Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Bridging the Gap between Privacy Incidents and PETs

Master Thesis

Lena Csomor

April 5, 2023

Advisors: Prof. Dr. Kenny Paterson, Dr. Anwar Hithnawi,
Alexander Viand, Shannon Veitch

Applied Cryptography Group
Institute of Information Security
Department of Computer Science, ETH Zürich

Abstract

The loss of privacy, which the pervasiveness of the internet has enabled, has severe consequences for both individuals and society as a whole, and demands countermeasures. A serious lack of protection for the average internet consumer exposes them to potentially significant consequences both in the online and the physical world. Such issues have been studied to an extent from the security perspective but are far less understood regarding privacy.

This thesis introduces a novel, user-centric framework for realistic threat modeling, which is based on our own systemization of past privacy incidents, and serves to study the modus operandi of real-life attacks and Privacy Enhancing Technologies (PETs). To accomplish this, we developed a model of privacy-based attacks and utilized it to survey news articles reporting privacy incidents. Through our analysis, we established a comprehensive taxonomy of privacy-based attacks and systemized their data flows. By making these complex data flows manageable, comparable, studyable, and categorizable, we enable future researchers to build upon our work.

In a small study on relevant PETs and their impact on data flows, our resulting framework proved to be highly applicable and yielded valuable insights. We surveyed existing notions of privacy and approaches to threat modeling in PETs and discussed the gaps between assumed versus real-life attack vectors. This should help current and future research identify knowledge gaps on attack vectors, which parties, roles and actions to consider as well as to become aware of the harms to avoid. Overall, our framework and study offer a promising foundation for further research into privacy-based attacks and effective threat modeling.

Acknowledgements

I have way too many great people I need to thank because the number of humans involved in this project just kept growing.

I thank my supervisor Alex Viand, who has put up with me for a long time now. Thanks for believing in our ideas, staying curious and not being afraid to do this project. Thanks for always being patient and believing in me, too.

I also want to thank Dr. Anwar Hithnawi for her delightful optimism and believing in this work just as much.

A big thank you goes to Shannon Veitch who joined us along the ride and helped us determine the right path and untangle our minds.

Thank you to Bailey Kacsmar, who answered our call all the way from Waterloo, shared her wisdom generously and kept us from doing stupid things.

We all owe a thank you to Prof. Kenny Paterson, who allowed us to do this work and supported this fun bunch.

Finally, I thank my boyfriend, my family and my cat for their unwavering emotional support, not only during this thesis but the whole journey.

Contents

Chapter 1

Introduction

Privacy has gained visibility and importance in recent years, and both regulatory and consumer expectations are shifting towards demanding better protection for sensitive data. The loss of privacy, which the pervasiveness of the internet has enabled, has severe consequences for both individuals and society as a whole and demands countermeasures [21, 49, 31, 134].

Conflicting interests in terms of usability, personalization and privacy requirements make it difficult for the average consumer, as well as regulators or providers, to safely navigate the online world [79, 117, 29, 87]. While technological advances try to address these issues and protect the user, it has been repeatedly shown how these efforts fail to efficiently capture user's needs and concerns [102, 33]. For example, apps following best practices in online security might still lack desired features, such as the ability to lock the app with a password. This arises from mismatches between the threat models of privacy technologies and the attack vectors used by real-life attackers that users might be confronted with [50]. This mismatch results in a severe lack of protection for the average internet consumer, exposing them to significant consequences both in the online and the physical world. Such issues have been studied to an extent from the *security* perspective [102, 33] but are far less understood with regarding *privacy*.

Privacy Technologies. Nonetheless, a plethora of privacy technologies are currently being developed to aid the end-user, targeting a variety of settings and guarantees. These tools and techniques offer protections ranging from ad-hoc best effort (e.g., pseudonymization) to formally provable guarantees (e.g., differential privacy). Many approaches remain vigorously debated, with no sign of a consensus on what is and is not considered sufficient. The guarantees given by modern privacy-preserving technologies are highly non-trivial: formal definitions tend to be complex and are almost exclusively used in research, while intuitive ad-hoc guarantees make it hard to

understand real-world robustness. In addition, these guarantees are usually stated with respect to very specific threat models that focus on very specific aspects, do not compose with each other, and, as a result, fail to address privacy holistically.

These contained, abstract approaches stand in harsh contrast to the current online ecosystem's overwhelming complexity. Today's level of connectedness and the eroding lines between the physical and online world offer a vast, new attack surface that, from a privacy viewpoint, is under-researched. Preventatively assessing the real-life threats is challenging, as there is no "ground truth". User concerns might not correspond with the actual expected risks and harms due to either misunderstandings of the underlying technology or biased perspectives of the likelihood of different harms. For example, most users are very worried when entering credit card numbers when it is in fact background information such as addresses or family relations (e.g., mother's maiden name) that enable financial exploitation by undermining fraud detection mechanisms.

User-centric Privacy. While developers' design choices might be driven by technological aspects, regulatory demands, and liability concerns, it is unclear to what extent compliance actually prevents relevant privacy harms to the user. Consequently, the current state of technology forces trade-offs between safety, usability and functionality upon developers and users alike. They are then left guessing which "degree of privacy" could be appropriate to avoid potential harm.

We argue that for effective protection mechanisms, knowing the threats it should defend against precisely is crucial. Identifying privacy violations to their full extent has so far proven difficult, as both privacy as a concept and what is perceived as harm depends on context, social and cultural norms, and other factors that are difficult to assess objectively and may change over time.

However, some more significant consequences of such violations can easily be found in both media and research. Lay people's concerns cover a wide range of potential consequences, including legal action, economic discrimination, being stuck in a "bubble", or social embarrassment [87]. Concerns generally address physical, social, and psychological harms [78]. Unfortunately, a plethora of real-life examples of such harms can be found, e.g., Uber employees misusing company infrastructure to stalk ex-partners [67], or location data being legally purchased and used to identify and publicly harass specific individuals [21]. On a larger scale, national mass surveillance has become nearly ubiquitous [116, 31]. As famously stated by the former head of the NSA, Michael Hayden, governments "kill people based on metadata" [49].

There has been highly impactful work on how it is the result of inappropriate data flows that such harms become possible [92]. We observe that being aware of harms and consequences gives an idea of what the user might require protection from, but designing effective protection mechanisms requires knowing where the data flow could be altered. Thus, understanding the modus operandi of privacy-based attacks in terms of both data flows and consequences prove to be the crucial puzzle piece in effective threat modeling. This knowledge is the first step towards successfully navigating of trade-offs and incentives while working towards adequate privacy for the user.

Contributions. This thesis aims to identify, study, and systematize privacy violations and associated harms arising from the current lack of privacy of end-users and to analyze it in the context of existing privacy-preserving technologies. Towards this, we will provide a formal definition of privacy incidents, survey known incidents, and develop a visualization of attack data flows as well as a taxonomy of privacy violations and arising harms. Based on this, we compare attack patterns to the intended data flows of privacy technologies to discuss their effectiveness against real-life attacks, which fosters an understanding of weaknesses and blind spots in current design decisions made for PETs. This should help current and future research identify knowledge gaps on attack vectors, which parties, roles, and actions to consider, and to become aware of the harms to avoid. Our work re-evaluates and contextualizes existing approaches to threat modeling and discusses the gaps between assumed versus real-life attack vectors. This provides a new perspective on threat modeling that incorporates a user's lived experience, such as the influence of context and how privacy threats cause harm.

We begin by discussing existing approaches to privacy and providing our own formal model for privacy-based attacks in Chapter 4. This will then be employed to study and systemize past privacy incidents, which allow for the derivation of attack patterns and the beforementioned taxonomy in Chapter 5. To provide context on the current situation, in Chapter 6, we assess current threat models in PETs and their intended data flows with the framework derived in the previous chapters to demonstrate its practicality and to discuss the gaps and overlaps between the modus operandi of real-life attack and the threat modeling choices in PETs. We then sketch our ideas of the future of threat modeling in Chapter 7.

Chapter 2

Background

2.1 Introduction to Privacy

Privacy is a fundamental human right [2], but determining the appropriate amount of privacy has long been a subject of philosophical and legislative debate. In many cases, it is difficult to even determine whether privacy has been violated due to the cultural and personal aspects of privacy, which are context-dependent and evolve over time. Privacy is crucial to the social and self-aware nature of humans, and it is an essential aspect of almost every area of our lives. When privacy is violated, individuals may feel uncomfortable, exposed, and face serious consequences such as social exclusion or hate. Having privacy provides us with freedom, including the freedom to keep secrets, maintain different levels of closeness with different people, and make decisions about our lives. The primary goal of this thesis is to enhance the protection of privacy in today's world.

The internet has significantly impacted our privacy, as information can be retained and exposed indefinitely, regardless of its truthfulness. Few laws protect our privacy, and even fewer are enforceable. This thesis will examine the implications of this and possible solutions.

More formally, in this thesis, we distinguish between three related notions: privacy theory, privacy model, and privacy taxonomy. A privacy theory proposes an explanation or definition of the concept of privacy. A privacy model aims to represent the concept of privacy on a chosen level of abstraction. A privacy taxonomy is a set of definitions arranged in hierarchical order, which often overlaps or is part of a model. Note that these terms are often used interchangeably or overlap significantly in general usage [131].

2.2 Policies and Laws

Privacy is a human right captured in the European Convention of Human Rights of 1950 [2], and since then, this right has been incorporated into various local legislations with different implementations.

The General Data Protection Regulation (GDPR) [123] applies to all services operating in Europe and aims to offer a more comprehensive and expansive protection than previous legislations. It encompasses the important key principles of "fair and lawful processing, purpose limitation, and data minimisation and data retention" [9]. Data protection has to be implemented by design and by default, and consent to the processing of personal data has to be "freely given" [37]).

However, current practices of obtaining user consent and communicating privacy policies (which need not even be legally binding) are often inadequate and incomprehensible, rendering the use of the word "consent" questionable [86, 93]. Further, it is worth noting that many services are still non-compliant with GDPR and fail to employ the necessary measures to comply with the law [46, 65].

The California Consumer Privacy Act (CCPA) [8] is one of the most progressive privacy legislations in the U.S. It provides Californian residents with the right to know, delete, and opt-out of the sale of personal information, as well as non-discrimination for exercising these rights. However, it has several exceptions, including medical information and consumer credit reporting information, and it only applies to Californian residents [34, 8].

The U.S. aims for stronger legal protections of children and health data, as seen in laws such as the Children's Online Privacy Protection Rule (COPPA) [6] and the Health Insurance Portability and Accountability Act (HIPAA) of 1996 [10]. COPPA imposes additional requirements on online services that engage with children under 13 years old, while HIPAA sets strict regulations on health data. However, these laws often fail to address new technologies, leaving users vulnerable and unprotected [1]. The need for technical solutions that can be deployed faster than new legislation and at the same time support existing law's enforcement is highlighted by these ongoing issues [122]. However, it is important to note that these laws are a product of a cultural understanding of privacy and may shape the perception of users on these topics, potentially influencing the results of user studies.

2.3 Privacy-enhancing Technologies

Privacy-enhancing technologies cover a wide range of different settings and provide an even wider variety of guarantees. To describe patterns within a

set without revealing individual parts, statistical noise has proven a powerful tool. Differential Privacy [41] uses formal analysis of statistics to provide formal guarantees of privacy. However, these guarantees are expressed in a non-intuitive way, as bounds on changes in the probability of certain observations. Further, for data sharing and processing under strong privacy prerequisites, multiple technologies have evolved. Secure Multi-Party Computation (MPC) [44] allows mutually distrusting parties to compute over private data without having to release the underlying data. This can be used to replace existing or required trusted third parties. Fully Homomorphic Encryption (FHE), which allows computations over encrypted data without the interaction required in MPC [58], and Zero-Knowledge Proofs (ZKP), which allow one to prove knowledge of a secret without revealing it [18], can similarly be used to construct protocols that require less trust and expose less private information. In recent years, there has been an increasing number of applications of these and similar techniques to Machine Learning. In addition, techniques such as Federated Learning [80] have emerged that trade-off formal guarantees for more practical performance. Furthermore, there is a long history of practical protection systems that provide less formal guarantees but are usually significantly more efficient. These range from simple access control systems [48] to notions of anonymity [107], pseudonymization techniques [91] and other, frequently more ad-hoc approaches.

2.4 Threat Models

Threat modeling is a crucial process that involves identifying potential threats and vulnerabilities in a system, assessing their likelihood and impact, and devising countermeasures to mitigate them [5]. This process is carried out at different levels of technical abstraction throughout the development lifecycle, which includes the design, implementation, and deployment of technologies.

A comprehensive threat model typically includes a description of the technology itself and all the assumptions made about it, potential threats to the technology or its users, countermeasures to address those threats, and some form of validation to test the efficacy of the model, the threats, and the countermeasures. This involves identifying potential adversaries and their capabilities, as well as distinguishing between trusted and untrusted components that may be vulnerable to attack.

Frameworks such as LINDDUN [4] and the MITRE Privacy Threat Taxonomy [17] have been developed specifically to address privacy concerns during the deployment of existing technologies. However, these frameworks are not designed for use during the design stage of technologies.

Our research aims to fill this gap by developing a threat modeling frame-

work that supports the design of PETs. Unlike existing frameworks, which are focused on identifying vulnerabilities in existing systems, our framework is designed to help researchers compare their chosen privacy guarantees and adversary capabilities to real-world attack scenarios.

While some formal definitions exist in the threat modeling used for the design of PETs, there is no fixed set of notions or guarantees used in the process. This is because the field is constantly evolving to provide new guarantees in order to meet the needs of researchers who are developing new technologies. Our framework aims to support this process by providing a systematic, user-centric approach to identifying and mitigating threats that may arise in the real-world.

2.5 Qualitative Analysis and Coding

Qualitative analysis describes a group of research methods that investigate non-numerical, often unstructured data, usually in small quantities [106]. It is used to gain an understanding of a certain phenomenon that can not be investigated with numerical methods, such as people's perception or social reality. It mostly makes use of natural language sources such as interviews, books or reports. The available analytic strategies are coding, pattern thematic analysis and content analysis. This work will focus only on coding.

Qualitative coding is a process of categorizing unstructured data to identify themes and patterns for analysis. This involves assigning codes to excerpts of text such as paragraphs, sentences, phrases, or words. Qualitative coding enables researchers to systematically derive categorizations and identify patterns and themes.

There are two main approaches to qualitative coding: inductive coding and deductive coding. Inductive coding is a ground-up approach where codes are derived from the data without any preconceived notions. This is useful for exploratory research or when new theories, ideas, or concepts need to be developed. Deductive coding, on the other hand, is a top-down approach where codes are predetermined based on an existing research framework or theory. This is beneficial when a pre-determined structure is needed for the final findings, such as in program evaluation studies. A hybrid approach that combines inductive and deductive coding can also be used.

Since this research method operates on unstructured data, often in little quantities, and with natural language that leaves room for interpretation, this is not an exact science. However, it is sufficient to investigate trends and distributions, reveal patterns and analyze problems that could otherwise not be investigated at all, as they cannot be formulated numerically. Thus, while the results need to be interpreted with care, this research method is essential for complex problems with social components.

Chapter 3

Related Work

3.1 Privacy in Technologies

Security and privacy guarantees are commonly investigated for specific technologies, but the threats they identify can have far-reaching consequences beyond the individual tool. Machine learning [99], for example, is susceptible to adversarial inputs, inference and inaccurate models, which can result in wrongful convictions [68], disclosure of sensitive information [55], compromised research results [99], and even fatalities [90]. Private learning can mitigate some of these risks, but it introduces trade-offs between model accuracy, complexity, and resilience to attacks. Similar challenges exist in other technologies, such as drones [90], where optimizing security and privacy necessarily reduces them for society and vice versa. Our work examines online data flows from a user-centric perspective and acknowledges the recurring theme of trade-offs between privacy, usability, and effectiveness, and their potential extended consequences.

Navigating security and privacy threats requires metrics to evaluate relevant properties. While efforts have been made to quantify security risks from a corporate perspective [129], our work will mainly focus on end-user privacy and protective technologies, such as PETs. Privacy metrics are highly fragmented, but Wagner and Eckhoff [124] provide a framework to help researchers evaluate metrics based on adversary model, data source, input and output measures, and categorize PETs into privacy domains: communication systems, databases, location-based services, smart metering, social networks and genome privacy. Our analysis of past privacy incidents observes attacks in each of these domains, making this thesis valuable for evaluating PETs in similar scenarios.

3.2 User Harm, Risk and Behavior

As people increasingly view their devices as extensions of themselves [79], and their perceptions of privacy continue to evolve [63], it's crucial to comprehend their understanding and perceptions of threats, risks, and protective measures, to meet their needs in privacy and security [69].

Effective privacy research requires acknowledging that users' experiences are not uniform. It is crucial to assess the varying, potentially harmful impact of technology on different demographics. Previous research has identified at-risk demographics, whose risks for harm vary depending on the context and technology involved [125, 118]. Moreover, study participants from diverse demographics exhibit different understandings and perceptions of technology, influenced by factors such as technical literacy, experience, and socioeconomic status [94, 52, 81, 57, 85]. These insights reinforce our observations of past privacy incidents' targets and assist in identifying potential biases.

Laypeople's lack of understanding of online threats and harms, as well as their magnitude is concerning, as noted by Howe et al. [69]. Prior work found differences in the perceived severity of harms, noting that "expert participants had different insights on the potential implications of certain harms that general population social media users might never see or interact with if not impacted personally" [110]. Our qualitative analysis of numerous past events is intended to offer a framework for researchers, enabling them to minimize their own biases regarding the understanding and severity of online threats and harms.

As we seek to incorporate the actual needs of users, as well as their perception, it is imperative to examine their decision-making procedures when confronted with trade-offs regarding privacy. The privacy paradox, a phenomenon that describes discrepancies between users' self-reported attitudes towards privacy and their actual privacy behaviors [114, 59, 28], may be explained by the idea of "bounded rationality" [11]. This theory suggests that individuals may not be able to process the large amount of information required to make completely rational decisions about their privacy. Additionally, various contextual factors, such as spacial, temporal, and social considerations, have been demonstrated to significantly impact decision-making. [74, 105, 121]. Therefore, to address privacy and security concerns in an effective, user-centric manner as we do in this work, it's crucial to consider not only the diverse demographics and their varying relationship with technology, but also the social and contextual factors that affect their decision-making.

3.3 Mental Models and Misconceptions

The aim to create effective privacy features has motivated researchers to evaluate the disparity between their own and the users' perceptions of privacy [94]. Oates et al. [94] found that laypeople associate privacy with physical spaces they can control, while tech experts often think of data, indicating a divide between designer and user. Users rely on mental models to fill gaps in their knowledge about the services they use daily [73], but these models are often incomplete due to lack of technical expertise [71, 94]. Identifying weaknesses in both user awareness and technical protection mechanisms through real-life events allows us to work towards bridging this divide.

We observe that misconceptions about privacy are ubiquitous among internet users and can lead to unsafe practices. Gerber et al. [60] identified common privacy misconceptions, such as the belief that having nothing to hide or being unimportant would prevent data collection. Users also tend to overestimate their risk awareness [66] and underestimate dangers [127]. Kang et al. [76] investigated internet users, and found that most participants were aware that their data is shared among companies, but were unclear about who had their data, and had misconceptions about the sensitivity and visibility of data types, and the parties who could access their data. Other studies have identified similar misconceptions in older adults [52] and about web tracking and targeted ads [121, 87]. However, it is often unclear whether users perceive companies' data accesses and sharing practices as privacy violations, and if they expect consequences.

Researchers found that privacy-protecting technologies currently available do not adequately cater to the safety and comfort needs of users, both in the context of web tracking [87] and publicly shared online data [111]. To address this, Schnitzler et al. [111] suggested the development of adversarial models that take into account users' privacy fears and unintentional exposure in real data publishing scenarios, as well as the development of secure data sharing mechanisms that can function under such models. We have taken up this challenge by proposing realistic threat models that incorporate general concepts of online data flows and PETs from both technical and user perspectives.

3.4 User Threat Models

Realistic threat models can only be achieved when we incorporate the various ways in which technology is utilized and exploited, particularly in use cases that were not intended by design. Some user groups that are particularly prone to unexpected behavior are older adults and victims and perpetrators of domestic violence.

Older adults face a unique set of privacy challenges due to their dependence on others, their limited technical literacy, and frequent lack of ownership over the devices they use. As highlighted in [52], the need for frequent access to their personal data by medical professionals, caregivers, and family members poses challenges in balancing convenience and privacy. In addition, older adults are at a seemingly higher risk for medical and financial fraud and discrimination. The use of second-hand or public devices further amplifies the risks associated with limited technical literacy [52], which can lead older adults to avoid technology altogether.

Despite the challenges that older adults face, they are often overlooked in discussions of at-risk demographics because they may not be as active on the internet. However, given that society is rapidly aging, it is crucial to consider their situation in both user studies and technology development. Integrating older adults into the threat modeling and user experience processes could be a significant step towards better accessibility for all.

A detailed analysis of threat models in the field of Intimate Partner Violence [50] highlights the various methods used by attackers to exert control over their victims through access to personal data and devices. The attacker is formalized as an UI-bound adversary, "an authenticated but adversarial user of a victim's device or account who carries out attacks by interacting with the standard user interface, rather than through the installation of malicious or sophisticated softwaretools" [50]. The authors also note that attackers may leverage third parties to harm the victim, and that denying access to a device can be just as effective as compromising it. In our examination of past privacy incidents, we will categorize these attack vectors as "physical access" and recommend that interested readers consult this study for additional information.

Other works have also delved into privacy and security issues related to specific social use cases of technology. Wu et al. [130] provide a non-exhaustive list of such works, and investigate whether protective mechanisms exist for each use case. Their research further underscores the fact that current technology fails to meet users' needs due to inadequate threat models, which forces them to alter their behavior or exploit the system, potentially exposing them to additional security and privacy risks.

3.5 User Preferences and Choice

Users are generally opposed to the use of their data by companies [84, 76], but may be more willing to accept data usage if they perceive benefits [87, 121]. Despite this, it is still evident that users prefer to be asked and repeatedly be given a choice about their data when changing contexts [87, 121].

