which provides different abstraction levels. Such structure is helpful when modeling real-world systems that may have huge DFDs. The highest level of a DFD depicts all external entities and the key data flows between them and a system depicted as a single composite process. This level is known as the *Context Diagram*, which is partitioned into diagrams at a lower level. It may, in turn, be exploded to create a more detailed child diagram (i.e., a refined diagram). The low-level diagram processes are more specific and illustrate the logic required to generate the outgoing data flows.

Previous works have discussed *leveling* in DFDs (hierarchical modeling) and *consistency rules* [36, 102, 111]. There are two primary rules for ensuring consistency between an abstract level and its refined version. First, the balancing rule dictates that every process, data store and external entity on an abstraction level is shown on a refined level. Moreover, all data flows are determined at an abstract level must hold on its refined version (i.e., preservation of connectivity). Concerning refinement in DFDs, the only known study that defines a concept of refinement for DFDs is that by Ibrahim *et al.* [60]. They have codified some conventional structured DFD rules to ensure model consistency, but only between context and Level 0 DFDs. Thus, there is a necessity for a formal definition of refinement for DFDs for numerous arbitrary levels. Likewise, the enhanced DFDs (PA-DFDs) require a notion of refinement that preserves structural and functional properties and the underlying privacy concepts.

Modeling the purpose of data operations DFDs are primarily used for modeling operations on data such as storage, forwarding and processing (functional properties). In order to achieve privacy by design in software systems, each operation on personal data must be executed based on specific purposes. A fundamental principle of data protection in general, and the GDPR in particular, is that organizations must collect and process personal data for explicit and specified purposes and only for the purposes for which it was collected. This requirement is called purpose limitation, introduced in Article 5(1)(b) of the GDPR. The purpose limitation requirement has had an immediate and consequential effect. In January 2019, for instance, the French data protection commission penalized Google € 50 million for lacking the legal basis for personalizing its advertisements [40]. Analyzing data purposes not only facilitates compliance with legislation but also enables users to comprehend the data practices of companies better [17]. Thus, they can make educated decisions regarding the use of Internet-based services.

Our previous work on privacy-aware DFDs explored purpose limitation in a general sense without expanding on how the purpose of DFD activators and flows should be specified, validated, or inferred. Basin *et al.* [14] analyze the purpose specification principle and propose a methodology for checking GDPR compliance in business process models, which are related to DFDs in some way. However, their research also emphasizes the difficulties of capturing the concept of purpose at the level of software entities. Although we agree with this viewpoint, we believe it is necessary to attempt to evolve DFDs to model non-functional properties, such as purpose limitation.

Solving the DFDs' privacy-related shortcomings mentioned above leads us to construct a privacy-aware system design model. In this thesis, we propose such models. First, we automate the conceptual model of transforming DFDs into PA-DFDs and check its correctness. This automatic transformation approach represents an attempt to make it approachable to integrate privacy principles into a software design, even for regular software engineers without privacy expertise. Secondly, we provide a refinement framework for DFDs and PA-DFDs to preserve structural and functional properties and the underlying privacy principles. This refinement approach can be executed for numerous arbitrary levels of DFDs and PA-DFDs. To the extent of our knowledge, our refinement approach is the first formal definition of DFD refinement. Finally, to model the purpose of data operations in DFDs, we propose a formal approach that elaborates on how the purpose can be specified, verified, or inferred. Several approaches are complementary to our work [e.g. 93, 107, 14]. Nevertheless, as far as we know, our proposed framework represents the first attempt to annotate and analyze purposes in all steps of the personal data lifecycle (data collection, disclosure, usage, storing and retrieval).

II

Thesis Objectives and Structure

With the introduction of the GDPR and its requirements for privacy by design and privacy by default, organizations have started looking for mechanisms and tools that help engineers comply with the GDPR from the design stage. The initial objective of this project is to provide model-based approaches for protecting personal data by design and to gain legal compliance with some GDPR requirements. In particular, we focus on improving the currently adopted mechanisms to leverage them to enforce some of the GDPR's requirements by design. As mentioned before, we select an architectural approach (i.e., model-based) to implement privacy by design.

The first part of this thesis focus on developing an access control model to effectively capture the nature of information accessed and shared in OSNs. They are one of the most popular web-based services for people to communicate and share information with each other. OSNs might raise serious problems concerning users' privacy with all their benefits since their privacy policies settings and access control models still lack key elements. One privacy risk is caused by accessing and sharing co-owned data items. Another risk is caused by the privacy settings offered by OSNs that do not, in general, allow fine-grained enforcement, especially in cases where posted data items concern other users. We focus on enhancing the access control model by working on multiparty and fine-grained privacy policy features. Enriching the access control system in this way enables compliance by-design.