Policymakers support users' wishes for greater privacy control, but this is in opposition to advertising trade groups [84], creating obstacles that ultimately put users at risk. The disagreement leads to challenges regarding what users should be able to control, what the default settings should be, and who should design and deploy privacy mechanisms [84]. Nonetheless, privacy issues resulting from the transfer of browsing history, online behavior, and self-reported information to unknown third parties, the high re-identifiability of individuals in pseudonymized settings, and the possible harms remain largely unsolved. While the probability of harmful scenarios may be low, the significant user exposure increases the likelihood of some scenarios happening. Our analysis of past privacy incidents shows instances of all these problems, emphasizing the urgency of constructively addressing these privacy issues.

PETs have been introduced to address the conflict between privacy concerns and data processing desires, in the hope that PETs can reduce aversions against data usage and mitigate privacy issues at the same time [32, 75, 74]. Researchers thus follow the question under which circumstances users would agree to share their data [32, 75, 74]. Cummings et al. [32] investigated user perceptions of differential privacy and found that while some privacy concerns can be mitigated, user willingness to share data is mostly influenced by the specific description of the technology. Kacsmar et al. [74] found that private computation increased willingness to share data, but the context and transparency were again crucial factors in users' decision-making. The findings emphasize the crucial role of effective communication between technology designers and users in promoting the broader adoption of PETs.

Chapter 4

A Model for Privacy-based Attacks

The digitalization of everyday life has forced society to rethink their idea of privacy. This term is coined by cultural understanding, social expectations, law and philosophy. However, with technical advances, the longstanding spheres and spaces of people's lives had to undergo drastic changes. Various approaches try to describe and define this new reality with different concepts, the most common and popular of which we will list in the first part of this chapter.

We start with a deep dive into the different aspects that shape privacy, and have led to definition attempts in the past. Ultimately, experts have found that a single definition can not be sufficient to encompass the complexity of privacy and have turned to broader frameworks that allow them to model the space in form of taxonomies or flows. We describe the most important of these frameworks in this chapter.

In reviewing these ideas, we find that existing frameworks are capable of describing privacy violations, but not attacks based on them. This motivates our development of a new model for privacy-based attacks which appropriately captures the complexity of modern attacks. Further, we have found that many works in the intersection of technology and privacy rely on trivial models for attacks that can not capture the complexity of modern, internet-based, heavily connected attacks, and miss representing all relevant data flows, thus possibly overlooking both dangers and possibilities to counter them. The strength of our model lies in its applicability to various contexts such as research papers, informal reports and documentations, and the comparability between individual attacks it creates. It further appropriately captures the whole data flow that constitutes the attack. We manage to do so with a conveniently restricted set of roles and actions that ensures easy and fast utilization.

4.1 Theories of Privacy

In this thesis, we must clearly differentiate privacy from security, safety and other similar concepts in order to properly define our scope. Thus, we first need to understand privacy and the novel ways in which the privacy of the user is impacted by today's technological possibilities. This will then allow us to expertly discuss privacy violations and their consequences. Most theories of privacy stem from before or right at the beginning of the Information Age (historical period from mid-20th century to now) and have been discussed in terms of legislation and philosophy. We focus on work from the last century from western countries (North America and Western Europe), as the developments in that time and region are most relevant to us. This is because most of our chosen data sources in terms of research and current developments focus on this geographical region, and thus have to be understood in context of their cultural background.

Solove, a professor of law well-known for his work on privacy in the context of information technology, has examined the most impactful works on privacy in modern times and systemized them into 6 categories [113], which we introduce and summarize in the following paragraphs. In search of a definition for privacy he came to the impactful conclusion, nowadays shared by many in the field, that the philosophical idea of "family resemblance" can be applied to privacy [113]. This means that the many areas that make up privacy are not connected by one single characteristic, but rather by many overlapping similarities that do not share one common feature. This also holds for the 6 categories he identified, which considerably overlap, but neither alone would be sufficient, and they also do not all share the same features. As the online world, consisting entirely of data, is more restricted, we will also discuss these categories in our context and settle on a definition.

Right to be let alone. In their pioneering work "The Right to Privacy", Warren and Brandeis [126] are the first to advocate for a right to privacy in the U.S., where they define it as the "right to be let alone". One of the most famous achievements of this notion is that warrantless wiretapping of persons was ruled a violation of the U.S. Fourth Amendment, according to a Supreme Court decision in 1967 [3] invoking the right to be left alone. This overturned an earlier Supreme Court rule that denoted wiretapping to not be a constitution of search and seizure.

With today's communication frequently taking place over text messages, which frequently remain stored on devices unencrypted, or in some cases even on companies' servers, an individual's past communication remains accessible for the government in the future. At the same time, masses of tracking data are available for sale to the government as well. It is fairly easy to re-identify individuals in these data sets, thus rendering almost ev-

eryone subject to mass surveillance. We argue that the definition we choose for this work shall indeed encompass mass surveillance as a violation of privacy. However, the right to be let alone falls short of the aspect that not only such highly intrusive practices, but more subtle aspects of social lives, such as selective disclosure and the right to make certain decisions, require protection, as we will discuss next.

Limited access to the self. This concept can be understood as a protection from *unwanted* access in the form of physical access, personal information, or attention [20]. This concept is different from the right to be let alone as it is not equivalent to solitude, but opens the circle to a chosen set of people. The crucial aspect of choice is brought into this notion, for example by Ernest Van Den Haag, arguing that "the right to privacy entitles one to exclude others [...]" [27]. We also note that privacy comes into existence only in relationship with society [89], as without other people we would have no desire for privacy.

The notion is particularly interesting in the context of information inference through aggregation and non-consensual data collection, as people might willingly share certain data with certain parties without anticipating their collusion and capability to retrieve more information than the person gave originally. We do not reject this notion, but argue that, in our case, it requires more refinement. Ruth Gavison suitably argues that this concept contain elements of secrecy, anonymity and solitude [56], which we will again address in the following paragraphs.

Secrecy. As mentioned above, secrecy, the concealment of information, can be seen as a subset of limited access to the self. It further extends the right to be left alone, which can be seen as a protection from invasive practices. It denotes a privacy interest that for when information is obtained, it should not be used against the data subject's will [101]. In court, this idea has been used in cases such as Roe v. Wade and Whalen v. Roe, where privacy included the avoidance of disclosure, including information about "certain kinds of important decisions". The court ruled for individual independence in such matters, referring to the constitutional "zone of privacy". Further, the conclusion that what is no longer completely secret must be completely public has been contradicted frequently by experts, including a ruling of the U.S. Supreme Court, who has however not been consistent on the matter [120]. Thus, privacy should not only encompass nondisclosure, but also selective disclosure, as this is a lot closer to lived reality of most people [77].

This notion is not only relevant when it comes to bodily autonomy, but also regarding the non-consensual publication of content such as revenge porn, or the exaggerated spread of statements during a context collapse. Further, the right to use proper end-to-end encryption can also be debated in this

context. Selective disclosure, though a vital part of people's social networks, proves to be difficult on the internet. Information is retained for a long time while people's attitudes towards each other may change, and society, the corporate world and governments are keeping track.

Control over personal information. Alan Westin's pioneering work [128] was one of the first to consider data privacy and protection in 1967. His new definition of privacy reads 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"*. It can thus be seen as a subset of limited access to the self. Westin further elaborates that information should be treated as property, an idea nowadays well-established already in intellectual property laws.

However, this falls short of the realization that information can be generated in relationships and does not exclusively belong to one person or the other. Further, experts have argued that privacy does not need to be applicable to all information, but only that which is deemed "intimate" [70]. It also only includes information, and not the right to make certain decisions. Thus, for a general notion of privacy, Westin's definition is deemed insufficient.

For our work on information flows however, it encompasses very well the privacy expectations and available protection mechanisms, as the concept of control allows not only the interception but alteration of data flows. We can not anticipate which information will reveal "intimate" details in the future, and thus do not restrict ourselves to only "sensitive" data. We operate on a limited space, as many things that would not be "data" in the physical world are converted to data in the online realm, for example in the form of emojis to convey emotions. Further, users can mostly choose whether and how they want to be present online, while they can not escape the physical world. At the same time, the scope they operate in is versatile, as users can be both trapped in a bubble or exposed to the whole world wide web. This, and the fact that almost everything on the internet is preserved somehow, increases the importance of control and underlines the suitability of Westin's definition. A lack of control over their own data can have a detrimental impact on a user, thanks to the enhanced scope and exaggerated retention of the internet.

Control over data also gives a user a certain control over their narrative online, which further overlaps with the next category.

Personhood. Defining privacy as personhood is more abstract than the notions above, but is often used to underline the importance of privacy itself, and to examine which aspects of the self are generally perceived as private or worthy of protection. Based on philosopher Stanley Benn's ideas, Solove

notes that because surveillance leads to self-censorship and inhibition, personhood could be "defined in terms of the individual's capacity to choose" and relates this back to respect for personhood [113]. This independence to make choices has been upheld in several Supreme Court decisions, bringing forth the following explanation: "At the heart of liberty is the right to define one's own concept of existence, of meaning, of the universe and the mystery of human life. Beliefs about these matters could not define the attributes of personhood were they formed under compulsion of the State" [95]. Some have argued that in fact, these thoughts relate to liberty and autonomy and not to privacy [109]. This has been countered by noting that these concepts can overlap and do not need to be exclusive [36]. However, this cannot obscure the fact that there is no consensus on a definition for personhood.

Intimacy. As a definition for privacy, intimacy again relates back to our ability to limit the access of others to our self. It further implies that forming personal relationships is tied to privacy, going as far as the claim that "intimate relationships simply could not exist if we did not continue to insist on privacy for them" [62] Gerety defines intimacy as "the consciousness of the mind in its access to its own and other bodies and minds, insofar, at least, as these are generally or specifically secluded from the access of the uninvited" [61]. However, this notion fails to address issues of privacy that do not contain relationships, such as computer databases [103].

In conclusion, we have seen how privacy as a general concept is hard to grasp with a single definition. Thanks to our limited scope, we are however able to stick with Westin's definition, acknowledging that it can overlap with others. We rephrase it as follows:

Definition 4.1 *Privacy is an entity's ability to control how, when, and to what extent personal information about it is communicated to others.*

This definition will be used across various text-based media, including news articles, research papers, court protocols, and legal texts, all of which are related to privacy and technology. The definition is intentionally general so that it can be applied to a wide range of sources while also being concise enough for an entity to determine whether their privacy is being preserved. Due to the nature of technology, privacy violations can be more easily identified compared to non-digital contexts. Therefore, Definition 4.1 is well-suited for our purposes as it enables us to identify situations where an entity has lost control over their personal information, which we refer to as *privacy violations*.

From an outside perspective, a privacy violation can only be identified when the information is noticeably used in a way the entity would not allow, if it still had control. It can happen that an entity's privacy is violated without it knowing. It is important to note that a privacy violation can also be

noticeable. Solove ultimately suggests to conceptualize privacy within contexts, which means adopting multiple conceptions of privacy instead of a single one. This led him to later propose a privacy taxonomy, which we will consider in extension of our definition and discuss it in 4.2.1.

4.2 Models of Privacy

We find that recent takes on privacy are more versatile and have been developed with information technologies in mind, which is why we will explore them in extension to our definition. The provided frameworks or taxonomies are broader than single definitions and are meant to provide guidance where previous, rigid definitions have failed. The two famous examples we will discuss in the following help us better understand how privacy can be violated in the digital realm, both from a harm and a data flow point of view. They give a rough idea of what exactly it is that we are trying to avoid or protect, which we will later investigate in detail by analyzing real incidents. This will support us in the development of a model for privacy-based attacks.

We will first review Solove's taxonomy of privacy [112], which focuses on the consequences of different privacy violations. Then, we will discuss Nissenbaum's influental work about mass surveillance and information flows, which leads her to ultimately model privacy as "contextual integrity" [92].

4.2.1 Solove's Taxonomy of Privacy

Solove [112] argues that different privacy violations are of different severity, introducing both the need for multiple conceptions instead of one, and the idea to focus on privacy-related harms instead of a definition. He consequently provides a taxonomy as a framework for legislators and researchers.

The taxonomy consists of four categories of "harmful activities": information collection, information processing, information dissemination, and invasion. The first category, information collection, includes both surveillance and interrogation. The second category, information processing, concerns the extraction of additional information from data through aggregation or identification. The taxonomy also addresses insecurity regarding information handling, secondary use of information collected for a particular purpose, and exclusion of the data subject from the data processing.

The third category, dissemination of information, deals with spreading information beyond its intended context. In addition to breaches of confidentiality, distortion, and increased accessibility, Solove distinguishes between disclosure and exposure and incorporates specific threats such as blackmail and appropriation.

Finally, the fourth category, invasion, is an activity that does not necessarily require personal information but still enables intrusion and decisional interference. Solove views this taxonomy as a catalog of harmful activities, which can aid in the identification and prevention of privacy-related harms.

We find that in this taxonomy, a violation of privacy and the resulting harm are strictly tied together. There is no distinction between certain parties or actions, but only a view on the outcome of a specific scenario. We argue that for threat modeling, such thinking will not be sufficient. Practical threat modeling requires knowing where to intercept or alter data flows in order to prevent harm. Further, we find harm and privacy violations need not be so closely correlated, as privacy can be violated by one party - without notice of the user - and observable harm can be caused by another. While both actions may have their own legal implications, they are important to distinguish from a technical viewpoint.

We have identified that the privacy violation itself does not need to be noticeable when discussing definition 4.1. We argue that *privacy harm* is the noticeable consequence of a privacy violation. An easy example in the context of Solove's taxonomy is that information dissemination is clearly a privacy violation: an entity loses control over how its data is shared. This in and of itself does not need to cause damage to said entity or even be noticeable by anyone except the disseminator. If, however, said information dissemination results in reputational damage for the entity, then the reputational damage is a privacy harm caused by a privacy violation. In the following, we will often use the terms "privacy harm" and "harm" interchangeably, as all harms we will consider come from privacy-based attacks. We define:

Definition 4.2 *Privacy Harm is the noticeable consequence of a violation of privacy to the entity whose privacy was violated, where it is sufficient if notice is taken by anyone except the causer of the privacy violation and the causer of the harm.*

4.2.2 Nissenbaum's Privacy as Contextual Integrity

Nissenbaum [92] acknowledges that recent technologies have dramatically enhanced certain entities' capabilities in information collection, analysis and dissemination. Her work focuses especially on the example of public (mass) surveillance, and argues for an alternative benchmark of privacy. The ideas are purposefully limited to information about people and do not try to encompass privacy as a whole. Nissenbaum argues that the public discourse about privacy is failing to recognize the problems of public surveillance, and thus aims to create a framework for prescribing restrictions on collection, use and dissemination of data.

An integral realization is that "there are no arenas of life not governed by norms of information flow, not information or spheres of life for which 'any-

thing goes' " [92]. Nissenbaum proposes contextual integrity, which models information privacy as maintenance of an appropriate flow of information. An information flow is considered appropriate if it does not violate contextual information norms. These include the expectations and social norms surrounding the data type, data subject, the parties involved in sending and receiving data, as well as how the data is transferred. Nissenbaum summarizes these contextual information norms in norms of appropriateness and norms of distribution. Contextual integrity is maintained as long as both hold. Norms of appropriateness include the social acceptance of the revelation of certain information types in specific situations. Information that is considered appropriate to reveal in one context can constitute a privacy violation in another. Norms of distribution considers that there are expected norms on information flow, and that information cannot be freely revealed in any arbitrary situation because it has once been revealed within a specific context. The great strength of contextual integrity lies in its flexibility to adapt to changing norms and sociotechnical developments.

We observe that while following norms can provide flexibility in social and legal contexts, they hardly make sense formalized in threat models for technologies that might serve internationally, across different demographics, cultures and laws. This however nicely highlights the necessity of user's choice and granular control, as to allow the user to ensure their own appropriate information flows.

4.3 Our Model of Privacy-based Attacks

Solove and Nissenbaum's frameworks address the privacy violation and the resulting harm, but an effective strategy for preventing harm requires understanding the entire data flow, from its origin to its endpoint, in order to determine the most effective protection strategies. To efficiently abstract the data flow, parties, and actions in the attack, it is crucial to have an initial model of all possibly involved roles, for both comparability and mapping purposes between different attacks. We will draw from Solove's list of harmful activities and Nissenbaum's emphasis on appropriate information flows as inspiration. Finally, we desire our definition of privacy-based attacks to be an easily applicable model instead of a formal definition to facilitate the analysis. Rather than using oversimplified victim-perpetrator models, our approach aims to capture complex, real-world information flows through a role-action model. This will appropriately describe the various actors involved in privacy-based attacks, including their corresponding roles and actions. This model, and the analysis we can conduct based on it, will aid the development more nuanced and effective strategies for both analyzing and addressing privacy violations.

Scope. This model has been developed for privacy-based attacks under definition 4.1 and has not been tested beyond the scope of this work. Definition 4.1 is general enough to be applicable to all the topics we will consider, but concise enough to know whether it is intact for an entity. However, this definition can not be employed to cleanly separate security and privacy. For this work, we exclude all violations of privacy that exist due to a lack of or insufficient security, thus excluding not only security incidents but the overlap between security and privacy. Further details and the practical implications of this distinction are detailed in 5.1.3.

4.3.1 Development

To create our model, we start with the following simple observation visualized in Figure 4.1. This figure shows the data flow, parties and actions

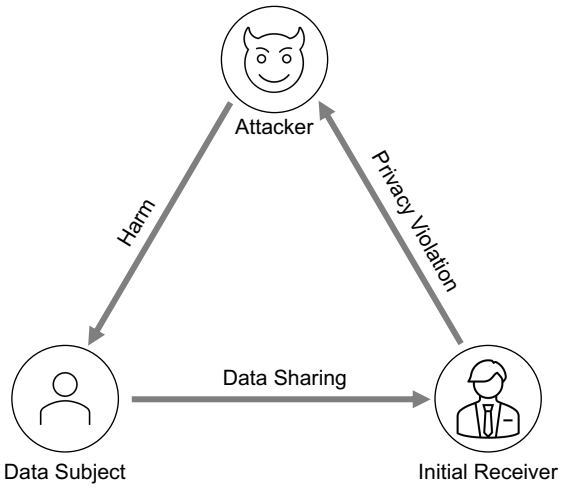

Figure 4.1: Trivial 3-Party Model Describing a Privacy-based Attack: Parties are indicated by nodes, data flows in the direction of the arrows while actions, denoted by the text above the arrows, can be initiated by either party

involved in a privacy-based attack. It is privacy-based, because it is enabled by a privacy violation, in the scope we have defined in 4.3. It is an attack because it causes harm to one of the parties. The arrows in the diagram indicate the direction of the data flow and are captioned with the type of action each pair of connected parties perform. In terms of the actions, which each require a passive and active part, the direction can go either way.

We have defined earlier how privacy violation and harm can be separated from each other which is why we already introduce three parties: a data subject, an initial receiver and an attacker. We show the necessity for three parties with an example: Alice (the data subject) shares an intimate picture

with Bob (the initial receiver), trusting him to keep it between them. Bob however shares the picture with Charlie (the attacker). This action is clearly a privacy violation, as Alice has had her picture shared beyond her intent and thus lost control over her data. Charlie, as a consequence of this privacy violation, uses the picture to blackmail Alice, which causes her distress (the harm) and reveals the privacy violation that has happened.

We will later provide clear definitions for these parties and actions, but first we will describe why this model is still insufficient. We argue that this model does not describe all roles and actions that influence and enable a privacy-based attack. Since we are interested in the data flow of the attack, and altering it to prevent harm, it is necessary for us to encompass all steps between data origin and harm. We thus need to identify the missing parts and provide a new model, and we do so by providing counterexamples where the trivial model fails to capture all relevant parties. Then, we create an extended model that can appropriately depict the attack, and repeat. We observe that it is not necessary that the attacker is also involved in the privacy violation, as the original violation can be only the enabling action of a longer data flow, where the attacker is at the end. Thus, we require an additional party between the initial receiver and the attacker. Further, the data subject themselves sharing their data is too strong of an assumption. It is possible that the data subject was not involved in the gathering and distribution of the data, or if the data subject was not legally capable of consenting to actions on their data. Thus, we require a party between the data subject and the initial receiver.

In line with definition 4.1, the term ”data” refers to an arbitrary type of personal information: it may be information about or of an entity, including pictures, whereabouts, relationships, communication, mental, physical, social or financial status as well as data commonly known as personal, personally identifiable or sensitive data.

4.3.2 Final Model

The above discussion finally allows us to derive a final model, and we provide a detailed description of the parties, actions and scope in the following. Note that this model has been developed for privacy-based attacks under the exclusion of security. If we allowed security in scope, an attacker could perform alterations to the data flow, interfere with any other party and their actions, which is not captured by our model.

We did not include further parties that simply pass on data, with their actions not being a “first”. For example, there could be multiple parties between the Initial Receiver and the Attacker, but they would all employ the same role of the Data Handler, and add nothing to the scenario, as the privacy violation has happened, but the attack not yet. There could of course

also be multiples of every party. There are several options how this would be categorized:

- They operate as a group, and can thus be viewed as one entity.
- They do not operate as a group, but after the first entity of those multiple parties nothing relevant is changed in the attack flow (think of these entities as a chain), in which case we would simply ignore every entity but the first.
- They do not operate as a group, and the different entities perform very different operations on the data, resulting in the model to diverge further as each party in turn interacts with different parties than the others (think of this attack flow as a tree), in which case we would consider each entity to be part of a different attack, and thus they should not occur in the same instance of the model.

Parties

- **Data Subject**: The entity who can be identified or is identifiable directly or indirectly by reference to the personal data in circulation.
- **Initial Sharer**: The first entity sharing the data in circulation. It is not the same as the data subject if the data subject was not involved in the gathering and distribution of the data, or if the data subject was not legally capable of consenting to actions on their data.

  Example: A child is photographed on private property and the picture is shared. Since the child cannot legally consent to having its picture taken, and by being on private property has not willingly exposed itself to the public, it has not been involved in the gathering or distribution of the data. Therefore, the initial sharer is the photographer.

  Counterexample: Someone takes a picture of an adult person on a bus and shares it. Since the person has willingly exposed themselves to the public, they have consented to "sharing" their presence, but only for the scope of the bus ride. The photographer thus directly violates the person's privacy through exaggerated retention (taking the picture) and disclosure (sharing the picture) of data, and in that case holds the role of the initial receiver.
- **Initial Receiver**: The first entity who receives the data in circulation and causes a privacy violation.
- **Data Handler**: One or multiple entities between the initial receiver and the attacker, who receive and pass on data.
- **Attacker**: The entity who uses the data in circulation to cause harm on the data subject.

Actions

- **Data Access**: The receiving entity takes action to obtain the data from an adjacent party. This can either be the case between Data Subject and Initial Sharer, or between Data Handler and Attacker. This can but does not need to be a privacy violation.

  Examples: An entity might be buying data, legally requesting data access, or using physical access to acquire data.

  Counterexample: A party hands data to the receiving entity unprompted. In that case, the receiving entity remains passive in initializing the interaction. This example would qualify as Data Sharing.

- **Data Sharing**: The giving entity takes action to provide the data to another party. This is the case when the Initial Sharer initializes contact to the Initial Receiver and sends them data.

- **Privacy Violation**: An action or inaction that causes the data subject to lose control about how, when, and to what extent personal information about it is communicated to others.

  Example: This can happen through unauthorized collection, use or disclosure of data, or the retention and processing on inaccurate data.

  Counterexample: We assume parties have control over themselves and their devices, and consider otherwise as a security problem. This excludes stories where spyware is secretly installed on devices belonging to the subject, or tracking devices being planted on a person or their belongings. It does however not necessarily include access to the subject and their devices, such as a person reading through their spouse's chats when they've been given access to the device to read a recipe online, as this is an unauthorized collection of data and not a security breach.

- **Harm**: Action that causes physical, mental or other kind of damage or worsening of the state of the affected party.

This concludes our definitions for privacy and privacy-based attacks, which now fulfill our prerequisites. We thus propose the following model in Figure 4.2. The model includes the parties and actions discussed above in the order they occur in a privacy-based attack, which starts with the data subject. The data flows along the arrows, in a way that is detailed by the associated action. We can use this model to identify even partially described privacy-based attacks by matching the definitions of parties and actions to what is described in the data source. In Chapter 5, we demonstrate that this model is in fact exhaustive in capturing all parties and actions involved in privacy-based attacks by using it to analyze more than 100 attacks.

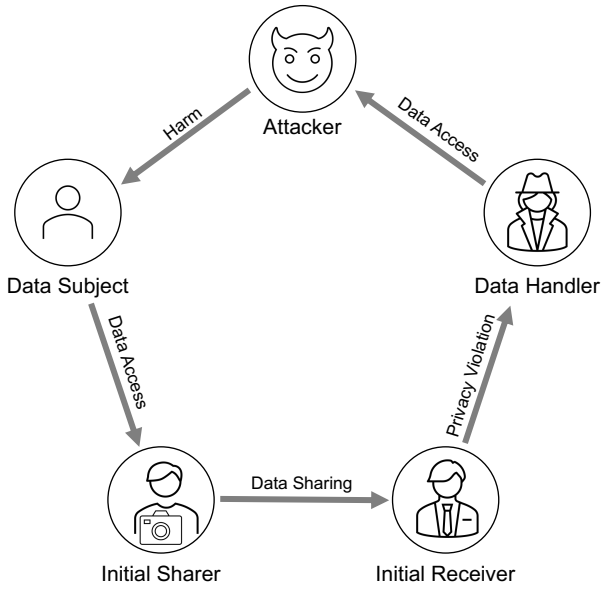

Figure 4.2: Final 5-Party Model Describing a Privacy-based Attack: Parties are indicated by nodes, data flows along in the direction of the arrows while actions, denoted by the text above the arrows, can be initiated by either party

Chapter 5

Analysis of Past Privacy Incidents

A fundamental contribution of this thesis is the systemization of past privacy incidents. We use the term privacy incidents interchangeably with privacy-based attacks. This entails scenarios that qualify as privacy-based attacks according to the model and definitions provided in Chapter 4. We focus on privacy incidents, which are incidents that contain a privacy-based attack by our definition and are confirmed to have happened by a data source.

This chapter systematizes real-world attacks by visualizing their modus operandi and listing the adverse consequences for users. The objective of this chapter is to enhance the understanding of the attack landscape, which can guide the development of better protective measures. We go beyond re-evaluating and contextualizing existing threat modeling approaches by providing a new perspective that integrates the user's lived experience. Our comprehensive framework enables realistic and user-centric threat modeling and helps identify and prevent threats, making it relevant for developing effective PETs. As a result, this work is a valuable contribution to the field of privacy research and PET development.

In the following sections, we will introduce the methodology used to achieve this systemization in Section 5.1, share the results of our work in Section 5.2 and discuss them in Section 5.3. We use the definitions and the privacy-based attack model derived in Chapter 4 to guarantee consistency throughout our work.

5.1 Methodology

In the following sections, we will provide a detailed account of our methodology. Firstly, we will outline our rationale for our choice of data source, specify the scope, and explain the approach used to collect the data. Subsequently, we will delve into the particulars of the systemization procedure,

which involves filtering the data and subjecting it to hybrid coding for qualitative analysis. We were able to deduce the categories to represent the parties and actions in the privacy-based attack model, as well as relevant context by identifying common themes and patterns.

We argue that this methodology is a robust and effective way to systemize past privacy incidents toward understanding their data flows and context. To create a comprehensive collection of incidents that allow for a rigorous investigation of privacy-based attacks, we carefully selected both our data source and scope. The hybrid coding approach we employ for qualitative analysis is particularly effective because it allows us to identify important context and nuances, which other, more rigid methods may miss. Thus, our systemization reflects the nature of privacy-based attacks and facilitates the design of effective protective measures for users.

5.1.1 Scope

In the following, we will focus solely on privacy and not security (such as defense against data breaches, attacks by hackers, or malware), since intact security can help prevent privacy violations but is not always sufficient. This is especially true when data is made public, sold, or directly exploited by the data collector. We are interested in privacy threats that do not require security breaches, as these kinds of threats are less studied than their security counterparts and can make life dangerously easy for adversaries.

We rely on our definitions of privacy, privacy violations, and model for privacy-based attacks, as provided in Chapter 4, to accurately investigate privacy-related topics. The colloquial use of privacy does not provide a well-defined concept, as it is often influenced by cultural understanding, laws, and social expectations and thus is not suited for our qualitative analysis. We require these definitions because they are well-suited for investigating PETs and personal information and allow us to guarantee reproducible and extendable results. They are general enough to apply to all the topics we will be exploring yet concise enough to provide a clear understanding of whether privacy is intact for an entity.

5.1.2 Data Source

We collect data with the goal of understanding the patterns and relationships in the data flow of privacy-based attacks. This requires a sample of attacks that reveal the underlying patterns. The time constraint of this thesis compels us to maintain a reasonable sample size and use a data source with an appropriate level of abstraction, providing enough information to observe the data flow but with as few additional details as possible. To ac-

complish this, we must select a data source with an adequate sample size that can be assumed to cover a substantial part of possible attack paths.

For these reasons, we decided to employ newspaper articles about privacy-based attacks as our data source. Thus, they will serve as a proxy for the real world since no single data source can capture every aspect of real-world complexity. News articles are available in high numbers and account for a variety of topics, with reports rarely longer than a few pages. Further, they also mention incidents that have not been brought to prosecution and, therefore, also contain minor or not necessarily illegal attacks. This is important because, currently, legislation is often not yet suited for technology-enabled attacks [96].

To guarantee the inclusion of stories from a broad scope, we will focus on the digital archives of the American newspapers 'The New York Times" and "Vice" in this initial effort. The New York Times is one of the most read newspapers in the world and holds more Pulitzer prices than any other organization [47], while Vice's content is historically more focused on pop culture and covers more serious topics with varying depth, objectivity claims and political stances [83].

5.1.3 Data Collection

In the following paragraphs, we outline the phases that we have employed to ensure that our data source provides an insightful and relevant sample of privacy-based attacks. By insightful, we mean that the sample should encompass a diverse range of attacks, while by relevant, we imply that the sample should cover a significant proportion of the possible attacks that could be identified using this specific data source.

Filtering

We obtain a relevant sample of articles through a three-stage filtering procedure that we apply to the online archives of the selected news agencies. The process consists of three consecutive stages: keyword search and two levels of content filtering. The steps described below are designed to be applied to any news archive, making it easy to replicate and extend our work.

Our filtering procedure is highly efficient because, at each stage, we reject a large portion of the available candidates. This helps us save time while still obtaining a comprehensive and meaningful sample.

Keyword First, we apply the keyword "privacy" as a search term in the online archives of both newspapers in the time frame between October 2019 and October 2022, which resulted in collectively 6547 candidates.

Applying a keyword without the category filters provided on the online archive pages guarantees a broad sample. This is because technology is not always in the focus in articles that contain privacy incidents, especially if people of public interest, severe crimes like murder, or large-scale lawsuits are involved. Consequently, articles could end up in other categories than expected. We assume that if an article focuses on events caused or enabled by personal data, the article will either contain the word "privacy" or be tagged accordingly. We chose the mentioned time frame to retain the attacks' practical scope and relevance thanks to their recency.

In the next step, we define detailed prerequisites for our analysis and apply them to the news articles, first only to the abstract, then to the full content of the remaining articles.

Requirements Our goal requires us to find attacks that involve privacy violations. Thus, the articles we choose for our data set should contain those. We, therefore, formulate high-level inclusion and exclusion criteria, which allow quick filtering and give a good indication of whether an article contains a technology-enabled privacy incident.

As *inclusion criteria*, the article must reference a privacy violation, problem, threat, or incident; an attacker or harm that pertains to privacy; and finally, the issue must be caused or facilitated by technology. If none of these topics were mentioned, then the article does not mention vital components of a privacy-based attack, and either does not contain the type of incident we search for or lacks data and can not be thoroughly analyzed. Examples are listed in Table 5.1. If the article does not include each of the five content requirements, then we exclude it from the sample.

We further provide six *exclusion criteria* for effective filtering, and articles are discarded if they contain any of the following:

- **Legislation and Policies**: Articles containing debates over or changes of legislation or policies are not incidents. Many of these articles mention hypothetical or potential scenarios, which we are not interested in since they could distort our analysis with imaginary attacks.
- **Security topics**: Many newspapers do not properly separate security and privacy. Articles on security-related content, such as data breaches, hacks, and malware, are explicitly excluded. We only consider scenarios where the subject of the article has control over their self, including their own devices, as mentioned in the description of Section 4.3.2.
- **Tech & Tools**: Reports on new technologies are not incidents. Many of these articles mention hypothetical or potential scenarios, which we

Content	Example Indicators
Threats	Alarming, concern, threat, fear
Problems	Problem, worry, issue, questionable, controversial
Relation to Privacy	Data-privacy, data handling, privacy problem, privacy concern, internet privacy, online privacy, data disclosure, data sharing, data selling, data protection, data collection, tracking
Harms	Identity theft, surveillance, doxxing, stalking
Relation to Technology	Use of app, tool, device or software, online activity like online tracking, data type like messages, browser history, location data

Table 5.1: Indicators for Required Content

are not interested in since they could distort our analysis with imaginary attacks.

- **Tutorials**: "How to be safe. . . ", "How to protect your privacy. . . ", "The best xy for z" are not incidents. These kinds of articles sometimes mention incidents later in the article, but these are often found in separate articles.

- **Political Debates**: Frequently heavily opinionated, one again has to heed hypothetical (drastic, apocalyptic) scenarios.

- **Duplicates**: We consider an article a duplicate if all the actors and the harm are the same. We keep the article with more details in our data set, which is usually the one published the latest. This does not distort our analysis, as the incident itself remains the same, just with fewer "unknowns".

An initial iteration considers only the title and abstract of the article and discards those that do not meet the criteria described above. 298 candidates persist after this step. In a second iteration, we employ the criteria on the whole text of the remaining articles. We are left with 102 articles for our analysis. This is a good data set size for our methodology since more common attack patterns will show as trends, but rare attacks will not become invisible in analysis or visualization. A detailed enumeration can be found in Table 5.2.

Newspaper	Keyword "Privacy"	Title & Abstract	Full Text
New York Times	4030	184	58
Vice	2517	114	44
Total	6'547	298	102

Table 5.2: Number of Relevant News Articles by Filter Stage and Newspaper

5.1.4 Qualitative Analysis

Our ultimate objective is to derive coherent definitions and consistent patterns across the dataset toward a systemization of the modus operandi of real-life attacks. We aim to produce comprehensive results that can serve as valuable reference for future researchers who wish to expand on our findings. To achieve this, we adopt a methodology that minimizes assumptions and prioritizes clarity and coherence in our results.

Our choice of data source, in combination with our type of goal, demands qualitative research methods that provide us with a comprehensive understanding of attack distribution and components. Qualitative analysis enables us to retain the context and intricate details of attacks, revealing previously unknown patterns.

Qualitative Coding. We effectively deduce categories and identify patterns from the available data by employing qualitative coding, a prominent tool in qualitative analysis [106]. During this process, we map "codes" to specific excerpts from the data. In our approach, we use the terms "codes" and "categories" interchangeably to better reflect their role in our methodology. These codes usually comprise summative words or brief phrases that describe the excerpt. We identify excerpts that mention relevant parties or actions in the attack by rereading the articles' content and then tagging them with descriptive codes. Further, we identify and reflect standout features for additional context, which is essential to provide insightful conclusions about the nature of the attacks being investigated. These codes enable us to detail and contextualize each party and action in our privacy-based attack model, ensuring that the resulting codebook is comprehensive.

Hybrid Coding. Our coding approach is a hybrid, incorporating both deductive and inductive methods. Inductive coding involves grouping data excerpts into themes and inferring codes from them, while deductive coding is used when the researcher already knows which codes to apply to the data.

In our study, we begin with a set of deductive codes based solely on the parties and actions in the privacy-based attack model. We then extend this set inductively to explore more detailed subcategories and the socio-technical context of the attacks.

The codes have a hierarchical structure, where the deductive level tells us where we are in the model, and the first inductive level is meant to be mutually exclusive and as exhaustive as possible, given our bottom-up approach to derive them. Further lower-level categories need neither be exclusive nor exhaustive and are simply meant to provide additional context. We purposely avoided extending the codebook with additional codes that we were only expecting to see but did not, as doing so could have distorted our interpretation and made it harder to recognize any limitations in our approach in terms of bias. The only exception is for categories indicating a lack of detail in an article, which we noted for the sake of completeness.

Conflict Resolution. As it is important to avoid unclear or incoherent codes, we aim to capture attacks with as few assumptions as possible and proactively counteract conflicting codes. If we find that codes do not work well (e.g., do not fit in any of our predefined categories, or very similar situations being sorted into different categories), that will be because of incorrect assumptions. We resolve this by only creating codes that describe what is (not) there and group them into higher-level codes later when we are ready to derive definitions for the patterns we are observing. If a pattern proves to be wrong (by counterexample, e.g. finding a code that should fit in the pattern but does not), we break it back down into the lower-level codes and start over. This ultimately gives us coherent definitions and consistent codes over the whole data set.

The hybrid approach allows us to retain comprehensiveness and relevance thanks to the privacy-based attack model that is used for the deductive set of codes and gain new insights, contextual relations, and attack patterns through the inductive part. The codes and their number of occurrences enable the close observation of characteristics and the inference of patterns in the attacks. For this work to be extended, we suggest that a future researcher refer to our codebook in Section 5.2.2.

5.2 Results

In this section, we delve into analyzing data flows in privacy-based attacks. Our goal is to identify where and how users require protection from harm, focusing specifically on data flows enabled through privacy violations.

In the following subsections, we will review our observed data flows in multiple steps, starting with the code definitions in Section 5.2.1 and noteworthy

observations regarding the model of privacy-based attacks in Section 5.2.2 and Section 5.2.3. This will provide context for the data flow visualizations discussed later on in Section 5.2.6. We emphasize that our qualitative analysis is intended to provide an overall understanding of the distribution rather than a detailed frequency analysis. Moreover, we will highlight standout features in our coding to provide insightful conclusions about the nature of the attacks investigated.

This section is crucial to fully grasp the implications of our research findings and how they can inform threat modeling and design decisions for PETs. By identifying areas where data flows facilitate attacks, we aim to support the development of effective protective measures that prioritize users' privacy. We further re-evaluate and contextualize existing approaches to threat modeling and provide a new perspective incorporating the user's lived experience. Moreover, an inspection of our analysis offers insights into how and where to control the user's data flow and the potential consequences that may arise from design tradeoffs.

5.2.1 Codebook

In the following paragraphs, we present the codebook (Section 5.2.2) that we derived during our qualitative analysis of privacy-based attacks. This codebook serves as a guide for identifying and categorizing different aspects of privacy-based attacks and is derived from 114 news articles, from which we deduced 122 code categories.

The codebook serves as a valuable tool, as it allows for the systematization of attack data flows from other data sources in the same manner as we did. The top-level categories in the codebook should be universally applicable to data sources in the English language, as they are tied to our model of privacy-based attacks, which has been developed independently from the qualitative analysis. For instance, this facilitates surveys of trend developments in attacks as the data flow distributions change over time. Note that the lower-level categories are more closely tied to the narrative of news articles. Consequently, they may not be as easily applied to other sources, and further extensions of these categories could be required.

The codebook not only allows for the collection of context in attacks but also enables realistic and user-centric threat modeling, which is essential for creating more effective PETs. By defining roles and attributes of actors/parties involved in attacks, assigning fitting and realistic actions, and understanding how context influences roles, actions, threats, and parties, the codebook provides a comprehensive framework for identifying and protecting against threats. The codebook's ability to collect important context from attacks with a minimal yet almost exhaustive set of options helps to improve the understanding of the attack landscape, which can guide the development

of better protective measures. This makes the codebook, together with our model for privacy-based attacks, and the data flow visualizations created as a result, a relevant contribution to the field of privacy research and PET development.

5.2.2 Parties

We were able to exhaustively describe all parties with a surprisingly small but descriptive set of categories. We found the narrative of articles to heavily distinguish between the victim and all other involved parties, which is reflected in our coding. Thus, we discuss them separately in the following paragraphs.

Data Subject.

Due to the stark differences in reporting on "celebrities" and "common people", we chose to categorize data subjects accordingly and emphasize again the significance of demographics in risk assessments in Section 3.2. This is reflected in our data which shows that 87% of the cases involved common people while 13% involved celebrities, which is noteworthy as the proportion of celebrities in the general population is much lower. This suggests a disproportionate number of reports on harm inflicted upon celebrities, which is unsurprising as the fate of celebrities is widely considered "newsworthy".

Other Parties.

For all remaining parties, we identified that the categories of "personal connection", "stranger", "company", and "government" were the most expressive, achieve a practical level of abstraction, and preserve critical characteristics. Frequently mentioned parties and their features, such as brand names, relationship status, or governing bodies, were repeatedly sorted into groups until a practical number of common features could be isolated, resulting in these categories. For us, this proved to be exhaustive, with only one article lacking sufficient details to determine any code at all, as can be seen in Table 5.4.

The versatility of privacy-based attacks is showcased by our findings of cases for all combinations of parties and categories. The respective combinations are not uniformly distributed. The data reveals that the majority of initial receivers are companies (72.3%), which are often used as intermediaries for other companies (43% of attackers) and governments (32.5% of attackers). This emphasizes the high value and power that can be extracted from user data and the interests of both corporations and governments in accessing it. We will provide more insight on this and the small number of individ-

Area	Code	Description
Data Subject	Celebrity	Entity of broad public recognition within one or multiple communities (politicians, musicians, athletes, etc.)
	Common People	Not a celebrity as defined above
	Other	
Initial Sharer	Self	The data subject
	Personal Connection	Acquainted with the data subject
	Stranger	Unknown to the data subject
	Company	Business Organization
	Government	Governing authority, including its legislative, judicial, and executive branches
	Unknown	Lack of Detail in Article
First Sharing	Voluntary	Knowingly and without pressure
	Necessary	Compulsory Opt-In
	- No Opt-Out Possibility	Not required for functionality but still compulsory
	- Required for Service	Required for functionality
	- Forced to Use Service	Use of service is compulsory to function within society
	Involuntary	Data is gathered through coercion
	Unknowing	User is not aware data is shared
	- Taken without force	Data is taken from a (passive) user, not shared willingly
	- Unknowing Opt-In	User was not comprehensibly asked for consent
	Unknown	Lack of Detail in Article
Initial Receiver	Personal Connection	Acquainted with the data subject
	Stranger	Unknown to the data subject
	Company	Business Organization
	Government	Governing authority including its legislative, judicative and executive organs
	Unknown	Lack of Detail in Article

Continued on next page

continued from previous page

Area	Code	Description
Privacy Violation	Accuracy of Data	Data is dated, not accurate or incomplete
	- Exaggerated Retention	Out-of-date information is kept and used
	Collection of Data	Information is collected without or outside of authority
	- Without Consent	Collector has never obtained any consent to gather any data
	Disclosure of Data	Unauthorized disclosure
	- To Third Parties	Disclosure is restricted to specific parties
	- Publish Data	Disclosure is not restricted
	- Outside Intended Scope	Limited disclosure outside agreed or expected scope
	Use of Data	Data already in possession is used without authority
	- Without Consent	Data is taken with consent and processed without
	- Forcefully acquired data	Data is taken and processed without consent
	- Outside intended scope	Processing outside agreed or expected scope
Data Handler	Personal Connection	Acquainted with the data subject
	Stranger	Unknown to the data subject
	Company	Business Organization
	Government	Governing authority including its legislative, judicial and executive branches
	Unknown	Lack of Detail in Article
	Same as Initial Receiver	
Attacker Motivation	Intentional	A goal is reached through purposefully causing harm, or the harm itself is the goal
	- False Positive	Attack hits the wrong target
	- Financial Gain	Harm is caused or knowingly accepted for profit
	Collateral	Negligence, incompetence, ignorance or selfishness cause damage
	None	No bad intentions or incompetence

Continued on next page

continued from previous page

Area	Code	Description
Attacker Data Access	Existing	No need to create access as attacker, e.g. app data access by distributor, employee access to business data
	Financial	Buying data, including paying services to use SDKs
	Public	Data is publicly available
	Legal	Access through legal system, e.g. court orders, warrants, special privileges
	Physical	Direct access to device or victim
	Unknown	Lack of Detail in Article
Attacker	Personal Connection	Acquainted with the data subject
	Stranger	Unknown to the data subject
	Company	Business Organization
	Government	Governing authority including its legislative, judicial and executive branches
	Unknown	
	Same as Initial Receiver	
	Same as Data Handler	
Attacker Skill	High	Vast resources, power or money, e.g. capability to write own software, request warrants or re-identify people in databases
	Medium	Skilled but not very resourceful, e.g. deploy existing software, easy social manipulation or impersonation
	Low	Neither skilled nor resourceful, e.g. online trolls, using popular plug-and-play technologies, abusing existing privileges in a company
	Unknown	Lack of Detail in Article