Due to the currently restrictive privacy regulations, software engineers are expected to design privacy preserving architectures for technological systems. Several techniques have been offered to aid system engineers (e.g., designers) in incorporating essential privacy concepts. However, those techniques are mostly based on security-oriented privacy concepts like anonymity and linkability rather than regulation-oriented requirements such as purpose limitation, retention time, or accountability.

In the second part of the thesis, we focus on proposing a framework that automatically embeds privacy requirements at the design level. We adopt DFDs as design models (i.e., software architecture) that can be developed to help software designers to verify if a design is compliant with the privacy regulations. We start this line of research by extending previous work that transforms DFDs into PA-DFDs through high-level graphical rules; neither a complete algorithm nor a reference implementation is given. Therefore, we provide algorithms to check and automatically transform DFDs into PA-DFDs. Moreover, we augment this transformation framework with a tool. We prove the correctness of our transformation model structurally.

The typical strategy to pragmatically sidestep massive DFDs is to begin with a high-level design consisting of composite processes, which are then refined into more detailed processes using a top-down methodology. Thus, we observe the necessity of relating various levels of abstraction. To do this, we provide a precise definition of refinement and a rigorous procedure for determining and obtaining suitable refinements that preserve essential attributes (e.g., privacy requirements).

As a first step towards our objective of modeling privacy requirements, particularly purpose limitation, in DFDs, we develop a rigorous mathematical framework for annotating DFDs with purpose labels and privacy signatures. This research is needed since our prior work on PA-DFDs considers purpose limitation only superficially, without elaborating on how the purpose of a DFD should be specified, verified, or inferred.

In summary, the research questions we address in this thesis are:

- What is the privacy risk caused by accessing and sharing co-owned data items in OSNs? (Paper A)
- What is the privacy risk driven by coarse-grained privacy policy settings offered by OSN, especially in cases of co-owned data items? (Paper A)
- How can OSNs' users collectively manage to view and share their co-owned data items? (Paper A)
- How can we alleviate some of the effort and the level of privacy expertise that software engineers need to design privacy-friendly systems using DFDs? (Paper C)
- How can we automatically incorporate certain privacy principles (i.e., GDPR provisions) in DFD to obtain privacy-aware DFDs? (Paper C)
- How to evaluate our enhanced transformation approach and prove the correctness of its properties? (Paper C)
- How can we preserve structural and functional properties between different levels of abstraction of DFDs? (Paper D)
- How can we define refinement for different levels of abstraction of privacy-aware DFDs? (Paper D)
- How do we develop a rigorous methodology to check and obtain suitable refinements to preserve relevant properties (e.g., privacy requirements)? (Paper D)
- How can we formulate and analyze purposes in design models such as DFDs? (Paper E)
- How do we perform automated checks for the purpose limitation principle in purpose annotated DFDs? (Paper E)

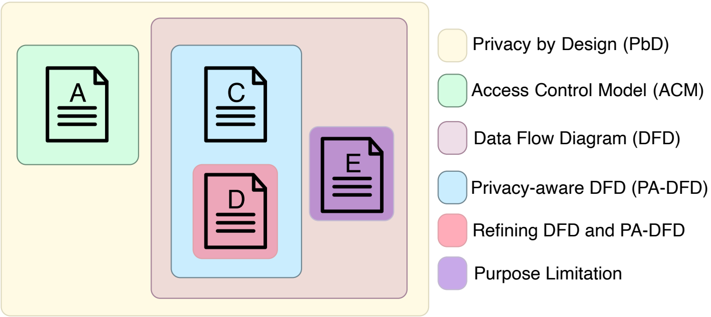

Figure II.1: Abstract illustration of the relationship between the papers

Papers

The papers included in this thesis cover several different topics, all related to privacy by design (PbD). Figure II.1 captures the relationship between the papers and their subjects. Paper A focuses on enhancing an access control model for accessing and sharing co-owned data items in OSNs.

Papers C, D and E provide privacy aware frameworks that leverage the notation of DFDs as design models. Papers C and D generally focus on developing the transformation framework (DFD to PA-DFD). Paper D defines a refinement and a rigorous process for determining and obtaining suitable refinements that preserve essential attributes. Paper E develops a formal method for modeling purpose limitation (PL) in design models (DFDs). Papers C, D and E deliver algorithms, reference implementations and case studies of the corresponding frameworks.