Continued on next page

continued from previous page

Area	Code	Description
Target	Targeted	Victim is selected as aim of attack
	Filtered	Group with shared attribute (which is the filter), or a subset thereof, is aim of attack
	Untargeted	No or very little aim is employed for attack (e.g. mass surveillance)
	Unknown	Lack of Detail in Article
Data Type	Visual Content	e.g. Pictures, Videos, Social Media Posts
	Technical Info	e.g. IP Address, Timestamps
	Biometric Data	Unique physical characteristics like face or fingerprints
	All Data on Device	Adversary has full access to unlocked device
	DNA	Digitalized version or physical presence of DNA
	Communication	e.g. Text messages, email, phone calls
	Public Records	Publicly available data
	Usage	Implicit or explicit data from app usage, e.g. tracking, behavioral data, browser history, financial transactions
	Location Data	e.g. location history, path, current location
	Medical Information	e.g. Patient record, prescriptions
	Other Personal Data	e.g. Criminal record, address, sexual orientation, self-declared information
	Unknown	Lack of Detail in Article

Continued on next page

continued from previous page

Area	Code	Description
Harm	Targeted Ads	Online behavioral advertisement, online tracking or interactions resulting in personalized forms of advertisement
	Mass Surveillance	Digital or physical, continuous or selective forms of mass surveillance
	Legal Prosecution	Legal investigation or conviction
	Financial Harm	Direct or indirect noticeable financial repercussions
	- Job Loss	Getting or handing in notice
	- Fin. Discrimination	e.g. reduced bonus or salary, job loss, hardship getting a job
	- Data Loss	Losing data of monetary value, e.g. access to accounts and devices
	Social Harm	Societal Consequences including but not limited to the below
	- Doxxing	Unwanted publication of private or identifying information
	- Online Exposure	Being ridiculed for involuntarily published data
	- Defamation	Reputational Damage
	- Fear for Reputation	Fear of reputational damage
	Physical Harm	Physical harm to property, self, or personal connections
	- Break-in	Forced entry to a locked space
	- Restriction	e.g. No-Fly order, travel ban
	- Murder	Attacker is directly or indirectly responsible for a murder
	Psychological Harm	All kinds of trauma, including but not limited to the harms listed in the following
	- Harassment	Online or offline harassment and hate
	- Intrusion in private life	e.g. unsolicited phone calls, reveal of highly sensitive data
	- Stalking	Online or offline forms of unwanted harassment or persecution
	- Mental Illness	e.g. depression, anxiety, PTSD, suicidality
	- Emotional Distress	e.g. fear, shame, embarrassment, feeling violated
	- Threats	Attacker threatens and frightens the Data Subject

Continued on next page

continued from previous page

Area	Code	Description
Potential Harm	Financial, Psychological, Social, Physical Harm, Mass Surveillance, Legal Prosecution, Targeted Ads	All the same harms are possible
Legality	Legal	Privacy Violation is legal
	Illegal	Privacy Violation is illegal
	Legal after lawsuit	Privacy Violation ruled legal
	Illegal after lawsuit	Privacy Violation ruled illegal
	Ongoing lawsuit	Privacy Violation caused ongoing lawsuit
	Lawsuit settled	Privacy Violation caused lawsuit that has been settled without court decision
	Unknown	Lack of Detail in Article
Victim Behavior	Noticeable actions	Victim does something noticeable, e.g. performing onstage, having a criminal record, going to a protest
	Everyday online interactions	Victim uses internet and devices as expected
	Everyday offline interactions	Victim behaves inconspicuously
	Compliance with law enforc.	Victim hands data to law enforcement in order to comply
	Unknown	Lack of Detail in Article

Table 5.3: Past Privacy Incident Analysis Codebook: The codes' hierarchical structure is indicated by indentation. Depending on the complexity of the news article and the mentioned attack, multiple instances of the same code category may have been present in an article, and occurrences of codes in a specific are may not sum up to the amount of news articles analyzed. Single entities can take on multiple roles.

ual initial receivers (8.4%) and attackers (22.8%) in the following discussion (Section 5.3).

Party	Category				
	Personal Connection	Stranger	Company	Government	Unknown
Initial Sharer (not self)	2.7% (3)	0.9% (1)	2.7% (3)	2.7% (3)	0.9% (1)
Initial Receiver	3.4% (4)	5% (6)	72.3% (86)	19.3% (23)	0
Data Handler	2.7% (3)	13.3% (15)	59.3% (67)	24.8% (28)	0
Attacker	3.5% (4)	19.3% (22)	43% (49)	32.5% (37)	0

Table 5.4: Occurrences of categories per party (except Data Subject) in absolute number and percentage relative to the other categories' occurrences for an individual party

5.2.3 Actions

Our codebook reveals that the different actions in the privacy-based attack models each individually require a set of categories. We will detail their characteristics in the following, as well as their distributions, to reveal insights and trends which will be of importance in user-centric threat modeling with our framework.

First Sharing

In the categories for the "First Sharing" action, it is interesting to see that the categories (voluntary, involuntary, unknowing, and necessary), deduced out of news articles' narratives, reflect the user's ability to choose, an important subject in a user's comfort, as discussed in Section 3.5.

The significant amount of unknowing data sharing (46.6%) and the requirement for substantial data sharing by services to work as intended (33.9%) have implications for user-centric PETs. These findings, detailed in Table 5.5, highlight the need for protective technologies deployed on the user's end to both educate the user about their situation and potential alternatives, and to provide them with more control and choice. Further investigations are necessary to understand why and how this data is distributed, which will help determine the most effective and comprehensive way to empower users with more control over their data. Intentional and voluntary sharing of data was rarely found to cause harm in the inspected articles (7.6%).

Our research suggests that the initial sharing of data typically involves a lower incidence of unknowing sharing and a higher incidence of coercion than we would typically expect *outside* of an attack scenario. Specifically, 11% of reports mentioned coercion in this context. It is worth noting that the phenomenon of coercion is inherently linked to an attack context since

it implies the use of violence or the threat thereof. In contrast, unknowing sharing may not always result in harm, but the fact that it remains a significant factor in attacks still highlights the need for increased attention to this type of data sharing.

Privacy Violations

The coincidence of our deduced top-level categories for privacy violations, *accuracy*, *collection*, *disclosure*, and *use of data* with the Canadian legislation on the matter [97] provides additional confidence in the thoroughness of our categories beyond our empirical study.

If we compare these categories, detailed in Table 5.6, to Solove's taxonomy [112], we find, however, that one of his categories appears to be missing: invasion. This is a consequence of us using a more complex model for privacy-based attacks, and this part of the taxonomy can still be found in our results. In our model, invasion is represented by "Involuntary" first sharing, which also works for Solove's respective subcategories like decisional interference.

Our analysis shows that disclosure to third parties constitutes the most common type of privacy violation, accounting for 47.9% of the cases, proving that it is enough to take data outside of its intended context to do harm without making it completely public. For threat modeling, this implies that designers must look at the potential consequences of data being shared in unintended contexts, and how this could be avoided most effectively. Consequently, this issue brings up the topic of contextual integrity, as discussed in Section 4.2.2.

It is not only the presence but also the absence of subcategories in Table 5.6 that is intriguing, as in the case of "Accuracy of Data". This category is rare (2.5%), and we can identify only the subcategory "Exaggerated Retention", but not other inaccuracies like invented or falsified data. This is surprising, as inaccurate data could potentially cause significant harm, such as false accusations and reputational damage. Since inaccurate data was so infrequent in our data set, we suggest that investigating this specific privacy violation would require a much larger data set. Beyond potential biases in reporting, we hypothesize that data traded between companies and governments could for instance be either extremely accurate or completely wrong, with little in between. If the data is extremely accurate, it may not be prone to errors that could result in a privacy violation based on inaccurate data. On the other hand, if the data is entirely wrong, it may not be useful or harmful to anyone, thus reducing the likelihood of any reported privacy violations. However, this is just one potential explanation and further analysis is needed to fully understand the reasons behind the limited number of reported data inaccuracies. Given its assumed potential for harm, we strongly encourage future work to investigate the matter.

Attacker Data Access

We were easily able to deduce a closed set of actions for how the attacker can access data, as the possible paths repeated very frequently and the news articles mostly abstained from technical details. Thus, we adopted what the articles deemed to be relevant features in the attack path. Note that technical access that did not exist prior to the attack is not listed in Table 5.7, as in all observed cases, we found such an access to be a security problem and thus out of our scope. Since formal threat models usually only consider accesses by technical means, this set reveals a broad palette of available options for attackers that are usually neglected.

Very powerful paths exist outside the sphere that is usually considered in threat modeling. A lot of data that is barely "anonymized" or "pseudonymized" is sold freely and legally on the internet, even though the relative simplicity of re-identification attacks has been proven repeatedly [82]. Our study shows the potential for abusing financially available data, with 21.7% of attacks using financial means to get to the data. The same data and more is available in its original state if government bodies are interested (19.2% of attacks). Finally, the often neglected physical access to victim and device shows within a non-negligible 16.7% of attacks. It is striking how frequent and easy all of these access types are, and we strongly suggest all of them be taken into account in developing future protective technologies.

Finally, the very frequent abuse of existing accesses (33.3%) presents a challenge for threat modeling as it implies that trust in any party should be limited, requiring potentially more complex solutions. There are several potential reasons why this is the most frequent type of data access by attackers. One explanation may be a lack of awareness by the user that the data is shared at all, which is consistent with the frequency of unknowing sharing that we observed in our research. Alternatively, the user may have misplaced trust in a party that should not have been trusted in the first place or that becomes corrupted over time while the user is engaged with that party. In either case, these findings highlight the need for caution and possibly enhanced privacy protections when sharing data, particularly with third-party entities. Acknowledging and integrating this problem in threat modeling is crucial to protect the user since users can not be expected to make informed decisions about whom to trust, as discussed in Section 3.3.

Our research further revealed an interesting finding that no physical accesses occurred due to shared devices, despite the fact that this is a known concern, as highlighted in a previous study on older adults [52]. While there are several possible explanations for this finding, including the size of our dataset or possibly a decline in the prevalence of shared devices, it is also possible that this result may reflect a news bias towards more sensational or high-profile privacy violations, as discussed in Section 5.3.4. Nonetheless,

First Sharing			
Upper Category	**Occurrences**	**Category**	**Occurrences**
Voluntary	7.6% (9)		
Necessary	33.9% (40)	No Opt-Out Possibility	0.8% (1)
		Required for Service	17.8% (21)
		Forced to Use Service	3.4% (4)
Involuntary	11% (13)		
Unknowing	46.6% (55)	Taken without Force	20.3% (24)
		Unknowing Opt-In	21.2% (25)
Unknown	0.8% (1)		

Table 5.5: Occurrences of categories in action "First Sharing": Upper category denotes category of highest hierarchical level according to the codebook, category denotes second hierarchical level, entries denote absolute numbers and percentage relative to the other categories' occurrences

Privacy Violation			
Upper Category	**Occurrences**	**Category**	**Occurrences**
Accuracy of Data	2.5% (3)	Exaggerated Retention	1.7% (2)
Collection of Data	6.6% (8)	Without Consent	6.6% (8)
Disclosure of Data	71.1% (86)	To Third Parties	47.9% (58)
		Publish Data	9.9% (12)
		Outside Intended Scope	8.3% (10)
Use of Data	19.8% (24)	Without Consent	3.3% (4)
		Forcefully Acquired Data	6.6% (8)
		Outside Intended Scope	9.9% (12)

Table 5.6: Occurrences of categories in action "Privacy Violation": Upper category denotes the category of the highest hierarchical level according to the codebook, category denotes the second hierarchical level, entries denote absolute numbers and percentage relative to the other categories' occurrences

it is important to recognize that the potential for abuse of shared devices remains a significant privacy risk that should not be ignored, as detailed in [52], particularly as some especially vulnerable groups in the population, like the elderly or poor, may rely on shared devices in their daily lives.

Harm

Our study has a rigorous focus on privacy and we aim to eliminate subjective judgment in determining what constitutes damage wherever possible. By following the narrative of each article, we classified consequences that

Attacker Data Access	
Category	**Occurrences**
Existing	33.3% (40)
Financial	21.7% (26)
Public	6.7% (8)
Legal	19.2% (23)
Physical	16.7% (20)
Unknown	2.5% (3)

Table 5.7: Occurrences of Categories in Action "Attacker Data Access": Entries denote absolute numbers and percentage relative to the other categories' occurrences

were implied as negative as harm, thus reducing subjectivity. This method is possible due to the structure of news articles, where mentioning of some kind of damage or adverse consequence is necessary for an interesting story. While this method is also biased, it is a static bias that is independent of the coder.

Apart from a high-level overview in Table 5.8, we provide a more comprehensive understanding of harms resulting from online events, as we expanded our list of harms by incorporating previous research from various data sources. In doing so, we also aimed to address the potential for certain harms to be overlooked due to their rarity or lack of widespread reporting. However, we excluded harms from prior work that clearly fell outside the scope of our research, such as the harms of intentionally viewing disturbing content online or engaging in the illegal sale of weapons or drugs. These types of harm were not related to privacy-based attacks, which was the focus of our investigation.

It is important to note that harm resulting from online events can be interpreted on different levels of abstraction. For example, domestic violence may not be explicitly listed in our tables, because articles *about* domestic violence may mention other negative consequences, such as threats of physical violence, distress, and harassment, which we would have flagged as harm. Domestic violence can be represented as a subset of attackers with a personal connection to the data subject.

We hope that the broad list of consequences provided in Table 5.9 will facilitate reflection on the scope within which protective technologies should operate and increase understanding of the pressing need for effective protective measures. The provided references point to papers in related research areas that mention and investigate the type of harm indicated in the respective row. These papers may also provide additional information about how these harms occur in the online space. We further include how often we

Category	Harms Occurrences	Potential Occurrences
Financial Harm	6.3% (10)	1.9% (1)
Social Harm	13.1% (21)	48.1% (26)
Physical Harm	3.1% (5)	18.5% (10)
Psychological Harm	32.5% (52)	9.3% (5)
Legal Prosecution	8.1% (13)	18.5% (10)
Mass Surveillance	26.9% (43)	1.9% (1)
Targeted Ads	10.0% (16)	1.9% (1)

Table 5.8: Occurrences and Potential Occurrences of Harms per Category: Potential Occurrences mean the harm is mentioned by the news article as potential or hypothetical outcome additional to the harm that has actually happened, Entries denote absolute numbers and percentage relative to the other categories' occurrences in the same column.

have encountered this harm in our privacy incident analysis. This additionally fosters an understanding of the bias in our dataset, as some harms might frequently come up in papers, but not in the reports.

5.2.4 Data Types

Several noteworthy observations can be made from Table 5.10 which highlight a plethora of challenges for effective protective measures. Firstly, location data is the most frequently abused data type (25.9%, 37 occurrences) and is usually obtained through purchase (18 occurrences). This highlights the careless and dangerous practices surrounding this type of data and the need for effective protective measures. Although efforts to address the issue of re-identification associated with location data [35] are on the rise, the data is still vulnerable to misuse. Secondly, genetic data has not (yet) been abused or sold in any of the observed attacks despite frequent media concerns [104] regarding the abuse of genetic data through commercial services. All observed cases of access to genetic data have been through physical (4) or legal means (5). Thirdly, biometric data has been inferred through publicly available visual content (2), emphasizing the need for evolving face obfuscation technologies [26, 25, 45]. Fourthly, in four cases, all data on a device was available to an attacker through physical or legal access. This type of data access has been widely neglected in the design of PETs until recently [12]. Lastly, communication data has been accessed through legal means in four instances, highlighting both the government's power to access intimate and private information and the need for a discussion about when and how this should be prevented, which has been discussed in prior work [12].

Category	Harm	Paper	Priv. Inc. Occ.
Physical	Break-In, Burglary, Robbery, Property Damage	[60, 52]	1
	Distance Restriction, Arrest		2
	Murder		2
	Detainment	[125]	0
	Incarceration	[125]	0
	Deportation	[125]	0
	Sexual Violence	[125]	0
	Destruction of Device	[50]	0
Psychological	Intrusion in Private Life		16
	(Sexual) Harrassment, Bullying	[60, 110, 50, 118]	4
	Stalking	[50]	10
	Mental Illness	[52]	6
	Emotional Distress	[125, 52, 84, 87, 50]	14
	Threats	[125, 110, 118]	2
	Spam, Trolling, Dogpiling	[60, 110, 118, 52]	0
	Blackmail	[50]	0
Financial	Job (Opportunity) Loss	[60]	5
	Discrimination	[87, 52, 60]	2
	Data Loss	[50]	2
	Quit Education		1
	Financial Loss	[52, 60]	0
	Negative Reviews	[118]	0
	Fraud	[52]	0
Social	Doxxing, Outing, Deadnaming	[118, 50]	3
	Exposure, Public Humiliation	[60, 118]	8
	Fear of Reputational Damage		3
	Defamation	[125, 52, 50]	7
	Damaged Trust in Government	[125]	0
	Hate Speech	[118, 110]	0
	Incitement	[118]	0
Other	Legal/Criminal Prosecution	[87, 60, 125]	13
	Surveillance	[125, 118, 52, 50]	43
	Targeted Ads	[87, 60, 52]	16

Table 5.9: Types of Harm by Category: Column "Paper" lists citations that mention or research the same type of harm, Priv. Inc. Occ. (Privacy Incident Occurrences) denotes the absolute number of occurrences in our analysis.

Data Types		Attacker Access Types					
Category	**Occurrence**	**Existing**	**Financial**	**Public**	**Legal**	**Physical**	**Unknown**
Visual Content	15.4% (22)	9	0	2	0	11	0
Technical Info	2.1% (3)	1	1	0	1	0	0
Biometric Data	4.2% (6)	4	0	2	0	0	0
All Data on Device	2.8% (4)	0	0	0	3	2	0
DNA	4.2% (6)	0	0	0	4	5	0
Communication	7% (10)	4	0	0	4	0	1
Public Records	1.4% (2)	0	1	0	1	0	0
Usage	11.2% (16)	8	3	1	2	0	0
Location	25.9% (37)	11	18	2	5	2	0
Medical Information	6.3% (9)	3	2	1	3	0	0
Other Personal Data	18.2% (26)	9	5	2	5	1	2
Unknown	1.4% (2)	1	1	0	0	0	0

Table 5.10: Occurrences of Data Types in Comparison to How They Are Accessed: In some attacks, multiple data types are involved, and in others, multiple access types, thus the sum of accesses and sum of data types is not the equal

5.2.5 Attack Context

To effectively prevent or deflect attacks, it is crucial to understand an attacker's resources, motivation, and whether the victim is a chosen target or one of many. Because most news articles lacked details about the attacker, we could only make rough estimates on their skills and resources. We observe that most attacks are untargeted (49.6%) and the data reaches the attacker either via existing access or through financial transaction, as shown in Table 5.11. Those untargeted attacks are usually performed by resourceful attackers, as seen in Table 5.12. Conversely, targeted attacks are mostly performed by low-skilled attackers that use physical access or publicly available data. Recognizing such prevalent attack patterns can aid in deciding whether the user should be protected as an individual or in a group setting.

An attacker's intention may influence their modus operandi and persistence in the face of obstacles. While we find most attacks to be intentional, there are also interesting categories of attacks where there is no inherently malicious party involved, as can be confirmed in Table 5.13. Some occur through negligence or incompetence (6%), while other attacks were found and reported through responsible disclosure, or identified before they could be abused (2.6%). This is especially relevant when considering which parties to trust, as through incompetence or negligence, even (in the colloquial sense) trustworthy parties can cause damage. Trust, in the case of threat modeling, is not only about malicious intent but also about the ability to function as intended, which is absent in the case of incompetence or negligence.

Target		Attacker Access Types					
Targeting Type	**Occurrence**	**Existing**	**Financial**	**Public**	**Legal**	**Physical**	**Unknown**
Targeted	32.7% (37)	7	7	4	10	10	1
Filtered	16.8% (19)	3	2	1	7	6	1
Untargeted	49.6% (56)	28	17	3	7	4	1
Unknown	0.9% (1)	1	0	0	0	0	0

Table 5.11: Occurrences of Targeted Attacks (in absolute numbers and percentages relative to other categories' occurrences) in Comparison to How Data is Accessed (in absolute numbers)

Attacker Skill		Targeting			
Attacker Skill	**Occurrence**	**Targeted**	**Filtered**	**Untargeted**	**Unknown**
High	51.3% (59)	11	6	42	0
Medium	23.5% (27)	11	8	8	0
Low	21.7% (25)	16	5	4	0
Unknown	3.5 % (4)	0	0	3	1

Table 5.12: Occurrences of Attacker Skills (in absolute numbers and percentages relative to other categories' occurrences) in Comparison to How Targeted an Attack is (in absolute numbers)

		Attacker Motivation	
Upper Category	**Occurrence**	**Category**	**Occurrence**
Intentional	91.5% (107)	Financial Gain	35.9% (42)
		False Positive	2.6% (3)
Collateral	6% (7)		
None	2.6% (3)		

Table 5.13: Occurrences of Attacker Motivation Categories: Upper category denotes category of highest hierarchical level according to the codebook, category denotes second hierarchical level, entries denote absolute numbers and percentage relative to the other categories' occurrences

5.2.6 Data Flows

Finally, we are equipped to inspect the data flows derived from our analysis, which is a core piece of this work. A single data flow is an individual attack we reviewed, described in terms of our model for privacy-based attacks. By visualizing all the data flows of the analyzed attacks at once, we are able to gather novel insights about not only the modus operandi of attacks but how we could alter data flows in order to prevent these attacks. Thus, our work creates a profound picture of possible attacks. Thanks to its extensiveness, together with our model of privacy-based attacks, it provides a novel framework for researchers to review their threat models and observe whether they have captured all relevant vectors of attack.

We will first review the data flows between the parties of the privacy-based attack model in Figure 5.1. The data confirm some of our previous observations, such as the amount of data going to companies and being shared with

other companies or the government. Further, many conclusions we drew and trends we observed in the sections before are now more easily accessible, such as the distribution of attacker data access types and initial receiver categories.

The visualization makes trends easy to observe and provides a new perspective on cases where one entity assumes multiple roles. For example, the most common data flow pattern we observe is where an individual shares their data with a company, which then shares it with another company that also acts as the attacker. Or, at a glance, we see a significant portion of data flows that originate with the government as the initial receiver end up with the government as both the data handler and attacker.

This visualization further allows us to reflect on where a designer would best employ protective mechanisms. One could use data minimization to avoid companies getting such a vast amount of data and prevent a large share of the observed attacks. Alternatively, measures can be taken to prevent data retention, preventing data from being repeatedly shared with other parties. Instead of listing myriads of options here, we aim to foster an understanding of how to interpret and use our results in a threat modeling process.

It is important to note that the severity of harm caused by an attack does not necessarily correlate with the breadth of the data flow. Therefore, we encourage users of this framework to not only consider the broadest data flows but also reflect on the smaller ones during the threat modeling process.

To provide a comprehensive understanding of how to modify data flows, we visualize the data flows in terms of actions, as presented in Figure 5.2. By examining the various actions associated with each data flow, we can identify appropriate measures to prevent these actions and, thus, potential attacks.

From the fraying of flows between attacker data access and harm, we can see that how the data was obtained does not necessarily influence how it can be used against the victim. On the other hand, we can observe how various ways of data disclosure make up a significant portion of flows, which underlines the power and potential of technologies such as Zero-Knowledge Proofs (ZKPs), Multi-Party Computation (MPC), Fully Homomorphic Encryption (FHE), and Federated Learning (FL) which allow parties to perform computations without sharing raw data. However, we want to point out that each depicted path in the diagram is not to be neglected, as its mere presence proves the existence of an attack possibility that could be avoided in the future if the protection of users' privacy is taken seriously.

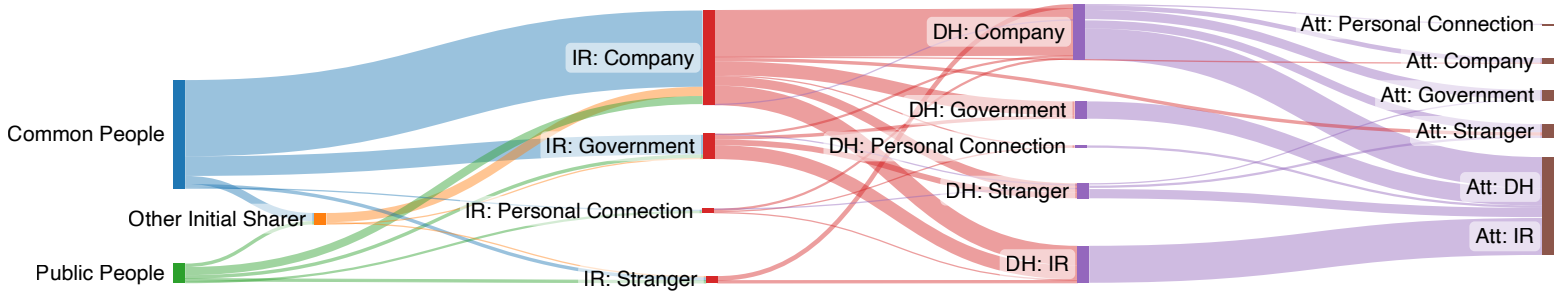

Figure 5.1: Data Flow between Parties in a Privacy-based Attack: The data flows from the data subject on the left to the attacker on the right. "IR" stands for initial receiver, "DH" for data handler, and "Att" for attacker. The width of a bar indicates its relative share, the broader the bar the bigger its share in our data set.

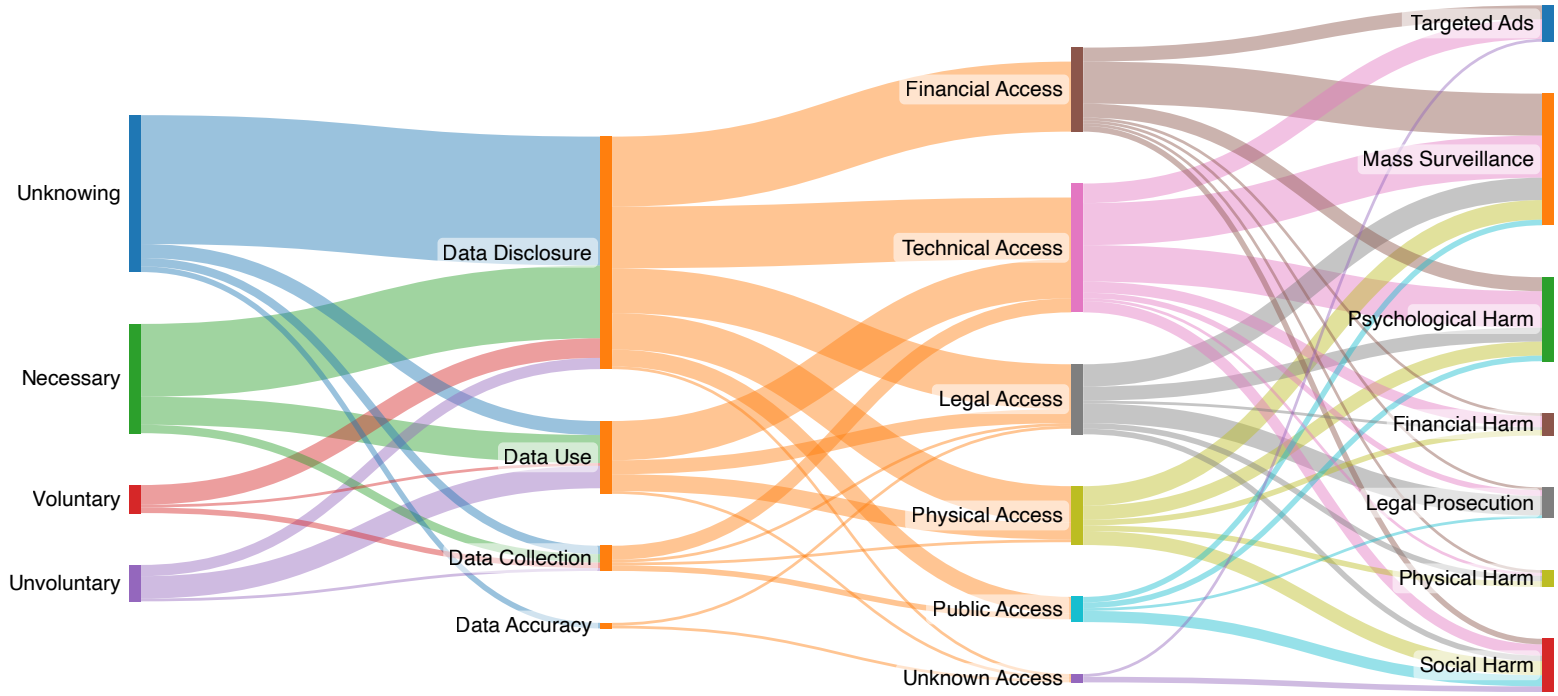

Figure 5.2: Data Flow Visualized through Actions in a Privacy-based Attack: The data flows from left to right. The width of a bar indicates its relative share, the broader the bar, the bigger its share in our data set. Action between Data Subject and Initial Sharer is omitted.

5.3 Discussion

Our analysis of past privacy incidents has provided us with valuable insights into the modus operandi of privacy-based attacks. Through our examination of the involved parties, actions, and contextual factors, we have been able to visualize the associated data flows in a meaningful way. We are confident that these findings can support researchers in adopting a more holistic and user-centric approach to threat modeling.

We contend that our approach, which focuses on the entire attack data flow, including its outcomes, is more effective than simply considering privacy violations in isolation. Attacks are now easier to observe, and our results clearly indicate where urgent attention is needed. We also facilitate identifying potential solutions because we can observe the data flows and reason about the efficacy of different measures. In the following sections, we will discuss additional observations that we made during our analysis and highlight problems that we were not able to address previously. While news articles shape these observations and could thus be limited in their truth, as discussed in Section 5.3.4, they are supported by recent court rulings, political debates, and legislative efforts that emphasize the issues we discuss.

5.3.1 Powerful Players

We have previously mentioned how all categories of entities are present in all parties, which confirms that they should be considered in threat modeling. Each of these categories implies a certain set of assumptions about resources, which hold for most cases and make sense to bear in mind when interpreting these results.

This includes, for example, a government having vast options in terms of money and skills, thus being able to conduct large-scale, high-effort operations if needed. A personal connection, however, might be able to build up social pressure through their relationship with the data subject or have background information, but as an attacker is likely to remain a "UI-bound adversary" [50], thus can only make use of features and information provided by the User Interface (UI) of a device or application.

In the following, we want to focus especially on companies and governments and will discuss the small portion of personal connections and strangers mostly in Section 5.3.4

Before we further elaborate on the power yielded by big companies and governments, we want to note that both are, in many cases, bound by certain well-known incentives. For governments, it can be detrimental if their citizens collectively lose trust in them. Especially in democratic countries, the ruling party will want to keep their citizens somewhat content to stay

in power. The same is true for companies that rely on having customers to make money. These thoughts should be kept in mind when reflecting on the next few paragraphs.

Nation States. We want to emphasize that solutions should be found to keep nation-state adversaries at bay. Our analysis shows nation states do make use of their power, which includes wielding warrants at companies who retain data [16], or simply buying the same data on the free market and thus even bypassing warrants [31, 116]. Government access to both past and present data about citizens is especially dangerous when long-standing fundamental rights break away and government bodies become too eager to find newly appointed "criminals" (e.g., women getting abortions [134, 98]). These possibilities to access data create unprecedented options for mass surveillance not only within the nation-states own but also on foreign soil. Further, we observe increased efforts in compelled self-incrimination of arrested people, increasing the urgency of solutions that protect personal data when unlocked devices are compromised. Physical access to unlocked devices has flown under the radar far too much. We argue that we should no longer treat personal devices as trusted, and assume that an adversary can unlock them. This poses tedious problems to solve, but they are nonetheless a crucial and necessary step in user protection.

Companies. We have mentioned how selling barely deidentified data is legal in many countries, even though re-identification has been proven to be easy [35, 21]. We want to emphasize that re-identification is not only easy, it is also actually being done by attackers in practice [21]. Looking at Figure 5.2, at the portion between data disclosure and financial access, we see how many attacks could be avoided if the sale of data was prevented, and how dangerous it is at present.

However, the problem does not stop there, as this data is analyzed and aggregated by data brokers, who allow their clients to buy data of specific groups of people, such as "pregnant women". We further want to point out how certain apps specifically target marginalized communities and sell the data they obtain [21], which is obviously dangerous, as has been shown repeatedly [125].

The deliberate blindness of app providers even includes accepting money to integrate certain SDKs which directly gather data for another company or government. In numerous cases, the respective developers either do not review what those SDKs do exactly, or turn a blind eye. This puts their user base at risk [31].

We argue that such practices should be known to everyone making threat modeling choices for underlying technologies or protective measures.

5.3.2 Further Challenges

In the following, we would like to highlight certain data points that we observed require special care, and present complex challenges in providing protection for the user.

Sensitive Information. Given the data types that have been abused in attacks, we find that some data types are not being treated as sensitive enough. While medical data, for example, is often protected by law [10], many other types are not. The dangers stemming from visual content, which can nowadays be used to extract biometric data, and from location data, which is captured by a plethora of services such as weather apps, dating apps, social media, in the metadata of every picture, should raise concerns about what nowadays is considered sensitive. These two data types also make up a considerable portion of the data abused in attacks, with 15.4% and 25.9%, respectively. Researchers have recently recognized these dangers, and efforts to obscure users' paths [15, 14] and faces [26, 25, 45] are being made.

Intangible Damage. In line with our user-centric approach, we want to point out that harm can be intangible, hard to assess by only technical means and only known to the victim if not shared on purpose, as is mostly the case with psychological harm.

In the review of our results, we see that psychological harm is not only one of the most frequent consequences (32.5% or 52 occurrences), but also the most commonly reported in combination with other harms, such as financial (2 out of 10), social (6 out of 21), and physical (1 out of 5) harm, as well as mass surveillance (3 out of 43) and targeted ads (2 out of 10). This makes sense given the nature of psychological harm: losing money or being socially excluded can cause distress, and, therefore, psychological harm can result as a further consequence of other types of harm. Therefore, it is crucial for both technology designers and legislators to take into account the potential psychological impact of data abuse and work towards solutions that minimize that harm. This can be as simple as legally acknowledging the damage but can also include emergency protocols or more complex solutions to give control back to the user.

Another harm that seems even harder to assess is mass surveillance, as its consequences to the general population are often hard to observe, even though they are well-researched. Surveillance is a versatile tool of power and of interest to many malicious entities. It is further not surprising that surveillance is one of the most common consequences of lack of privacy (26.9%), as it is relatively easy to achieve with today's vast amounts of data. Even if no individuals are targeted due to said surveillance, simply being aware of being watched has consequences, such as change of people's behav-

ior, self-censorship, or even anxiety [100]. Furthermore, mass surveillance often results in the warrantless collection of personal and intimate data, creating an environment where individuals are presumed guilty until proven innocent.

Such practices violate individual privacy rights and undermine trust in institutions. We urge for the development of technologies that limit the current possibilities of large-scale data collection, and work towards limiting the use of surveillance to situations where it is absolutely necessary and legally justified.

5.3.3 Incentives and Adoption

While we have covered many technical implications of our results for threat modeling, there are additional points to consider beyond the technological aspects.

When deciding where to deploy a technology to protect the user, it is important to consider not only the data flow but also the incentives that would encourage adoption. Technologies can also be developed in a way to offer those incentives, as is seen often in MPC. It is crucial to take into account commercial, political, social, and personal factors to ensure that the technology is not only theoretically effective but also practically adopted. This is not solely the responsibility of technology designers but also requires legislative action, which has the greatest potential to counteract financial interests.

The categories outlined in Section 5.2.2 and the visualizations in Figures 5.1 and 5.2 provide a comprehensive basis for discussing potential interventions. Therefore, it is necessary to consider all factors when creating solutions that are not only effective but also feasible in practice.

5.3.4 Limitations

The approach of relying on news articles to understand the harms of data sharing has several limitations. Firstly, news articles are not a perfect reflection of reality and can be biased towards extremes or certain communities. As a result, marginalized communities or harms that are not socially recognized as damage may be underrepresented in news reports. Furthermore, incidents that are not reported on are not be captured by this approach. To mitigate these limitations, it is important to seek outside perspectives and information from other sources, such as research on marginalized communities.

We did this explicitly with the harms in Section 5.2.3 with a limited number of related works. Some types of harm may not receive as much attention in news reports, but they can still have significant impacts on individuals and communities. For instance, studies regularly find that digital

abuse is anything but rare, especially in relationships, where in one case, it was found that 12% of participants had experienced intimate partner digital abuse [133], while in our analysis only 3.5% of attackers where a personal connection. Thus, we urge again to interpret our results more in terms of distribution and presence versus absence than to confide too much in the listed frequencies.

Finally, we would like to point out that the qualitative coding has been conducted by only one coder, where best practice would suggest using at least two for consensus and avoiding mistakes. While results have been carefully reviewed and discussed with all supervisors, given the nature of a master thesis and the background of the author, it was not possible to employ and train a second coder within the given time frame. Nevertheless, we encourage future work that aims to extend this framework to use two coders, at least, to adhere to best practices.

Furthermore, it is important to acknowledge that our study was primarily focused on U.S.-based newspaper agencies, which may have introduced some bias towards WEIRD communities (Western, Educated, Industrialized, Rich, and Democratic). To address this limitation, future research could expand the geographic and cultural focus of the framework to provide a more comprehensive understanding of privacy violations across different contexts and communities. Such an approach would enable a more nuanced understanding of the challenges and opportunities associated with privacy violations, and ultimately support the development of more effective strategies to address these issues.

Chapter 6

Analysis of Privacy Threat Modeling

Having gained a solid understanding of real-life privacy attacks, we now delve into the exploration of PETs as a means of preventing such attacks. Our investigation focuses on the impact of current PETs on data flows, from which we estimate their potential to safeguard users from harm and gather an understanding of their practical implications from a user's perspective.

We have previously found that certain data flows can enable an attack. Nissenbaum calls such data flows "inappropriate" [92]. Therefore, to successfully avert an attack, the data flow has to be changed, in our case by a PET. We now inquire how this can be achieved, and whether what the research community proposes can sufficiently protect the user.

However, comparing different PETs can be tedious due to the lack of consistency in threat modeling and guarantees (or even the lack thereof), as well as narrow scenario scopes. Fortunately, we can solve this problem by analyzing PETs using the same privacy-based attack model (Figure 4.2) that has endured over 100 past incidents.

Our previously developed framework provides us with the ability to create a meaningful comparison between individual PETs and to draw insightful conclusions from the current state of research. Unlike prior work [96], our approach avoids the need to categorize PETs for comparability, and allows us to take a user-centric perspective in estimating their impact. This sets our approach apart and provides us with a valuable way to compare and evaluate PETs.

Through this approach, we gain insights into the current state of research and its impact on data flows, as well as the user's privacy in terms of control over their own data. We identify common privacy violations and attacks that are being addressed, as well as those that remain unresolved. Our analysis also reveals overlaps and mismatches between real-life attacks and PETs, from which we derive suggestions and guidance for future research.

In the following sections, we will introduce the methodology used to achieve the comparison of PETs in terms of their impact on data flows data flows in Section 6.1, share the results of our work in Section 6.2, and discuss them in Section 6.3.

6.1 Methodology

In this section, we will provide a detailed description of our methodology. Firstly, we will clarify the scope of our study and explain our rationale for selecting a particular data source, namely published papers proposing new PETs. Next, we will elaborate on our data collection process by defining inclusion and exclusion criteria that reflect our scope and research question, as well as determining the appropriate size of the dataset. Our goal is to capture a wide range of relevant papers that accurately depict the breadth of the research space. Lastly, we will provide a comprehensive overview of our qualitative analysis approach.

We will use only inductive coding, which allows us to identify the properties of privacy-based attack models that are consistently present across all papers. It is important to note that the papers we have selected may not necessarily consider full privacy-based attacks but rather privacy violations, which may not encompass the entire data flow, and thus provide specifications for only a part of the parties and actions present in the privacy-based attack model. With inductive coding, we can efficiently integrate new properties as needed and adapt to circumstances where selected papers may lack an attacker or threat model, as well as a complete motivational scenario, resulting in a potential lack of context.

Our focus will be on identifying and systemizing the characteristics of PETs in terms of their impact on data flows and the user's privacy as comprehensively as possible. We contend that our methodology offers a valuable cross-sectional view of the PETs research field, as we have carefully chosen relevant papers based on the criteria described in Section 6.1.2. Additionally, we demonstrate in our use of data flows defined by the privacy-based attack model that it is a robust and effective means of facilitating comparisons.

6.1.1 Scope and Data Source

To ensure that our data sources align with our research objectives, we have chosen published research papers that propose PETs as subject of our study. This approach allows us to examine the original intentions behind the design of PETs and provides a representation of the work of our target audience. Moreover, we believe that research papers are more detailed, less biased, and more independent compared to whitepapers from corporations or other sources.

To select relevant papers from the literature on PETs, we have narrowed down our search to include only those papers that have won or been runners-up for either the Caspar Bowden Award for Outstanding Research in Privacy Enhancing Technologies [54] or the Usenix Test of Time Award [7]. The Caspar Bowden Award is named after Caspar Bowden, a privacy advocate who created the foundation to encourage researchers to develop technologies that protect people's privacy rights. The Usenix Test of Time Award recognizes papers that have made a lasting impact on their respective fields, and to qualify, they must be at least 10 years old and originally published at FAST, Usenix Security or NSDI.

This scope gives us a cross-section across the field of privacy research at a reasonable quantity, and we avoid overrepresenting short-lived "hot topics" that could have been prevalent had we only looked at the last few years. By including runners-up alongside the winners, we have increased the size of our dataset while still ensuring its relevance. Additionally, we believe that being a runner-up to one of these prestigious awards is a sufficient indication of a paper's significance in the field of privacy research.

6.1.2 Data Collection

We have established specific criteria to select papers from the list of Caspar Bowden and Test of Time Award winners and runners-up. First, the paper must propose a PET, or offer a thorough analysis of one. This ensures that the paper covers a design process, concrete privacy problem, or original ideas that are of interest to us and contains sufficient detail for qualitative analysis. We exclude papers that are merely an evolution of another paper or PET, as they often omit essential details by referring to the original paper, making them tedious to analyze qualitatively. Further, the paper needs to mention a real-world use case and mention privacy. This is because we are not interested in building blocks but in technologies that aim for real-world deployment since such technologies and the researchers behind them are the core focus of this thesis.

We do not expect the papers to explicitly state threat models or have running implementations since we want to avoid narrowing our data set further. Additionally, older papers may not have had the same technological capabilities for implementation as we do today. Adhering to these criteria, we were able to identify 17 papers published between 2002 and 2021. A list of these papers can be found in Table A.1 in the Appendix.

6.1.3 Qualitative Analysis

In Sections 2.5 and 5.1.4, we have introduced the foundations of qualitative analysis and coding. We will employ this analytical method again because

we are working with a natural language data source. By utilizing qualitative analysis and reusing the privacy-based attack model, we can achieve comparability even with a small data set and effectively analyze and interpret our data.

Having comparability is crucial to compare the PETs with each other and to our findings in Chapter 5. Our results will help us understand the current direction of PETs research in relation to real-life attacks, and contextualize our insights to understand how PETs can help protect against real-life privacy threats.

Focus. For this smaller analysis, our goal is not to systemize over a hundred complete data flows as in Chapter 5. Instead, we aim to understand how PETs modify data flows and the involved parties. Specifically, we seek to identify the interacting parties of the PETs, whose data is being protected, and how. Control is a key consideration, so we investigate who controls the data subject's data without the PETs and who would control it if the PETs was deployed. Additionally, we examine how the shared data changes when the PETs is applied. These points provide insight into how the situation changes thanks to the PETs in terms of data flow. Finally, if the paper presents this information, we will extract the capabilities that the authors assign to a potential attacker. We will focus on how the paper presents these points in terms of its intended use case. This analysis is not meant to gather the complete set of PETs' capabilities but aims to provide a broad overview of how the research community is motivating their work and the problems they claim to solve.