III

Summary and Contributions

This section provides summaries of the appended papers and outlines the personal contributions of each.

A A Collaborative Access Control Framework for Online Social Networks

Hanaa Alshareef, Raúl Pardo, Gerardo Schneider, Pablo Picazo-Sanchez

Our aim in this work is to provide a framework that empowers OSNs' users to collectively manage viewing and sharing their co-owned data items. As conflicting policies are commonly raised in multiple ownership privacy protection mechanisms, we proposed *Viewing* and *Sharing* aggregation-based algorithms which make a decision by solving potential conflicts between the different privacy settings of all the concerned users. This is achieved by taking into account the following aspects: the trust among users; the sensitivity level of users with respect to the concerned data item; and the weights of the following: (i) the types of *controllers* (those who are concerned in the decision that determines who can access a given data item and who cannot) and (ii) the types of *accessors* (those who are identified to access a given data item or not). We evaluated our solution by generating all possible combinations of components and performed experiments to show how the different components affect the decision on who should or should not access or share the data items. Furthermore, we provided proof-of-concept implementation into the open source OSN Diaspora.

Statement of contributions Hanaa contributed to proposing the collaborative access control model, formalizing the policies, and developing the collaborative access control algorithms. Moreover, she implemented the proof-of-concept prototype in Diaspora.

Appeared in: *Journal of Logical and Algebraic Methods in Programming (JLAMP), 2020.*

C Systematic Enhancement of Data Flow Diagrams with Privacy Checks

Hanaa Alshareef, Sandro Stucki, Gerardo Schneider

Recent legislation, such as GDPR, imposes tight restrictions on the handling of personal data. Privacy, like security, is a non-functional attribute; nonetheless, most software design tools, such as Data Flow Diagrams, are geared toward functional elements. A conceptual model was presented in prior research that DFDs may be developed into so-called privacy-aware DFDs to add specific privacy requirements to existing DFDs. In this paper, we develop algorithms that automatically convert DFD models into privacy-aware DFDs and a proof-of-concept implementation incorporated into a graphical tool for designing DFDs. This paper fulfills the practice of prior work that merely provided the concept of augmenting DFDs with privacy requirements and a very high-level transition between both models. Obtaining the algorithms (from the existing conceptual transformation) was not a simple operation, as several features of the transformation proved to be more intricate than anticipated and certain intuitions underpinning the high-level graphical transformation proved to be erroneous. We have addressed these conceptual flaws in our algorithms and evaluated them through theoretical evaluation and empirical evaluation. We demonstrated that the PA-DFDs produced by our transformation model possess desirable structural properties.

Statement of contributions Hanaa has contributed to building the transformation model, providing the algorithms and evaluating the model theoretically. Also, she was responsible for designing and implementing the tool and conducting the case studies.

Manuscript.

D Refining Privacy-Aware Data Flow Diagrams

Hanaa Alshareef, Sandro Stucki, Gerardo Schneider

When simulating real-world systems, B-DFDs and PA-DFDs may become excessively large. The typical technique to pragmatically overcome this challenge is to combine smaller processes using a bottom-up approach or to begin with a high-level design consisting of composite processes, which are then developed into more detailed processes using a top-down approach. In all instances, it is necessary to relate distinct levels of abstraction. In order to fulfill this, we propose a notion of refinement for both B-DFDs and PA-DFDs, formalizing the comparison of different degrees of abstraction of these design models. As far as we know, ours is the first formal definition of DFD refinement. For checking, finding and transforming refinements, we provide three different algorithms. We implement these algorithms in Python as part of a proof-of-concept (the tool called *DFD Refinery*) and apply it to a case study on an automated payment system.

Statement of contributions Hanaa contributed to structuring the refinement notion for DFD and PA-DFD and was also responsible for implementing DFD Refinery as a tool for our refinement framework and applying it to a practical application (an automated payment system).

Appeared in: *Proceedings of the 19th International Conference on Software Engineering and Formal Methods (SEFM 2021), online, December 2021.*

E Precise Analysis of Purpose Limitation in Data Flow Diagrams

Hanaa Alshareef, Katja Tuma, Sandro Stucki, Gerardo Schneider, Riccardo Scandariato

The primary purpose of DFDs is to model the functional features of a system. Recent research has demonstrated that DFDs can also be used to simulate non-functional properties, such as security and privacy, if they are annotated with the required security- and privacy-related requirements. Seven principles structure the approach to process personal data under Article 5(1) in the GDPR. Purpose limitation is one of these principles that are viewed as a fundamental building block for good data protection practice. Our prior work on privacy-aware DFDs (PA-DFDs) considered purpose limitation only vaguely, without elaborating on how the purpose of DFD activators and flows should be specified, validated, or inferred. This paper defines a rigorous mathematical framework for annotating DFDs with purpose labels and privacy signatures, checking the consistency of labels and signatures, and figuring labels from signatures. Our theoretical framework is implemented as a proof-of-concept tool comprising a domain-specific language (DSL) for providing privacy signatures and algorithms for validating and inferring purpose labels from such signatures. Finally, our framework and tool are assessed using the DFD of a fictional smart speaker system as a case study from the privacy literature.

Statement of contributions Hanaa contributed to modeling purposes in DFDs, providing the algorithms, and applying the case study. She was responsible for implementing the framework.

Appeared in: *The 17th International Conference on Availability, Reliability and Security (ARES 2022), Vienna, Austria, August 2022.*

Bibliography

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). `https://eur-lex.europa.eu/eli/reg/2016/679/oj`, 2016.

[2] A. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. In *International workshop on privacy enhancing technologies*, pages 36–58. Springer, 2006.

[3] S. Agostinelli, F. M. Maggi, A. Marrella, and F. Sapio. Achieving gdpr compliance of bpmn process models. In *International Conference on Advanced Information Systems Engineering*, pages 10–22. Springer, 2019.

[4] D. W. Allen, A. Berg, C. Berg, B. Markey-Towler, and J. Potts. Some economic consequences of the gdpr. *Allen DWE, Berg A, Berg C, Markey-Towler B and Potts J (2019)'Some Economic Consequences of the GDPR', Economics Bulletin*, 39(2):785–797, 2019.

[5] I. Altman. A conceptual analysis. *Environment and behavior*, 8(1):7–29, 1976.

[6] J. Anderson and F. Stajano. Must social networking conflict with privacy? *IEEE Security & Privacy*, 11(3):51–60, 2013.

[7] T. Antignac and D. L. Métayer. Privacy by design: From technologies to architectures. In *Annual privacy forum*, pages 1–17. Springer, 2014.

[8] T. Antignac, R. Scandariato, and G. Schneider. A privacy-aware conceptual model for handling personal data. In *ISoLA'16*, volume 9952, pages 942–957, Cham, 2016. Springer.

[9] T. Antignac, R. Scandariato, and G. Schneider. Privacy compliance via model transformations. In *EuroS&P Workshops'18*, pages 120–126, United Kingdom, 2018. IEEE.

[10] E. Arfelt, D. Basin, and S. Debois. Monitoring the gdpr. In *European Symposium on Research in Computer Security*, pages 681–699. Springer, 2019.

[11] U. G. Assembly. Universal declaration of human rights. *UN General Assembly*, 1948.

[12] C. Bartolini, S. Daoudagh, G. Lenzini, and E. Marchetti. Gdpr-based user stories in the access control perspective. In *International Conference on the Quality of Information and Communications Technology*, pages 3–17. Springer, 2019.

[13] C. Bartolini, S. Daoudagh, G. Lenzini, and E. Marchetti. Towards a lawful authorized access: A preliminary gdpr-based authorized access. *ICSOFT*, 2019:331–338, 2019.

[14] D. Basin, S. Debois, and T. Hildebrandt. On purpose and by necessity: compliance under the gdpr. In *International Conference on Financial Cryptography and Data Security*, pages 20–37. Springer, 2018.

[15] K. Bednar, S. Spiekermann, and M. Langheinrich. Engineering privacy by design: Are engineers ready to live up to the challenge? *The Information Society*, 35(3):122–142, 2019.

[16] E. Bertino and R. Sandhu. Database security-concepts, approaches, and challenges. *IEEE Transactions on Dependable and secure computing*, (1):2–19, 2005.

[17] J. Bhatia and T. D. Breaux. A data purpose case study of privacy policies. In *2017 IEEE 25th International Requirements Engineering Conference (RE)*, pages 394–399. IEEE, 2017.

[18] C. Bier, P. Birnstill, E. Krempel, H. Vagts, and J. Beyerer. Enhancing privacy by design from a developer's perspective. In *Annual Privacy Forum*, pages 73–85. Springer, 2012.