Inductive Coding. In Section 5.1.4, we introduced the concept of inductive coding, which we will be using exclusively this time, without additional deductive codes. The rationale behind this decision is that the papers we are analyzing differ significantly in style, topic, and age, and we cannot predict in advance what kinds of features we will be able to identify. These papers may not present complete attack paths, but rather focus on specific aspects of an attack. Hence, by utilizing inductive coding, we can incorporate all relevant features that fall within the scope of our research.

Despite using inductive coding, we still follow the privacy-based attack model to provide context for the attack snippets we find, and ensure that our results are comparable both with other papers and real-life attacks. This approach enables us to stick to a concise set of categories, creating new ones only when necessary and appropriate. This approach facilitates trend identification in a smaller dataset.

6.2 Results

We aim to provide a comprehensive understanding of how and where privacy-enhancing technologies (PETs) modify data flows and whether the progress made by the research community is effective in mitigating real-life attacks. To achieve this, we analyze and contextualize our observations on different PETs, comparing them with each other and with the results of our analysis in Chapter 5. This enables us to identify patterns and characteristics in the design of PETs, which we can discuss in the context of real-life attacks.

To facilitate these discussions, we propose a new classification option for PETs based on their impact on data flow. We found that the alterations to data flows and control over data subjects' data are non-trivial, and systemizing them makes discussions easier.

To review our results, we first present the derived codebook in Section 6.2.1. Subsequently, in Section 6.2.2, we provide noteworthy observations on the occurrence of codes, primarily focusing on discussing occurrence versus absence and interpreting distinctive distributions. This helps us identify patterns despite having a relatively small dataset. Finally, we analyze the modifications to the data flow in Section 6.2.4 using the insights gained from the previous subsections.

Our aim is to provide more context to the findings in Chapter 5 and facilitate a better understanding of areas where future research could be more effective. Additionally, we aim to highlight overlaps and differences in focus between research and real-life attacks.

6.2.1 Codebook

To comprehensively understand how PETs influence data flows and the user's privacy concerning their control over their own data, we developed a codebook through an inductive analysis of 17 highly relevant research papers on PETs. This codebook includes 47 distinct codes, providing a foundation for categorizing changes to data flows and identifying the entities in control. Our objective is to use the provided codes to determine the impact of PETs on data flows and the control over user data.

Although we do not assert that this codebook is exhaustive, or the codes mutually exclusive, it provides a reference for future research to delve more profoundly into the issues we discuss in this section. Further, the provided codes are a starting point for future work to further systemize data flows and create a new categorization for PETs.

Area	Code	Description
Privacy Violator Capabilities	Background Information	Additional information not contained in the data flow
	Nation State	Power and resources of a nation-state or its branches
	Insider Access	No need to create access, e.g. app data access by distributor, employee access to business data
	Data Analysis	Information inference through (statistical) analysis of data
	Other	
Averted Harm	Harm	Definition 4.2
	- Censorship	Taking down public messages, blocking access to webpages or the internet
	- Surveillance	Digital or physical, continuous or selective forms of surveillance
	- Prosecution	Legal investigation or conviction
	- Reputational Damage	Defamation
	Privacy Violation	Data subject has no control over their data
	- Semantic Inference	Creating new information from data
	- Data Leak	Accidental data disclosure
	- De-Anonymization	Identity and data can be linked
Targeting	Targeted	Victim is selected as aim of attack
	Untargeted	No or very little aim is employed for attack (e.g. mass surveillance)
	Both	Targeted and untargeted attacks are considered in this context
Target Audience	Data Processors	Entities who obtain data to retrieve information from it, e.g., journalists, researchers
	Data Collectors	Entities who retain data, e.g., databases, storage providers
	Service Providers	Entities who provide a service and/or interact with the end-user, can e.g., deploy client-side apps
	User	End-user of any service, the data subject

Continued on next page

continued from previous page

Area	Code	Description
Data Control before PET	Initial Receiver	Initial receiver has control over data subject's data
	Data Handler	Data handler has control over data subject's data
	Attacker	Attacker has control over data subject's data
	Public	Data subject's data is completely public
Data Control with PET	Initial Receiver	(New) initial receiver has control over data subject's data
	Data Subject	Data subject is in control over their own data
	Same as Before	The same entity as before the PET was deployed is in control over the data subject's data
Alteration	Revealed	Data flow is made visible by the PET to the data subject
	Intercepted	Data that has been shared before PET is either destroyed or never created thanks to the PET
	Narrowed	Data is still being shared but in a way that reveals less information about the data subject than without PET
	Redirected	Deployer circumvents or deceives original parties with the new data flow and deploys the PET without their agreement or interaction
	Added	Data flow that was previously impossible is enabled by the PET
Data Type	Identifiable Information	Data inherent to a person, e.g. name, age, address
	Behavioral Data	Data generated by a person, e.g. browser history, sensor data
Guarantees	Data Minimization	The PET allows the same/similar functionality with less data being shared
	Confidentiality	The PET allows data that was previously disclosed to be kept secret
	Deniability	Allows to deny knowledge of, interaction with or responsibility for data or data flow
	Identity Protection	The data subject's identity is kept secret or only selectively disclosed, anonymity

Continued on next page

continued from previous page

Area	Code	Description
	Detection	The PET detects data or data flows to inform the data subject about them
Protected Party	Owners and Sharers	Both the data subject and initial sharer(s) are protected by the PET
	Data Subject	Only the data subject is protected by the PET
Deploying Party	Data Subject	Data subject installs the PET on its own devices
	Initial Receiver(s) and Sharers	Initial receiver(s) and/or anybody sharing data installs or deploys the PET
	Both	Data subject and initial receiver deploy or install the PET

Table 6.1: PETs Analysis Codebook: The codes' hierarchical structure is indicated by indentation. Depending on the complexity of the mentioned use case, multiple instances of the same code category may have been present in a paper, and occurrences of codes in a specific category may not sum up to the amount of papers analyzed.

6.2.2 Parties

In the following paragraphs, we will highlight the most noteworthy properties we discovered in the parties involved in PETs data flows. We will analyze these parties' roles in the privacy-based attack model to provide readers with a better understanding of how the approach and narrative have changed between news articles and research papers. This highlights the usefulness of the privacy-based attack model and the results presented in Chapter 5 in this context.

Attacker Our analysis indicates that the "attackers" discussed in the literature primarily focus on violators of privacy rather than attackers as defined by our model, as they often do not consider full attacks in terms of the privacy-based attack model, but more consider themselves with partial attacks or solely privacy violations.

Our analysis revealed that some of the skills mentioned in research papers regarding an attacker or privacy violator's capabilities were also present in real-life attacks. These include the abuse of existing access and nation-state resources, which were identified in 4 and 2 occurrences, respectively in the papers, and our analysis of real-life attacks confirms their seriousness as threats. Moreover, data analysis skills and background information were identified as essential in 6 and 2 occurrences, respectively in the research papers, and these skills were also found to be crucial in real-life attacks. This indicates that some of the research papers accurately reflect attacker capabilities that are indeed relevant in real-life attacks.

However, our data did not reveal any PETs in our dataset that could effectively protect against personal connections' abuse, which we consider a significant gap, given the prevalence of digital abuse in relationships [133].

Initial Receiver. We discovered that the initial receiver of a PET plays a critical and versatile role in its deployment. In 10 papers, the initial receiver is responsible for deploying the PET, and have to be at least partially trusted. This has significant implications for the privacy of end-users, which we will explore in detail in Section 6.3.

Moreover, we observed that there can be multiple initial receivers, and they can also become targets themselves and even change roles during the deployment of PETs. In some scenarios, multiple initial receivers operate in a decentralized manner, such as in MPC settings, where parties have strong incentives not to disclose data to each other, ensuring the data subject's privacy. In other cases, trusted third parties are introduced, and they replace the initial receiver, becoming the new initial receiver, while the previous initial receiver becomes a data handler. This "shifting" of roles frequently happens, and we encourage future work to investigate whether it has any

impact on possible attack data flows. Additionally, initial receivers can become targets themselves, as seen in the case of investigative journalists who receive data from whistleblowers, giving them a strong incentive to deploy PETs correctly.

The privacy-based attack model has highlighted the crucial role of the initial receiver in deploying PETs and our knowledge from past privacy incidents highlights the need to consider their trustworthiness, context, and incentives for effective data protection. We will elaborate on this further in Section 6.2.4.

6.2.3 Actions and Context

In the paragraphs below, we will discuss our findings regarding the guarantees offered by PETs, the types of damage they claim to protect against and explain our categorization of data types and types of targeting in attacks.

We find that the consequences mentioned in the papers do not distinguish between privacy harm as per our Definition 4.2 and privacy violations. This is in line with our finding that papers do not mention full attack paths, but only snippets where a very specific problem is being solved, usually in the form of an interaction between two parties that is being transformed.

Since multiple technologies may be needed to fully avert certain attacks (an observation matching with [96]), we would encourage a more holistic point of view in the use case examples, as this kind of thinking could make PETs more interoperable, integrateable and thus more effective and more likely to be deployed.

Guarantees. To analyze the guarantees provided by PETs, we derived five categories: data minimization (2 occurrences), confidentiality (3 occurrences), deniability (3 occurrences), identity protection (12 occurrences), and tracking detection (1 occurrence). Although these categories are not mutually exclusive, they provide a comprehensive high-level understanding of the guarantees a PETs provides.

In line with the above guarantees, the most frequently mentioned harms and privacy violations that these PETs protect against are de-anonymization (9) and surveillance (6), followed by self-incrimination (2) and data leaks (3). Given that surveillance was also the most frequent harm in Chapter 5, this accurately reflects people's privacy concerns.

Furthermore, our findings show that the guarantees of identity protection and confidentiality offered by PETs are crucial for protecting against real-life attacks. Notably, the research community recognizes the inadequacy of pseudonymization for de-identification, and proposes PETs that offer

strong anonymity and unlinkability guarantees, which is crucial for preventing linking attacks by individuals and nation states, as revealed by our analysis of real-life attacks.

Moreover, unauthorized disclosure is a significant source of privacy violations, and PETs that provide strong confidentiality guarantees can help prevent such disclosures and improve user protection. In addition, we also found that deniability is an important property of PETs in combatting compelled self-incrimination, which is becoming increasingly relevant in light of recent geopolitical developments [12].

Finally, we highlight the importance of developing tools that allow users to detect data flows as a means of increasing their awareness and control over their personal data. This is particularly relevant given that users often prioritize convenience over privacy and are not fully aware of the data being collected by services they use, as demonstrated in our analysis of past privacy incidents. We argue that such tools not only enhance user protection but also serve an educational purpose by promoting greater awareness of data privacy issues.

The guarantees we identified address some of the most significant privacy concerns we uncovered earlier in this thesis. Nevertheless, it is crucial to recognize that a combination of these properties, implemented at various stages in the data flow, is necessary to provide effective user protection. Future work is needed to make stronger and more reliable statements about these issues.

Data and Attack Types. To address the limitations posed by a small sample size, we have classified protected data into two categories: identifiable information that is linked directly to an individual (8 occurrences), such as name, age, or address, and behavioral data such as browser history or sensor data (10 occurrences).

We have observed that both types of data are being safeguarded equally, which is a positive indication that the community recognizes the significance of protecting not only obvious sensitive data, but also preventing inference of user information, which is a prevalent and growing threat. It is important to note, however, that these two types of data can coexist in e.g. an application, and it thus may require multiple protection mechanisms to ensure user privacy and security.

Further, our analysis revealed that research papers typically focus on defending against either targeted or untargeted attacks. However, in Section 5.2, we mentioned filtered attacks, which can be seen as either a scaled-up targeted attack when executed repeatedly or a scaled-down untargeted attack when the attacker selects data based on certain victim characteristics.

Therefore, we argue that this is not a gap but a consequence of different narrative styles.

However, we found that more research papers mention targeted attacks (10 occurrences) than untargeted attacks (3 occurrences, 4 papers mention both), while our real-life observations showed the opposite trend. Not all data that is benign for an individual is also harmless for a larger group. Depending on the technique used, obfuscated data can prevent individual identification while still revealing trends, which could be exploited for mass surveillance and other nefarious purposes.

Therefore, we urge researchers to recognize the prevalence and potency of current untargeted attacks, as well as the minimal data required for these attacks, which could rely solely on distributions.

6.2.4 Alterations to the Data Flow

The insights presented in this section highlight the effectiveness of existing PETs but also expose several challenges that remain unaddressed. Most research papers focus only on specific properties of privacy, such as anonymity, rather than approaching privacy as a holistic concept, making it unclear if PETs provide adequate protection for a user's data flow.

To achieve an effective protection of user data, we argue it is crucial to adopt a more comprehensive approach that addresses the interoperability and integration challenges of PETs, formulates realistic attacker and threat models, and incorporates a user-centric perspective. The discussions presented in the following sections offer a fresh perspective on the impact of PETs, empowering researchers to make informed decisions on how to handle real-life attack scenarios and threats while considering the user's perspective.

In the upcoming paragraphs, we will examine the impact of PETs on the data flow, focusing on the points at which the data flow is altered, as shown in Figure 6.1, in order to identify areas where the research community's efforts align or diverge from real-life attack scenarios. We will begin by discussing the party that controls the data subject's data before and after the deployment of a PET, followed by an evaluation of how the PET alters the data flow, and finally, we will analyze the implications of these changes for user protection.

The Party in Control

Before we analyze our own results, we note that there is a limited number of points where privacy violations can be averted, which highlights the significance of Figure 6.1. These points include:

- The data never leaves the data subject, thus privacy remains untouched.

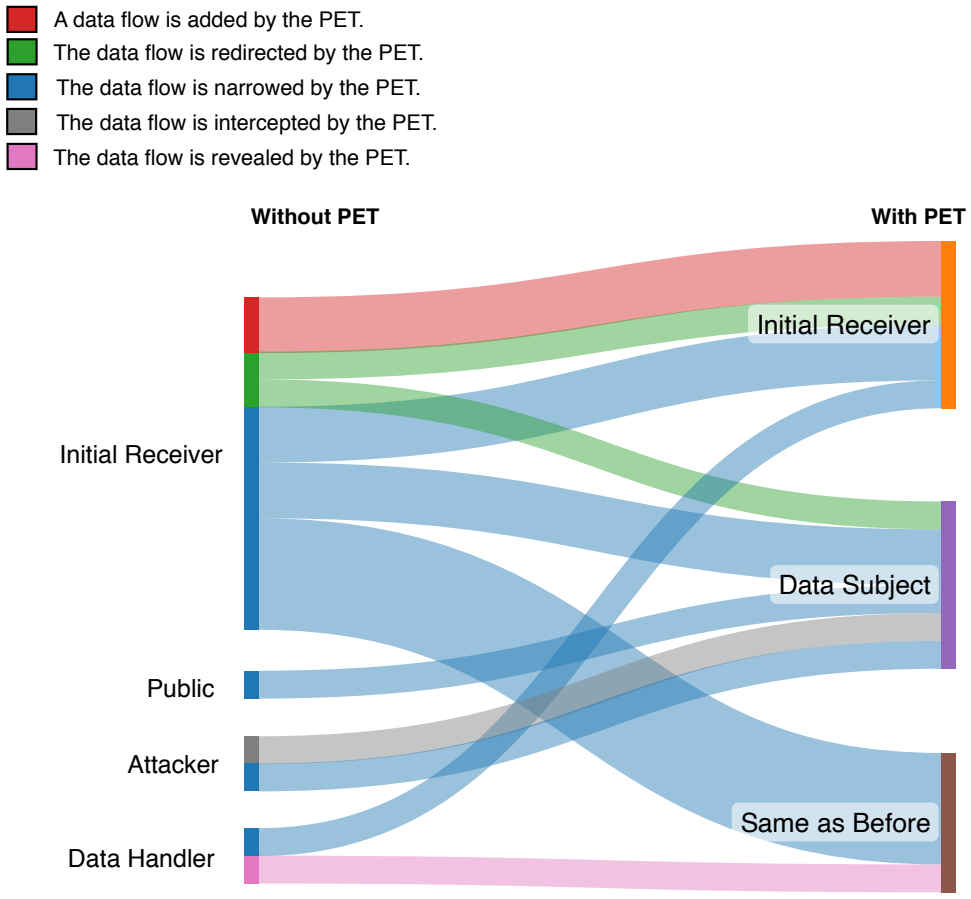

Figure 6.1: Alteration of Data Flows and Change of Parties in Control of Data Subject's Data in Analysis of PETs: Color of edge indicates the type of alteration, connection indicates which party in control of data without PET leaves control of data to which party when the PET is deployed. Width of an edge indicates relative portion of occurrences compared to total number of PETs analyzed.

- The data is shared in a secure manner between the data subject and the initial receiver. The initial receiver can be a trusted party, where only the channel between the two needs to be protected (as well as storage later on), or an untrusted party, where the data needs to be made non-sensitive.

- The privacy violation occurs when the data is transmitted from the initial receiver to the data handler. In such a case, privacy can be protected by not transmitting the data or by rendering it non-sensitive. However, this is a riskier option as the initial receiver can become corrupted, as demonstrated by real-life incidents such as governments acquiring confidential data from companies (as discussed in Chapter 5).

- If the data is controlled by the data handler or the attacker, it results in a definite violation of the data subject's privacy.

Figure 6.1 displays how the party in control over the data subject's data changes. It should be noted that except for the data subject, parties can be shifted or changed upon the deployment of a PET.

We observe that before the PET is deployed, the data subject is never in control over its data. This is intrinsic to our scope, since if the data subject was in control of the data before the PET there would be no privacy violation possible.

We find that usually only data subjects or initial receivers deploy PETs, and it is logical to deploy a PET while privacy is still fully intact. However, there is one exception, where the data flow is only made visible. This does not entirely keep the user's privacy intact, as by that point data has been shared, possibly without the user's consent. Nevertheless, the information allows the data subject to take action against further privacy violations. Until the data subject takes that action, it is however not in control over their own data. Thus, the PET itself does not bring back this control. We will discuss the issue of protecting privacy after it has been violated in Section 6.3.

Further, we notice that in most cases, it is not the data subject that ends up being in control over their data, so there is a gap between the deploying party and the protected party. We find PETs are deployed by the initial receiver who is usually thought to be some kind of service provider. This can be the same initial provider as in the original scenario, or a new one that has been introduced by the PET's requirements.

It may be concerning to consider how the privacy of the data subject can remain uncompromised if their data is controlled by the initial receiver. However, there are two crucial points to observe. Firstly, the data subject has the option to voluntarily share their data with the initial receiver and trust that it will not be misused. It is worth noting that even in our privacy-based attack model, the violation of privacy is only complete when the initial receiver abuses their privileges. Secondly, if the initial receiver is also responsible for the client-side application, we contend that they have control over the data subject's data, as they have the capability to un-deploy the PET and regain access to the data on their own. This possibly controversial idea will be discussed further in Section 6.3. We thus find that in many cases, a strong trust relationship between the data subject and the initial receiver is required.

Alterations to the Data Flow

We have identified 5 different ways in which PETs can alter data flows: interception, narrowing, addition, redirection, and revelation. These categories are non-trivial, and we will provide justifications for them in the following

paragraphs. They should serve as a useful starting point for future work, as they give us an idea of what is required of a PET to successfully counter data flows that enable attacks.

Interception. In the case of data flow interception, data originally shared between two parties is either destroyed in a timely manner or is never even created. This technique can provide deniability properties and effectively remove most attack possibilities further down the data flow. However, achieving interception is difficult since no trace can be left anywhere in the software stack.

Defining interception is also challenging, as some properties like Differential Privacy (DP), which allows for deniability and thus "intercepts" the direct data flow. However, we argue that as long as some data is shared that the data subject has been involved with somehow, there is a remaining possibility for some information to leak, for example in combination with background information on the attacker's side. For this reason, we would put such properties in the next category.

Narrowing. In the case of data flow narrowing, PETs offer a way to reduce the sensitivity of data by means such as encryption, or by sending only a subset of the data after performing precomputations locally on a user's device. A data flow can also be narrowed by introducing trusted third parties between the sharing parties, who then perform computations that render the data less sensitive, as is often done for anonymous authentication.

Narrowing the data flow seems to be the most common way in which PETs are designed according to our data, but their effectiveness depends highly on how they are deployed, as there are risks for data leaks if the problem is not viewed from a holistic viewpoint. For instance, a PET could hide the identity of a data subject but not the data they share, which might allow for easy re-identification.

Note that as long as a data flow is still existing, it is difficult to know for sure whether it is sufficiently narrowed and an attack practically impossible. We encourage future work to break down variations of data flow narrowing further to understand which guarantees can be made on a practical level and whether they sufficiently protect the user's privacy.

Redirection. Redirects are commonly employed in anti-censorship infrastructure to bypass or deceive the original infrastructure entirely. The original parties (apart from the data subject) do not willingly cooperate in the deployment of the PET and are replaced with new entities. This distinguishes it from the technique of narrowing the data flow, where the original parties are deliberately shifted, but not replaced entirely.

The approach of redirecting data flows is intriguing from a contextual point of view, as some paths are considered "riskier" than others. Therefore, the decision to redirect is guided by trust and context, both of which can also change over time.

Often, this is combined with additional measures that provide more protection than simply a change of route, and many of these mechanisms also lead to a minimization of the data that is leaked on its way. Note that such technologies also heavily rely on the available infrastructure and might involve many entities in deployment. A good example for a redirecting PET is Tor [38].

Addition. A data flow can be added when certain data flows have been prevented by regulations or incentives that led to data not being shared, and a PET provides a way to comply with these regulations and still share data. PETs can enable private data sharing and computation, for example in the case of Secure Multi-Party Computation (SMPC) protocols.

Revelation. Finally, as discussed earlier, a data flow can also be revealed. Although it may seem counterintuitive, this can sometimes be a strategy for privacy protection, as it gives the data subject the power to make informed choices about their data. However, these tools also bring additional privacy challenges, as they require access to some information from the data subject.

In conclusion, this study has provided a first overview of how our framework can be applied, and generated interesting insights already. By analyzing different PETs and contextualizing how they alter data flows, we found that the modifications to data flows and control over data subjects' data are complex, and systemizing them makes discussions more accessible. The proposed classification option for PETs based on their impact on data flow provides a useful framework for discussions and analysis of PETs. We hope that our study contributes to the ongoing efforts to enhance privacy and security in the digital world. Further research is necessary to deepen our understanding of PETs and their effectiveness in mitigating real-life attacks.