[19] M. Brodin. A framework for gdpr compliance for small-and medium-sized enterprises. *European Journal for Security Research*, 4(2):243–264, 2019.

[20] N. Bronson, Z. Amsden, G. Cabrera, P. Chakka, P. Dimov, H. Ding, J. Ferris, A. Giardullo, S. Kulkarni, H. Li, et al. {TAO}: Facebook's distributed data store for the social graph. In *Presented as part of the 2013 {USENIX} Annual Technical Conference ({USENIX}{ATC} 13)*, pages 49–60, 2013.

[21] B. Carminati, E. Ferrari, and A. Perego. Rule-based access control for social networks. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 1734–1744. Springer, 2006.

[22] A. Cavoukian et al. Privacy by design: The 7 foundational principles. *Information and privacy commissioner of Ontario, Canada*, 5:12, 2009.

[23] A. Cavoukian, S. Taylor, and M. E. Abrams. Privacy by design: essential for organizational accountability and strong business practices. *Identity in the Information Society*, 3(2):405–413, 2010.

[24] Y. Cheng, J. Park, and R. Sandhu. A user-to-user relationship-based access control model for online social networks. In *IFIP Annual Conference on Data and Applications Security and Privacy*, pages 8–24. Springer, 2012.

[25] P. Conroy, F. Milano, A. Narula, and R. Singhal. Building consumer trust: protecting personal data in the consumer product industry. *Deloitte Insights, November*, 13, 2014.

[26] M. Cramer, J. Pang, and Y. Zhang. A logical approach to restricting access in online social networks. In *Proceedings of the 20th ACM Symposium on Access Control Models and Technologies*, pages 75–86. ACM, 2015.

[27] J. Crampton and J. Sellwood. Path conditions and principal matching: a new approach to access control. In *Proceedings of the 19th ACM symposium on Access control models and technologies*, pages 187–198. ACM, 2014.

[28] B. Daigle and M. Khan. The eu general data protection regulation: an analysis of enforcement trends by eu data protection authorities. *J. Int'l Com. & Econ.*, pages 1–38, 2020.

[29] G. Danezis, J. Domingo-Ferrer, M. Hansen, J.-H. Hoepman, D. L. Metayer, R. Tirtea, and S. Schiffner. Privacy and data protection by design-from policy to engineering. *arXiv preprint arXiv:1501.03726*, 2015.

[30] S. Daoudagh and E. Marchetti. The gdpr compliance and access control systems: Challenges and research opportunities. 2022.

[31] S. Daoudagh, E. Marchetti, V. Savarino, J. B. Bernabe, J. García-Rodríguez, R. T. Moreno, J. A. Martinez, and A. F. Skarmeta. Data protection by design in the context of smart cities: A consent and access control proposal. *Sensors*, 21(21):7154, 2021.

[32] M. Davari and E. Bertino. Reactive access control systems. In *Proceedings of the 23nd ACM on Symposium on Access Control Models and Technologies*, pages 205–207, 2018.

[33] M. Davari and E. Bertino. Access control model extensions to support data privacy protection based on gdpr. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 4017–4024. IEEE, 2019.

[34] B. Debatin, J. P. Lovejoy, A.-K. Horn, and B. N. Hughes. Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of computer-mediated communication*, 15(1):83–108, 2009.

[35] J. W. DeCew. *In pursuit of privacy: Law, ethics, and the rise of technology*. Cornell University Press, 1997.

[36] T. DeMarco. Structure analysis and system specification. In *Pioneers and Their Contributions to Software Engineering*, pages 255–288. Springer, 1979.

[37] M. Deng, K. Wuyts, R. Scandariato, B. Preneel, and W. Joosen. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering*, 16(1):3–32, 2011.

[38] A. Dennis, B. H. Wixom, and R. M. Roth. *Systems analysis and design*. John wiley & sons, USA, 2018.

[39] E. 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 EC*, 23(6), 1995.

[40] EDPB. The cnil's restricted committee imposes a financial penalty of 50 million euros against google llc. `https://edpb.europa.eu/news/national-news/2019/cnils-restricted-committee-imposes-financial-penalty-50-million-euros_en#:~:text=On%2021%20January%202019%2C%20the,consent%20regarding%20the%20ads%20personalization`, 2019.

[41] EDPB. Guidelines 4/2019 on article 25 data protection by design and by default, 2019.

[42] E. Falkenberg, R. V. D. Pols, and T. V. D. Weide. Understanding process structure diagrams. *Information Systems*, pages 417 – 428, 1991.