6.3 Discussion

Our analysis of PETs has provided us with valuable insights in their impact on data flows. Through our examination of the involved parties, data alterations and other factors we have been able to discuss both gaps and and overlaps to our analysis in Chapter 5. While our findings require hardening by future work, through an extension of the data set and deeper investigations, we are confident that the result will provide a guide towards more holistic and user-centric PETs.

We contend that our approach in this work, which focuses on the comparison of our analysis of past privacy incidents to PETs, already provides a comprehensive overview of how our previously derived framework can be applied in thinking processes and decision-making. In the previous sections, we have successfully demonstrated how our privacy-based attack model and our taxonomy from Section 5.2.2 can be used for comparison and reflection.

In the following, we will discuss further certain topics that we have touched upon previously in more detail.

6.3.1 Trusting the Initial Receiver

It is clear from our analysis that PETs are often designed to be implemented by the initial receiver, who may retain control over the data subject's data, a situation that can seem counterintuitive. As a result of this situation, trust in the initial receiver is critical in many cases. Unfortunately, as we noted in Chapter 5, many initial receivers have demonstrated a lack of concern for their clients' privacy. Moreover, in situations where the data subject is an individual and the initial receiver is a large company or government entity, there is a significant power imbalance in terms of both resources and knowledge. These power imbalances further complicate issues of trust and highlight the importance of designing PETs with the data subject's privacy as a central focus.

To extend and reflect on our previous findings for a deeper understanding, we will now explore different scenarios to address the question of whether the initial receiver can be trusted in implementing PETs from the perspective of a user, where the initial receiver is either a company or government. We will provide three examples: (1) the PET is deployed only at the initial receiver who is a single entity, (2) the PET is deployed at multiple initial receivers, and (3) the PET is deployed both at the initial receiver and the data subject using a client-side application provided by the initial receiver. By analyzing these scenarios, we provide insight into various factors that contribute to the trustworthiness of initial receivers in implementing PETs.

Only at Single Entity Initial Receiver If the PET is solely deployed at the initial receiver, there is no way for the data subject to ensure that it is being used correctly. As a result, the data subject must rely on regulatory bodies such as the government. However, sufficient data protection legislation is required for this to occur. If the initial receiver releases data that has been altered by the PET, the data subject (or researchers) may attempt to verify whether the PET was utilised correctly. In general, this is the scenario in which the data subject has the least power.

Previous research has found that explaining the exact purpose of a PET improves user confidence and makes them more likely to share data [75].

However, this is just a social mechanism and does not offer the data subject any actual assurance that the initial receiver is not acting maliciously.

Therefore, to demonstrate trustworthiness, it is in the initial receiver's best interests to be as transparent as possible and to provide proof of their work, such as publishing the source code. However, a more interesting approach would be for the PET designers to facilitate this, by providing a mechanism that allows the initial receiver to prove to the public the correctness of their work.

Initial Receiver and Client-side Application In this scenario, the data subject has access to a client-side application, which is an improvement compared to the previous case. However, it is important to note that the initial receiver still has control over the data subject's data and can potentially remove the PET without the data subject's knowledge or consent. Again, the initial receiver can enhance their trustworthiness by releasing the source code for the application and allowing experts to verify if the application matches the source code.

However, despite this improvement, the data subject still lacks knowledge and is at risk. To address this, transparent and comprehensive verification mechanisms are necessary to enable informed decision-making. Context and incentives play a significant role in shaping how a situation is perceived, affecting a data subject's trust in a PET.

We urge researchers and deployers of PETs to take these situations and power imbalances seriously and to provide measures that enable the data subject to trust the proposed PET.

Multiple Initial Receivers Our analysis has shown that in cases where there are multiple initial receivers, they often have commercial or legal incentives to not share the data with each other. This implies that there is a greater likelihood that the initial receivers will deploy the PET correctly and not compromise the user's data. As a result, it is easier to trust the initial receivers in such situations. However, due to the limited data set we analyzed, we were not able to make further distinctions on the number of entities in the same role. Future work could explore the differences in challenges between central and decentralized deployment, the number of parties involved, and the level of their involvement.

We hope that this detailed thought process has deepened the understanding of how both our analyses can heavily interact in threat modeling processes. We argue that consistent use of both can provide future research with realistic, user-centric threat models.

6.3.2 After the Privacy Violation

We have observed that most PETs treat privacy as a binary concept and focus on maintaining specific properties to ensure privacy is "intact." However, we argue that privacy is a more holistic concept and that real-life attacks are often complex and require a chain of events to cause actual harm.

In a realistic scenario, data is typically collected first, then processed in some way, and often shared with other parties. This can lead to a potentially long chain of data handlers, who may also be able to infer new information from the data. Additionally, an attacker with a specific interest in this data is necessary for any actual harm to occur.

We recognize that once privacy has been compromised, the attack surface significantly expands, as an attacker may have multiple avenues to access certain data. However, if some of the actions that follow the privacy violation were prevented, the attack could become more challenging to execute, potentially deterring the attacker.

Therefore, we propose that it is worthwhile to explore the deployment of PETs even in situations where some privacy has already been compromised. It is important to consider that the scope of, for example, a disclosure is a crucial factor in determining the extent of harm that can be caused, and additional measures can still be taken to restrict the scope after a privacy violation. It is a common practice in corporate security to reduce the attack surface even without formal guarantees, and we suggest that this way of thinking could, in some contexts, also be beneficial in the context of privacy.

To achieve this, PETs need to be designed with a level of interoperability and context-awareness that allows them to understand their position in the data flow. Fortunately, there are existing efforts in this direction. For example, the UC framework [24] is a well-known example of a formal interoperability framework that is designed to ensure that protocols remain secure even when arbitrarily composed with other protocols.

6.3.3 A Long Way to Go

During our analysis of PETs, we have noted the absence of certain issues that have frequently arisen in our analysis of past privacy incidents. While our dataset of PETs may not be comprehensive, we want to draw attention to these gaps to raise awareness of these issues and provide avenues for future research.

Potential Gaps

Our analysis of PETs revealed a significant gap in addressing the issue of financial access to data, which is a prevalent problem in real-life attacks.

The lack of adequate pseudonymization and open sale of data makes it easier for attackers to exploit personal information. We urge researchers to prioritize this issue and explore potential solutions such as protecting the data subject's devices or, beyond PETs, advocating for laws against these practices.

We further observed a notable absence of focus on the issue of access to data by personal connections in our analysis of PETs. While such instances were infrequent in our examined attacks, previous research has shown that they are not uncommon in the general population [51, 133, 108]. Given the significance of our personal devices in our daily lives and how intimately we rely on them, we believe that this issue warrants greater attention. Additionally, we highlight the importance of safeguarding unlocked devices, as we identified one such example in our dataset.

Moreover, despite being the most commonly abused data type in our real-life attack dataset, we noticed a lack of attention to the issue of location data in the PETs analyzed. Location data is a particularly sensitive type of data as it can reveal a person's whereabouts over time, and just a few data points can be enough to re-identify an individual, as demonstrated by research on the uniqueness of human mobility patterns [35]. Thus, we advocate for the development of robust privacy-enhancing techniques that specifically address the protection of location data.

Our analysis of PETs revealed that most of them focus on protecting a data subject's identity, but other attributes also need to be safeguarded to prevent potential harm, as shown in prior research [125]. However, determining which data is sensitive and needs protection can be challenging, especially if the user is not aware of potential risks. Transparency and choice alone cannot resolve this issue, and we argue that the best approach is to share as little information as possible.

If information must be shared, users should have the option to delete it, as required by GDPR and other data protection laws [9]. However, we found few mechanisms for user control in the PETs we analyzed, despite the significant demand for such features, as reported in prior studies (see Chapter 3). Therefore, it is crucial to address these limitations and develop methods for user control and data protection.

Deployment

We cannot assume that the PETs we analyzed can be easily deployed as intended, as there are still many challenges that need to be addressed. For example, some applications are too slow to be practically useful, while others may amplify biases or require overly specific input for wide use. Additionally, many PETs require high infrastructure requirements and technical skills

for deployment and everyday use. Another issue is that many PETs are still limited in terms of scale and functionality. Finally, some technologies are difficult to deploy because they require a whole ecosystem to change, which is often unrealistic without legal requirements. For a detailed overview of the remaining deployment challenges of PETs, we refer to [96].

Finally, we must recognize that broader issues outside of PETs must also be addressed for widespread adoption. When designing a PET, it is important to consider not only how to deploy it, but also who will deploy it and how it will be used. This raises the critical question of how those who profit from personal data will deploy PETs. From our analysis of data flows and examples, we have observed that many privacy violations could have been avoided if companies did not sell user data. Therefore, when designing PETs, it is essential to consider the needs of both the company and the user and make an informed decision about who should deploy the PET and how it can cater to their specific needs.

It is worth noting that most of the privacy violations we encountered were legal. Hence, there is a need to examine whether laws on data protection are more effective than PETs. However, enforcing such laws remains challenging, with companies often keeping their inner workings secret during court trials. Therefore, we must also consider how to make data protection laws more enforceable. One potential solution could be to incorporate PETs into the enforcement process, as demonstrated by the US Census [122]. Ultimately, we must ask ourselves when laws are more efficient than PETs and what is needed to make them more enforceable.

6.3.4 Limitations

The small size of our dataset poses a challenge for drawing robust conclusions, and future work is required to harden our claims. Moreover, the conferences from which we selected papers may be WEIRD (Western, Educated, Industrialized, Rich, and Democratic), thus potentially limiting the diversity of perspectives and contexts represented in our analysis.

Additionally, the narrow scope required for publishing a paper often leads to the proposal of technical solutions that may not address privacy issues that occur within the entire ecosystem. It is crucial to recognize that attacks can transcend the boundaries of individual systems, and understand the importance of interoperability.

While academic papers may propose narrow solutions, our analyzed news articles often describe attacks that are only possible within a specific ecosystem, underscoring the need for more ecosystem-level research. Therefore, it is essential to continue investigating privacy attacks from a broader perspective to develop more effective solutions.

Chapter 7

Conclusion

This thesis introduces a novel framework, which is based on our own systemization of past privacy incidents, and serves to study the modus operandi of real-life attacks and PETs. This fosters an understanding of the requirements that existing and future PETs need to fulfill to protect the user's privacy efficiently. By making complex data flows manageable, comparable, studyable, and categorizable, we enable future researchers to build upon our work. Our approach has been demonstrated to be effective, providing compelling evidence of its practical application.

The Framework. Our taxonomy, in conjunction with our model of privacy-based attacks and observed attack data flows, provides a unique and comprehensive framework for researchers to review their threat models and ensure they have captured all relevant attack vectors. By integrating the user's lived experience, we offer a novel perspective that goes beyond re-evaluating and contextualizing existing threat modeling approaches. Our user-centric framework enables realistic and effective threat modeling, helping to identify and prevent threats, making it valuable for the development of PETs. In this way, our work offers new and innovative tools for the field of privacy research and PET development.

The Model. We have developed a sophisticated model for privacy-based attacks that aptly captures the intricacies of modern, internet-based, and highly interconnected attacks. The model not only streamlines the data flow, parties involved, and actions taken during an attack but also facilitates comparison across diverse attacks. Moreover, we have formally distinguished between the harm resulting from privacy violations and the violations themselves. By leveraging this model and conducting comprehensive analyses, we aim to facilitate the creation of more sophisticated and potent strategies for analyzing and mitigating privacy violations.

Our model has enabled us to successfully identify, study, and systematize privacy violations and their associated harms. As a result, we have developed a visualization of attack data flows and a comprehensive taxonomy of privacy violations and harms to better understand the modus operandi of real-life attacks.

The Taxonomy. Our taxonomy is a powerful tool that enables the creation of data flow visualizations from other data sources in the same manner as we did. It results from a codebook we derived by analyzing over 100 real-life attacks reported in news articles. The top-level categories in the codebook are universally applicable to data sources in the English language, as they are tied to our model of privacy-based attacks (Figure 4.2), which we have developed independently of the qualitative analysis in Chapter 4. The codebook's ability to collect important context from attacks with a minimal yet almost exhaustive set of options helps to improve the understanding of the attack landscape, which can guide the development of better protective measures.

For instance, our analysis identified notable differences in attacker attributes between individual-targeted attacks and untargeted attacks, as well as variations in the attacker's motivations. Understanding prevalent attack patterns can aid in determining whether users should be protected as individuals or in group settings. This knowledge can also inform the selection of effective protective mechanisms to avert specific attacks. Moreover, the attacker's intent is a relevant consideration when evaluating which parties to trust. Even otherwise trustworthy parties can cause harm due to incompetence or negligence. Therefore, in the context of threat modeling, trust should not only focus on malicious intent but also on the ability to function as intended.

The enhanced precision of data analysis employed by attackers, combined with the vast availability of large datasets, has rendered previously non-sensitive data types dangerous. Our research revealed location data as the most frequently exploited data type, often due to careless practices surrounding its collection and sharing. Additionally, we discovered that publicly accessible visual data could be leveraged to extract biometric information. Finally, when an attacker can access an unlocked device, it often results in unrestricted access to all data, an aspect largely overlooked in the design of existing protective measures. Our findings underscore the pressing need for a discourse on data sensitivity and guidelines for their secure sharing.

Beyond the attacks themselves, the codebook lists the consequences we identified, to facilitate reflection on the scope within which protective technologies should operate. Further, we hope to increase understanding of the pressing need for effective protective measures. We incorporated prior research into our analysis to ensure comprehensive coverage of potential on-

line harms, recognizing that rare or underreported harms may otherwise go overlooked.

Data Flows. The comprehensive approach of our framework that considers the entire attack data flow, along with its outcomes, is more effective than analyzing privacy violations in isolation. It enhances the observability of attack patterns and facilitates identifying critical areas requiring closer attention. Our results, including a visualization, allow us to observe the modus operandi of real-life attacks and identify areas that require urgent attention.

Formal threat models usually only consider accesses by technical means; our data shows that there is a palette of other options for attackers that are usually neglected. One of these options is buying barely "pseudonymized" data on the internet, which is legal in many cases. The same data and more is available in their original state if government bodies are interested and request access through the legal system. Further, we have found a non-negligible number of cases in which the attacker had physical access to a (potentially unlocked) device. These findings highlight the need for caution and stronger protections when sharing data, particularly with third-party entities. Acknowledging and integrating this problem into threat modeling is crucial to protect the user since we found evidence in prior work that they can not be expected to make informed decisions about whom to trust.

The attack data flows provide a clear and insightful view of possible attacks, and our visualization enables us to identify trends and gain a new perspective on situations, allowing us to determine where designers should best employ protective mechanisms. Our framework's extensive coverage makes it easier to identify unaddressed attack vectors. Our data displays the versatility of attacks as we find instances for all combinations of parties' categories. We emphasize that every combination and identified attack path deserves attention, as their existence indicates the potential for future attacks that can be prevented.

By carefully analyzing the actions associated with each data flow, we can identify effective measures to preempt potential attacks and develop user-centric protection mechanisms. For instance, the amount of data shared unknowingly by the user emphasizes the necessity for improved transparency, empowering users with greater control and choice over their information. Further, our data show that disclosing to third parties is the most common type of privacy violation. In combination with the frequent abuse of existing accesses, this presents a challenge for threat modeling as it implies that trust in any party should be limited, requiring potentially more complex solutions. Further, researchers must carefully consider the potential consequences of data being shared in unintended contexts and how this could be

avoided effectively.

Our approach not only fosters the discussion of potential solutions by observing data flows but also provides valuable insights into how and where the user's data flow could be controlled, as well as the consequences of design tradeoffs. This understanding of the attack landscape can help guide the development of more effective protective measures. We are confident that our findings can support researchers in adopting a more holistic and user-centric approach to threat modeling.

The Application. We showcase the practical applications of our framework by conducting a thorough analysis of existing PETs. Our investigation centers around the impact of these PETs on data flows and their potential to safeguard users from harm while taking into account the user's perspective. Our comprehensive framework enables us to make a meaningful comparison between individual PETs and draw insightful conclusions from the current state of research.

We argue that our methodology offers a valuable cross-sectional view of the PETs research field and propose a novel classification option based on their impact on data flow. These classifications can serve as a starting point for future work, as they highlight the necessary characteristics of a successful PET in countering data flows that enable attacks. This approach sets us apart and provides us with a valuable way to compare and evaluate PETs. Through our analysis, we gain insights into the current state of research and its impact on data flows and the user's control over their own data, thereby enhancing our understanding of privacy in practical settings.

We successfully identify common privacy violations and attacks that are being addressed, as well as those that remain unresolved. We found that some of the attacker capabilities and harms mentioned in the analyzed papers are relevant to real-life attacks. Notably, the research community recognizes the inadequacy of pseudonymization for de-identification, unauthorized disclosure as a significant source of privacy violations, the dangers of information inference, and deniability as an essential property in combatting compelled self-incrimination.

However, many critical situations and attacker types we identified in our framework remained unaddressed in our dataset. For example, no PET in our dataset could effectively protect against abuse by a personal connection, and the issue of financial access to data remained unaddressed. Further, while many PETs focused on hiding the data subject's identity, we argue that other attributes also need to be safeguarded to prevent harm, and if information must be shared, the data subject should have the option to delete it. Finally, it is crucial to recognize that a *combination* of guarantees, imple-

mented at various stages in the data flow, is most likely required to provide effective user protection.

We found that the alterations to data flows and control over data subjects' data are non-trivial, and systemizing them makes discussions more accessible. In the majority of analyzed PETs, the initial receiver had to be at least partially trusted and remained in control over the data subject's data. The privacy-based attack model has highlighted the crucial role of the initial receiver in deploying PETs, and our knowledge from past privacy incidents highlights the need to consider their trustworthiness, context, and incentives carefully to guarantee effective data protection.

Our investigation has revealed that research papers on PETs tend to focus on specific aspects of privacy, such as anonymity, rather than taking a holistic view of privacy as a concept. We acknowledge that this may be a consequence of the narrow scope that is required to publish a paper. However, using guarantees as a proxy for privacy instead of defining it, and treating it as a binary property, often results in only partial coverage of the attack paths and leaves it unclear whether PETs provide adequate protection for a user's data flow. These partial attack paths and the heavy focus on isolated guarantees sometimes move the data subject to the background, and other parties remain in control over its data. Further, some attack paths and contexts are not considered, and dangers potentially remain overlooked. Since multiple technologies may be needed to avert certain attacks fully, we would encourage a more holistic point of view in use-case examples so PETs become more interoperable, integrated, and effective, making them more likely to be deployed in real-world scenarios.

Our findings provide valuable insights into the gaps and overlaps in our analysis of past privacy incidents, and our analysis also uncovers mismatches between real-life attacks and existing PETs. Through these observations, we derive suggestions and guidance for future research enabling more holistic and user-centric PETs. While further investigations and data collection are necessary to strengthen our findings, we are confident that they will serve as a guide for future research, and although our approach has some limitations, we have shown that it yields interesting results.

Our approach, which compares PETs and analyzes past privacy incidents, demonstrates the versatility and utility of our framework for thinking processes and decision-making. We have demonstrated the efficacy of our privacy-based attack model and taxonomy as powerful tools for comparison and reflection, making our framework an essential and effective tool for PET discussions. We encourage further research to build upon our work to harden and extend this tool.

Future Work. Our work presents a solid foundation for others to build upon and take it to the next level. We have identified several areas where future work could extend our research. From a social perspective, we recommend conducting a user study to gather deeper real-life insights.

An important social aspect of protective mechanisms and whether they are effective is the user's perception of both the mechanism and their own risk for harm. A large-scale user survey could help identify these and provide a more comprehensive understanding of user perceptions of privacy violations and their consequences. We also need to understand whether users perceive these data accesses as intrusive and whether they expect harm from such practices. Adversarial models may further need to cover users' fears regarding privacy and unintended exposure, not just the dangers we have confirmed, to incorporate their perceptions. Ultimately, our goal is not just to mitigate "invisible" threats, but also to build user trust in PETs, which is critical for adoption.

Finally, we recommend advancing our framework by developing a standardized and modular formal language for threat modeling based on data flows. Such a language can effectively address emerging threats and be applied across diverse contexts and use cases.

The standardized vocabulary should provide a clear and accessible mapping between common privacy issues and both existing and envisioned privacy-preserving technologies that can be deployed to prevent them. Standardizing the vocabulary used in threat modeling processes is crucial for ensuring comparability and interoperability across different technologies, organizations, and industries.

Furthermore, we strongly advocate for the widespread use of formal threat and attacker models in the design of privacy-enhancing technologies. Together with the proposed language, this approach would render such models realistic enough to identify further potential gaps where necessary protection mechanisms have not yet been invented.

By implementing these recommendations, we can significantly enhance the privacy of data flows, creating a safer and more trustworthy digital environment for both individuals and organizations.

Appendix A

Appendix

List of Papers for PETs' Analysis

Title	Pub. Year	Award	Reference
Design and implementation of the idemix anonymous credential system	2002	Caspar Bowden Nominee	[23]
K-anonymity: a model for protecting privacy	2002	Caspar Bowden Nominee	[115]
Detecting Web Bugs with Bugnosis: Privacy Advocacy through Education	2003	Caspar Bowden Nominee	[13]
Rapid Mixing and Security of Chaum's Visual Electronic Voting	2003	Caspar Bowden Nominee	[64]
Tor: The Second-Generation Onion Router	2004	Usenix Test of Time Winner, Caspar Bowden Nominee	[38]
Private social network analysis: how to assemble pieces of a graph privately	2006	Caspar Bowden Runner-up	[53]
Blacklistable anonymous credentials: blocking misbehaving users without ttps	2007	Caspar Bowden Runner-up	[119]
An Ad Omnia Approach to Defining and Achieving Private Data Analysis	2008	Caspar Bowden Winner	[40]

Continued on next page

continued from previous page

List of Papers for PETs' Analysis			
Title	**Pub. Year**	**Award**	**Reference**
Adversarial stylometry: Circumventing authorship recognition to preserve privacy and anonymity	2012	Caspar Bowden Winner	[22]
Eternal Sunshine of the Spotless Machine: Protecting Privacy with Ephemeral Channels	2012	Caspar Bowden Runner-up	[39]
Telex: Anticensorship in the Network Infrastructure	2011	Caspar Bowden Runner-up	[132]
A Scanner Darkly: Protecting User Privacy from Perceptual Applications	2013	Caspar Bowden Winner	[72]
Zerocoin: Anonymous Distributed E-Cash from Bitcoin	2013	Caspar Bowden Runner-up	[88]
RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response	2014	Caspar Bowden Runner-up	[43]
Riposte: An Anonymous Messaging System Handling Millions of Users	2015	Caspar Bowden Winner	[30]
Students and taxes: A privacy-preserving study using secure computation	2016	Caspar Bowden Runner-up	[19]
DatashareNetwork: A Decentralized Privacy-Preserving Search Engine for Investigative Journalists	2020	Caspar Bowden Runner-up	[42]

Table A.1: List of the 17 papers used in the qualitative study in chapter 6, Pub. Year indicates the year the paper was published

Bibliography

[1] FTC policy statement signals increasing scrutiny on the protection of sensitive personal health information. https://www.wiley.law/newsletter-Sep-2021-PIF-FTC-Policy-Statement-Signals-Increasing-Scrutiny-on-the-Protection-of-Sensitive-Personal-Health-Information. Accessed: 2023-2-3.