[43] K. Fatema, C. Debruyne, D. Lewis, D. OSullivan, J. P. Morrison, and A.-A. Mazed. A semi-automated methodology for extracting access control rules from the european data protection directive. In *2016 IEEE Security and Privacy Workshops (SPW)*, pages 25–32. IEEE, 2016.

[44] R. Fogues, J. M. Such, A. Espinosa, and A. Garcia-Fornes. Open challenges in relationship-based privacy mechanisms for social network services. *International Journal of Human-Computer Interaction*, 31(5):350–370, 2015.

[45] P. W. Fong, M. Anwar, and Z. Zhao. A privacy preservation model for facebook-style social network systems. In *European Symposium on Research in Computer Security*, pages 303–320. Springer, 2009.

[46] P. W. Fong and I. Siahaan. Relationship-based access control policies and their policy languages. In *Proceedings of the 16th ACM symposium on Access control models and technologies*, pages 51–60. ACM, 2011.

[47] D.-G. for Communication. Special eurobarometer 431: Data protection. `http://data.europa.eu/euodp/en/data/dataset/S2075_83_1_431_ENG`, 2015.

[48] M. d. C. Freitas and M. Mira da Silva. Gdpr compliance in smes: There is much to be done. *Journal of Information Systems Engineering & Management*, 3(4):30, 2018.

[49] C. Gates. Access control requirements for web 2.0 security and privacy. *IEEE Web*, 2(0), 2007.

[50] D. Gollmann. Computer security. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(5):544–554, 2010.

[51] S. Gürses, C. Troncoso, and C. Diaz. Engineering privacy by design. *Computers, Privacy & Data Protection*, 14(3):25, 2011.

[52] I. Hadar, T. Hasson, O. Ayalon, E. Toch, M. Birnhack, S. Sherman, and A. Balissa. Privacy by designers: software developers' privacy mindset. *Empirical Software Engineering*, 23(1):259–289, 2018.

[53] Help Net Security. Only 20% of companies have fully completed their GDPR implementations. `https://www.helpnetsecurity.com/2018/07/16/complete-gdpr-implementation/`, 2018.

[54] J.-H. Hoepman. Privacy design strategies. In *IFIP International Information Security Conference*, pages 446–459. Springer, 2014.

[55] C. J. Hoofnagle, B. van der Sloot, and F. Z. Borgesius. The european union general data protection regulation: what it is and what it means. *Information & Communications Technology Law*, 28(1):65–98, 2019.

[56] D. J. Houghton and A. N. Joinson. Privacy, social network sites, and social relations. *Journal of Technology in Human Services*, 28(1-2):74–94, 2010.

[57] B. Howard. Analyzing online social networks. *Commun. ACM*, 51(11):14–16, Nov. 2008.

[58] M. Howard and S. Lipner. *The security development lifecycle*, volume 8. Microsoft Press Redmond, 2006.

[59] iapp. Cnil fines company 20k euros for illicit employee surveillance. `https://iapp.org/news/a/cnil-fines-company-20k-euros-for-illicit-employee-surveillance/`, 2019.

[60] R. Ibrahim and Y. Siow Yen. Formalization of the data flow diagram rules for consistency check. *IJSEA*, 2010.

[61] P. Ilia, I. Polakis, E. Athanasopoulos, F. Maggi, and S. Ioannidis. Face/off: Preventing privacy leakage from photos in social networks. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pages 781–792. ACM, 2015.

[62] C. R. Institute. Seizing the gdpr advantage: From mandate to high-value opportunity, 2018.

[63] M. Johnson, S. Egelman, and S. M. Bellovin. Facebook and privacy: it's complicated. In *Proceedings of the eighth symposium on usable privacy and security*, page 9. ACM, 2012.

[64] S. Joseph and M. Castan. *The international covenant on civil and political rights: cases, materials, and commentary.* Oxford University Press, 2013.

[65] P. Kamocki and A. Witt. Privacy by design and language resources. In *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC), May 11-16, 2020, Palais du Pharo, Marseille, France*, pages 3423–3427. European Language Resources Association, 2020.

[66] K. E. Kendall and J. E. Kendall. *Systems analysis and design.* Prentice Hall, 1999.

[67] D.-K. Kim, P. Mehta, and P. Gokhale. Describing access control models as design patterns using roles. In *Proceedings of the 2006 Conference on Pattern Languages of Programs*, pages 1–10. Association for Computing Machinery, 2006.

[68] B. Kostova, S. Gürses, and C. Troncoso. Privacy engineering meets software engineering. on the challenges of engineering privacy bydesign. *arXiv preprint arXiv:2007.08613*, 2020.

[69] P. G. Larsen, N. Plat, and H. Toetenel. A formal semantics of data flow diagrams. *Formal aspects of Computing*, pages 586–606, 1994.

[70] L. Lensdorf and U. Elteste. Real estate company fined € 14.5 million in germany for violating gdpr principle of privacy by design. `https://www.insideprivacy.com/eu-data-protection/real-estate-company-fined-e-14-5-million-in-germany-for-violating-gdpr-principle-of-privacy-by-design/`, 2019.

[71] H. Li, L. Yu, and W. He. The impact of gdpr on global technology development, 2019.

[72] J. Lindqvist. New challenges to personal data processing agreements: is the gdpr fit to deal with contract, accountability and liability in a world of the internet of things? *International journal of law and information technology*, 26(1):45–63, 2018.

[73] Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Analyzing facebook privacy settings: user expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 61–70. ACM, 2011.

[74] M. Madden. Privacy management on social media sites. *Pew Internet Report*, pages 1–20, 2012.

[75] M. Madden. Public perceptions of privacy and security in the post-snowden era. `http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/`, 2014.

[76] M. Madejski, M. Johnson, and S. M. Bellovin. A study of privacy settings errors in an online social network. In *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*, pages 340–345. IEEE, 2012.

[77] M. Madejski, M. L. Johnson, and S. M. Bellovin. The failure of online social network privacy settings. 2011.

[78] Y.-S. Martin and A. Kung. Methods and tools for gdpr compliance through privacy and data protection engineering. In *2018 IEEE European symposium on security and privacy workshops (EuroS&PW)*, pages 108–111. IEEE, 2018.

[79] Mention. The twitter engagement report 2018,users tagged, 2018. `https://mention.com/en/reports/twitter/users-tagged/`.

[80] E. C. of Human Rights. European convention on human rights. `https://www.echr.coe.int/Pages/home.aspx?p=basictexts&c`.

[81] E. Papadogiannakis, P. Papadopoulos, N. Kourtellis, and E. P. Markatos. User tracking in the post-cookie era: How websites bypass gdpr consent to track users. In *Proceedings of the Web Conference 2021*, pages 2130–2141, 2021.

[82] C. Perera, M. Barhamgi, A. K. Bandara, M. Ajmal, B. Price, and B. Nuseibeh. Designing privacy-aware internet of things applications. *Information Sciences*, 512:238–257, 2020.

[83] L. Piras, M. G. Al-Obeidallah, M. Pavlidis, H. Mouratidis, A. Tsohou, E. Magkos, A. Praitano, A. Iodice, and B. G.-N. Crespo. Defend dsm: a data scope management service for model-based privacy by design gdpr compliance. In *International Conference on Trust and Privacy in Digital Business*, pages 186–201. Springer, 2020.

[84] W. Presthus, H. Sørum, and L. R. Andersen. Gdpr compliance in norwegian companies. In *Norsk konferanse for organisasjoners bruk av IT (NOKOBIT). Svalbard, Norway*, pages 1–14, 2018.

[85] T. Priebe, E. B. Fernández, J. I. Mehlau, and G. Pernul. A pattern system for access control. In *Research Directions in Data and Applications Security XVIII*, pages 235–249. Springer, 2004.

[86] M. Rost and K. Bock. Privacy by design and the new protection goals. *DuD, January*, 2009:1–9, 2011.

[87] M. Saltarella, G. Desolda, and R. Lanzilotti. Privacy design strategies and the gdpr: A systematic literature review. In *International Conference on Human-Computer Interaction*, pages 241–257. Springer, 2021.

[88] P. Samarati and S. C. d. Vimercati. Access control: Policies, models, and mechanisms. In *International School on Foundations of Security Analysis and Design*, pages 137–196. Springer, 2000.

[89] R. S. Sandhu and P. Samarati. Access control: principle and practice. *IEEE communications magazine*, 32(9):40–48, 1994.

[90] M. Scott. Cambridge analytica helped ‘cheat’ brexit vote and us election, claims whistleblower, 2018.

[91] F. H. Semantha, S. Azam, K. C. Yeo, and B. Shanmugam. A systematic literature review on privacy by design in the healthcare sector. *Electronics*, 9(3):452, 2020.

[92] A. Shostack. *Threat modeling: Designing for security*. John Wiley & Sons, USA, 2014.

[93] L. Sion, D. V. Landuyt, K. Wuyts, and W. Joosen. Privacy risk assessment for data subject-aware threat modeling. In *IEEE Security and Privacy Workshops*, 2019.

[94] L. Sion, D. Van Landuyt, K. Yskout, and W. Joosen. Sparta: Security & privacy architecture through risk-driven threat assessment. In *2018 IEEE International Conference on Software Architecture Companion (ICSA-C)*, pages 89–92. IEEE, 2018.

[95] S. Sirur, J. R. Nurse, and H. Webb. Are we there yet? understanding the challenges faced in complying with the general data protection regulation (gdpr). In *Proceedings of the 2nd International Workshop on Multimedia Privacy and Security*, pages 88–95, 2018.

[96] A. Skendžić, B. Kovačić, and E. Tijan. General data protection regulation—protection of personal data in an organisation. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1370–1375. IEEE, 2018.

[97] D. J. Solove. *Understanding privacy*, volume 173. Harvard university press Cambridge, MA, 2008.

[98] S. Spiekermann. The challenges of privacy by design. *Communications of the ACM*, 55(7):38–40, 2012.

[99] F. Staff. Protecting consumer privacy in an era of rapid change–a proposed framework for businesses and policymakers. *Journal of Privacy and Confidentiality*, 3(1), 2011.

[100] J. M. Such and N. Criado. Multiparty privacy in social media. *Commun. ACM*, 61(8):74–81, 2018.

[101] J. M. Such, J. Porter, S. Preibusch, and A. Joinson. Photo privacy conflicts in social media: A large-scale empirical study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3821–3832. ACM, 2017.

[102] Y. Tao and C. Kung. Formal definition and verification of data flow diagrams. *Journal of Systems and Software*, pages 29–36, 1991.

[103] G. A. Teixeira, M. M. da Silva, and R. Pereira. The critical success factors of gdpr implementation: a systematic literature review. *Digital Policy, Regulation and Governance*, 2019.

[104] The European Commission. Internal Market, Industry, Entrepreneurship and SMEs. `https://ec.europa.eu/growth/smes_en#:~:text=Small%20and%20medium%2Dsized%20enterprises%20(SMEs)%20are%20the%20backbone,every%20sector%20of%20the%20economy`, 2018.

[105] C. Tikkinen-Piri, A. Rohunen, and J. Markkula. Eu general data protection regulation: Changes and implications for personal data collecting companies. *Computer Law & Security Review*, 34(1):134–153, 2018.

[106] W. Tolone, G.-J. Ahn, T. Pai, and S.-P. Hong. Access control in collaborative systems. *ACM Computing Surveys (CSUR)*, 37(1):29–41, 2005.

[107] K. Tuma, R. Scandariato, and M. Balliu. Flaws in flows: Unveiling design flaws via information flow analysis. In *ICSA'19*, pages 191–200. IEEE, 2019.

[108] S. D. Warren and L. D. Brandeis. Right to privacy. *Harv. L. Rev.*, 4:193, 1890.

[109] A. F. Westin and O. M. Ruebhausen. *Privacy and freedom*, volume 1. Atheneum New York, 1967.

[110] E. Wiese, A. Wykowska, J. Zwickel, and H. J. Müller. I see what you mean: how attentional selection is shaped by ascribing intentions to others. *PloS one*, 7(9):e45391, 2012.

[111] M. Woodman. Yourdon dataflow diagrams: a tool for disciplined requirements analysis. *Information and Software Technology*, 30(9):515–533, 1988.

[112] K. Wuyts and W. Joosen. Linddun privacy threat modeling: a tutorial. *CW Reports*, 2015.

[113] K. Wuyts, R. Scandariato, and W. Joosen. Empirical evaluation of a privacy-focused threat modeling methodology. *J. of Syst. and Soft.*, 96:122–138, 2014.

[114] K. Wuyts, L. Sion, and W. Joosen. Linddun go: A lightweight approach to privacy threat modeling. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 302–309, 2020.

[115] K. Wuyts, D. Van Landuyt, L. Sions, and W. Joosen. Linddun privacy engineering. `https://www.linddun.org/`, 2020.

[116] H. Xu. Reframing privacy 2.0 in online social network. *U. Pa. J. Const. L.*, 14:1077, 2011.

[117] C. Zhang, J. Sun, X. Zhu, and Y. Fang. Privacy and security for online social networks: challenges and opportunities. *IEEE network*, 24(4):13–18, 2010.