[2] International standards. https://www.ohchr.org/en/special-procedures/sr-privacy/international-standards. Accessed: 2023-2-3.

[3] Katz v. united states. https://constitutioncenter.org/the-constitution/supreme-court-case-library/katz-v-united-states. Accessed: 2023-3-17.

[4] LINDDUN. https://www.linddun.org/. Accessed: 2023-3-20.

[5] Threat modeling - OWASP cheat sheet series. https://cheatsheetseries.owasp.org/cheatsheets/Threat_Modeling_Cheat_Sheet.html. Accessed: 2023-3-20.

[6] Children's online privacy protection rule ("COPPA"). https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa, July 2013. Accessed: 2023-2-3.

[7] USENIX test of time awards. https://www.usenix.org/conferences/test-of-time-awards, September 2013. Accessed: 2023-3-17.

[8] California consumer privacy act (CCPA). https://oag.ca.gov/privacy/ccpa, October 2018. Accessed: 2022-10-13.

[9] What are the main aspects of the general data protection regulation (GDPR) that a public administration should be aware of? https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/public-administrations-and-data-protection/what-are-main-aspects-general-data-protection-regulation-gdpr-public-administration-should-be-aware_en, January 2018. Accessed: 2022-10-13.

[10] Health insurance portability and accountability act of 1996 (HIPAA). https://www.cdc.gov/phlp/publications/topic/hipaa.html, June 2022. Accessed: 2023-2-3.

[11] Alessandro Acquisti and Jens Grossklags. Privacy and rationality in individual decision making. *IEEE Secur. Priv.*, 3(1):26–33, January 2005.

[12] Martin R Albrecht, Jorge Blasco, Rikke Bjerg Jensen, and Lenka Mareková. Collective information security in large-scale urban protests: The case of hong kong. *Proceedings of the 30th USENIX Security Symposium*, May 2021.

[13] Adil Alsaid and David Martin. Detecting web bugs with bugnosis: Privacy advocacy through education. In *Privacy Enhancing Technologies*, pages 13–26. Springer Berlin Heidelberg, 2003.

[14] Miguel E Andrés, Nicolás E Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Geo-indistinguishability: differential privacy for location-based systems. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, CCS '13, pages 901–914, New York, NY, USA, November 2013. Association for Computing Machinery.

[15] Roland Assam and Thomas Seidl. A model for Context-Aware location identity preservation using differential privacy. In *2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, pages 346–353. ieeexplore.ieee.org, July 2013.

[16] Johana Bhuiyan. Facebook gave police their private data. now, this duo face abortion charges. *The Guardian*, August 2022.

[17] Cara Bloom. Privacy threat modeling. https://www.usenix.org/system/files/pepr22_slides_bloom.pdf, 2022. Accessed: 2023-3-20.

[18] Manuel Blum, Paul Feldman, and Silvio Micali. Non-interactive zero-knowledge and its applications. In *Proceedings of the twentieth annual ACM symposium on Theory of computing*, STOC '88, pages 103–112, New York, NY, USA, January 1988. Association for Computing Machinery.

[19] Dan Bogdanov, Liina Kamm, Baldur Kubo, Reimo Rebane, Ville Sokk, and Riivo Talviste. Students and taxes: A privacy-preserving study using secure computation. *Proc. Priv. Enhancing Technol.*, 2016(3):117–135, July 2016.

[20] Sissela Bok. Secrets: On the ethics of concealment and revelation. *Mod. Law Rev.*, 50(1):125–128, 1987.

[21] Michelle Boorstein, Marisa Iati, and Annys Shin. Top U.S. catholic church official resigns after cellphone data used to track him on grindr and to gay bars. *The Washington Post*, July 2021.

[22] Michael Brennan, Sadia Afroz, and Rachel Greenstadt. Adversarial stylometry: Circumventing authorship recognition to preserve privacy and anonymity. *ACM Trans. Inf. Syst. Secur.*, 15(3):1–22, November 2012.

[23] Jan Camenisch and Els Van Herreweghen. Design and implementation of the idemix anonymous credential system. In *Proceedings of the 9th ACM conference on Computer and communications security*, CCS '02, pages 21–30, New York, NY, USA, November 2002. Association for Computing Machinery.

[24] Ran Canetti. Universally composable security: A new paradigm for cryptographic protocols. *Cryptology ePrint Archive*, 2000.

[25] Varun Chandrasekaran, Chuhan Gao, Brian Tang, Kassem Fawaz, Somesh Jha, and Suman Banerjee. Face-Off: Adversarial face obfuscation. In *Proceedings on Privacy Enhancing Technologies*, 2021.

[26] Thomas Cilloni, Wei Wang, Charles Walter, and Charles Fleming. Ulixes: Facial recognition privacy with adversarial machine learning. *Proceedings on Privacy Enhancing Technologies*, 2022.

[27] Russell L Ciochon. *Privacy and Personality*. Routledge, 1 edition, September 2017.

[28] Jessica Colnago, Lorrie Cranor, and Alessandro Acquisti. Is there a reverse privacy paradox? an exploratory analysis of gaps between privacy perspectives and privacy-seeking behaviors. *Proc. Priv. Enhancing Technol.*, 2023(1):455–476, January 2023.

[29] Kovila P L Coopamootoo, Maryam Mehrnezhad, and Ehsan Toreini. *"I feel invaded, annoyed, anxious and I may protect myself": Individuals Feelings about Online Tracking and their Protective Behaviour across Gender and Country*. USENIX Association, 2022.

[30] Henry Corrigan-Gibbs, Dan Boneh, and David Mazières. Riposte: An anonymous messaging system handling millions of users. In *2015 IEEE Symposium on Security and Privacy*, pages 321–338. ieeexplore.ieee.org, May 2015.

[31] Joseph Cox. How the U.S. military buys location data from ordinary apps. https://www.vice.com/en/article/jgqm5x/us-military-location-data-xmode-locate-x, November 2020. Accessed: 2022-10-10.

[32] Rachel Cummings, Gabriel Kaptchuk, and Elissa M Redmiles. "i need a better description": An investigation into user expectations for differential privacy. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, CCS '21, pages 3037–3052, New York, NY, USA, November 2021. Association for Computing Machinery.

[33] Sauvik Das, Adam D I Kramer, Laura A Dabbish, and Jason I Hong. The role of social influence in security feature adoption. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, pages 1416–1426, New York, NY, USA, February 2015. Association for Computing Machinery.

[34] Lydia de la Torre. A guide to the california consumer privacy act of 2018. November 2018.

[35] Yves-Alexandre de Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. Unique in the crowd: The privacy bounds of human mobility. *Sci. Rep.*, 3:1376, 2013.

[36] Judith Wagner DeCew. The scope of privacy in law and ethics. *Law Philos.*, 5(2):145–173, 1986.

[37] Martin Degeling, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. We value your privacy ... now take some cookies. *Informatik Spektrum*, 42(5):345–346, October 2019.

[38] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The Second-Generation onion router. In *13th USENIX Security Symposium (USENIX Security 04)*, 2004.

[39] Alan M Dunn, Michael Z Lee, Suman Jana, Sangman Kim, Mark Silberstein, Yuanzhong Xu, Vitaly Shmatikov, and Emmett Witchel. Eternal sunshine of the spotless machine: Protecting privacy with ephemeral channels. *Proc USENIX Symp Oper Syst Des Implement (OSDI)*, pages 61–75, 2012.

[40] Cynthia Dwork. An ad omnia approach to defining and achieving private data analysis. In *Privacy, Security, and Trust in KDD*, pages 1–13. Springer Berlin Heidelberg, 2008.

[41] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*, pages 265–284. Springer Berlin Heidelberg, 2006.

[42] Kasra Edalatnejad, Wouter Lueks, Julien Pierre Martin, Soline Ledésert, Bruno Thomas, Laurent Girod, and Carmela Troncoso. DatashareNetwork: A decentralized Privacy-Preserving search engine for investigative journalists. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 1911–1927, 2020.

[43] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. RAPPOR: Randomized aggregatable Privacy-Preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, pages 1054–1067, New York, NY, USA, November 2014. Association for Computing Machinery.

[44] David Evans, Vladimir Kolesnikov, and Mike Rosulek. A pragmatic introduction to secure multi-party computation. *Found. Trends® Priv. Secur.*, 2(2-3):70–246, 2018.

[45] Ivan Evtimov, Pascal Sturmfels, and Tadayoshi Kohno. FoggySight: A scheme for facial lookup privacy. *Proceedings on Privacy Enhancing Technologies*, 2021.

[46] Ming Fan, Le Yu, Sen Chen, Hao Zhou, Xiapu Luo, Shuyue Li, Yang Liu, Jun Liu, and Ting Liu. An empirical evaluation of GDPR compliance violations in android mhealth apps. In *2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE)*, pages 253–264, October 2020.

[47] Kohut Fellows and Mitofsky Graduate Fellows. New york times. https://ropercenter.cornell.edu/new-york-times. Accessed: 2023-3-17.

[48] David Ferraiolo, Janet Cugini, and Richard Kuhn. Role-Based access control (RBAC): Features and motivations. In *Proceedings of the 11th Annual Computer Security Applications Conference*, pages 241–248. IEEE, December 1995.

[49] Ferran, Lee. Ex-NSA chief: 'we kill people based on metadata'. https://abcnews.go.com/blogs/headlines/2014/05/ex-nsa-chief-we-kill-people-based-on-metadata, May 2014. Accessed: 2022-10-10.

[50] Diana Freed, Jackeline Palmer, Diana Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. "a stalker's paradise": How intimate partner abusers exploit technology. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, number Paper 667 in CHI '18, pages 1–13, New York, NY, USA, April 2018. Association for Computing Machinery.

[51] Diana Freed, Jackeline Palmer, Diana Elizabeth Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. Digital technologies and intimate partner violence: A qualitative analysis with multiple stakeholders. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW):1–22, December 2017.

[52] Alisa Frik, Leysan Nurgalieva, Julia Bernd, Joyce Lee, Florian Schaub, and Serge Egelman. Privacy and security threat models and mitigation strategies of older adults. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*, pages 21–40, 2019.

[53] Keith B Frikken and Philippe Golle. Private social network analysis: how to assemble pieces of a graph privately. In *Proceedings of the 5th ACM workshop on Privacy in electronic society*, WPES '06, pages 89–98, New York, NY, USA, October 2006. Association for Computing Machinery.

[54] Caspar Bowden Legacy Fund. The caspar bowden award for outstanding research in privacy enhancing technologies. https://petsymposium.org/award/. Accessed: 2023-3-17.

[55] Muskan Garg. Mental health analysis in social media posts: A survey. *Arch. Comput. Methods Eng.*, 30(3):1–24, January 2023.

[56] Ruth Gavison. Privacy and the limits of law. *Yale Law J.*, 89(3):421–471, 1980.

[57] Christine Geeng, Mike Harris, Elissa Redmiles, and Franziska Roesner. "like lesbians walking the perimeter": Experiences of U.S. LGBTQ+ folks with online security, safety, and privacy advice. In *Proceedings of the 31st USENIX Security Symposium*, August 2022.

[58] Craig Gentry. *A fully homomorphic encryption scheme*. PhD thesis, Stanford, CA, USA, 2009.

[59] Nina Gerber, Paul Gerber, and Melanie Volkamer. Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. *Comput. Secur.*, 77:226–261, August 2018.

[60] Nina Gerber, Verena Zimmermann, and Melanie Volkamer. Why johnny fails to protect his privacy. In *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, June 2019.

[61] Tom Gerety. Redefining privacy. *Harv. C.R.-C.L. L. Rev.*, 12(2):233–296, 1977.

[62] Robert S Gerstein. Intimacy and privacy. In *Philosophical Dimensions of Privacy: An Anthology*, pages 265–271. Cambridge University Press, November 1984.

[63] Angelica Goetzen, Samuel Dooley, and Elissa M Redmiles. Ctrl-Shift: How privacy sentiment changed from 2019 to 2021. *Proceedings on Privacy Enhancing Technologies*, 2022.

[64] Marcin Gomułkiewicz, Marek Klonowski, and Mirosław Kutyłowski. Rapid mixing and security of chaum's visual electronic voting. In *Computer Security – ESORICS 2003*, pages 132–145. Springer Berlin Heidelberg, 2003.

[65] Danny S Guamán, Jose M Del Alamo, and Julio C Caiza. GDPR compliance assessment for Cross-Border personal data transfers in android apps. *IEEE Access*, 9:15961–15982, 2021.

[66] Marian Harbach, Sascha Fahl, and Matthew Smith. Who's afraid of which bad wolf? a survey of IT security risk awareness. In *2014 IEEE 27th Computer Security Foundations Symposium*, pages 97–110, July 2014.

[67] Alex Hern. Uber employees 'spied on ex-partners, politicians and beyoncé'. *The Guardian*, December 2016.

[68] Kashmir Hill. Another arrest, and jail time, due to a bad facial recognition match. *The New York Times*, December 2020.

[69] Adele E Howe, Indrajit Ray, Mark Roberts, Malgorzata Urbanska, and Zinta Byrne. The psychology of security for the home computer user. In *2012 IEEE Symposium on Security and Privacy*, pages 209–223, May 2012.

[70] Julie C Inness. *Privacy, Intimacy, and Isolation*. Oxford University Press, 1992.

[71] Michael J Jacobson. Problem solving, cognition, and complex systems: differences between experts and novices. *Complexity*, 6(3):41–49, January 2001.

[72] Suman Jana, Arvind Narayanan, and Vitaly Shmatikov. A scanner darkly: Protecting user privacy from perceptual applications. In *2013 IEEE Symposium on Security and Privacy*, pages 349–363, May 2013.

[73] David Jonassen and Young Hoan Cho. Externalizing mental models with mindtools. In Dirk Ifenthaler, Pablo Pirnay-Dummer, and J Michael Spector, editors, *Understanding Models for Learning and Instruction*, pages 145–159. Springer US, Boston, MA, 2008.

[74] Bailey Kacsmar, Vasisht Duddu, Kyle Tilbury, Blase Ur, and Florian Kerschbaum. Comprehension from chaos: What users understand and expect from private computation. November 2022.

[75] Bailey Kacsmar, Kyle Tilbury, Miti Mazmudar, and Florian Kerschbaum. Caring about sharing: User perceptions of multiparty data sharing. https://bkacsmar.github.io/files/caringsharingpaper.pdf, August 2022. Accessed: 2022-11-10.

[76] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. "my data just goes everywhere:" user mental models of the internet and implications for privacy and security. In *Eleventh Symposium on Usable Privacy and Security (SOUPS 2015)*, pages 39–52. usenix.org, 2015.

[77] Kenneth L Karst. "the files": Legal controls over the accuracy and accessibility of stored personal data. *Law Contemp. Probl.*, 31(2):342–376, 1966.

[78] Sabrina Karwatzki, Manuel Trenz, Virpi Kristiina Tuunainen, and Daniel Veit. Adverse consequences of access to individuals' information: an analysis of perceptions and the scope of organisational influence. *Eur. J. Inf. Syst.*, 26(6):688–715, November 2017.

[79] Jennifer King. How come i'm allowing strangers to go through my phone? smartphones and privacy expectations. March 2012.

[80] Jakub Konečný, H Brendan McMahan, Daniel Ramage, and Peter Richtárik. Federated optimization: Distributed machine learning for On-Device intelligence. October 2016. only available on arxiv.

[81] Priya Kumar, Shalmali Milind Naik, Utkarsha Ramesh Devkar, Marshini Chetty, Tamara L Clegg, and Jessica Vitak. 'no telling passcodes out because they're private': Understanding children's mental models of privacy and security online. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW):1–21, December 2017.

[82] Boris Lubarsky. RE-IDENTIFICATION OF "ANONYMIZED DATA". https://georgetownlawtechreview.org/wp-content/uploads/2017/04/Lubarsky-1-GEO.-L.-TECH.-REV.-202.pdf, 2017. Accessed: 2023-3-17.

[83] Jane Martinson. The virtues of vice: how punk magazine was transformed into media giant. *The Guardian*, January 2015.

[84] Jonathan R Mayer and John C Mitchell. Third-Party web tracking: Policy and technology. In *2012 IEEE Symposium on Security and Privacy*, pages 413–427, May 2012.

[85] Allison McDonald, Catherine Barwulor, Michelle L Mazurek, Florian Schaub, and Elissa M Redmiles. "it's stressful having all these phones": Investigating sex workers' safety goals, risks, and practices online. In *Proceedings of the 30th USENIX Security Symposium*, August 2021.

[86] Maryam Mehrnezhad, Kovila Coopamootoo, and Ehsan Toreini. How can and would people protect from online tracking? *Proc. Priv. Enhancing Technol.*, 2022(1):105–125, January 2022.

[87] William Melicher, Mahmood Sharif, Joshua Tan, Lujo Bauer, Mihai Christodorescu, and Pedro Giovanni Leon. (do not) track me sometimes: Users' contextual preferences for web tracking. *Proceedings on Privacy Enhancing Technologies*, 2016(2):135–154, April 2016.

[88] Ian Miers, Christina Garman, Matthew Green, and Aviel D Rubin. Zerocoin: Anonymous distributed E-Cash from bitcoin. In *2013 IEEE Symposium on Security and Privacy*, pages 397–411, May 2013.

[89] Barrington Moore. *Privacy: Studies in Social and Cultural History*. M.E. Sharpe, 1984.

[90] Ben Nassi, Ron Bitton, Ryusuke Masuoka, Asaf Shabtai, and Yuval Elovici. SoK: Security and privacy in the age of commercial drones. In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 1434–1451, May 2021.

[91] Thomas Neubauer and Johannes Heurix. A methodology for the pseudonymization of medical data. *Int. J. Med. Inform.*, 80(3):190–204, March 2011.

[92] Helen Nissenbaum. Privacy as contextual integrity. *Wash Law Rev.*, 79(1):119, 2004.

[93] Thomas B Norton. The Non-Contractual nature of privacy policies and a new critique of the notice and choice privacy protection model. *Fordham Intellectual Property, Media and Entertainment Law Journal*, 27(1):181, 2016.

[94] Maggie Oates, Yama Ahmadullah, Abigail Marsh, Chelse Swoopes, Shikun Zhang, Rebecca Balebako, and Lorrie Faith Cranor. Turtles, locks, and bathrooms: Understanding mental models of privacy through illustration. *Proc. Priv. Enhancing Technol.*, 2018(4):5–32, October 2018.

[95] O'Connor. Planned parenthood of southeastern pa. v. casey, 505 U.S. 833 (1992). https://www.law.cornell.edu/supct/html/91-744.ZO.html. Accessed: 2023-3-20.

[96] OECD. EMERGING PRIVACY ENHANCING TECHNOLOGIES, 2023.

[97] Office of the Privacy Commissioner of Canada. PIPEDA in brief. https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/pipeda_brief/, January 2018. Accessed: 2023-3-2.

[98] Grace Oldham. Facebook and Anti-Abortion clinics are collecting highly sensitive info on Would-Be patients – the markup. https://themarkup.org/pixel-hunt/2022/06/15/facebook-and-anti-abortion-clinics-are-collecting-highly-sensitive-info-on-would-be-patients, June 2022. Accessed: 2022-7-11.

[99] Nicolas Papernot, Patrick McDaniel, Arunesh Sinha, and Michael P Wellman. SoK: Security and privacy in machine learning. In *2018 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 399–414, April 2018.

[100] Jon Penney. Internet surveillance, regulation, and chilling effects online: A comparative case study. May 2017.

[101] Richard A Posner. *The Economics of Justice*. Harvard University Press, August 1983.

[102] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. How I learned to be secure: a Census-Representative survey of security advice sources and behavior. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, pages 666–677, New York, NY, USA, October 2016. Association for Computing Machinery.

[103] Priscilla M Regan. *Legislating Privacy: Technology, Social Values, and Public Policy*. Univ of North Carolina Press, 1995.

[104] Eric Rosenbaum. 5 biggest risks of sharing your DNA with consumer genetic-testing companies. https://www.cnbc.com/2018/06/16/5-biggest-risks-of-sharing-dna-with-consumer-genetic-testing-companies.html, 2018. Accessed: 2023-3-17.

[105] Mahsa Saeidi, Mckenzie Calvert, Audrey W Au, Anita Sarma, and Rakesh B Bobba. If this context then that concern: Exploring users' concerns with IFTTT applets. *Proceedings on Privacy Enhancing Technologies*, 2022.

[106] Johnny Saldaña. Coding and analysis strategies. In *The Oxford Handbook of Qualitative Research*. July 2014.

[107] Pierangela Samarati and Latanya Sweeney. Protecting privacy when disclosing information: K-anonymity and its enforcement through generalization and suppression. https://dataprivacylab.org/dataprivacy/projects/kanonymity/paper3.pdf, 1998. Accessed: 2022-10-15.

[108] Nithya Sambasivan, Amna Batool, Nova Ahmed, Tara Matthews, Kurt Thomas, Laura Sanely Gaytán-Lugo, David Nemer, Elie Bursztein, Elizabeth Churchill, and Sunny Consolvo. "they don't leave us alone anywhere we go": Gender and digital abuse in south asia. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, number Paper 2 in CHI '19, pages 1–14, New York, NY, USA, May 2019. Association for Computing Machinery.

[109] Michael J Sandel. *Democracy's Discontent*. Harvard University Press, October 2022.

[110] Morgan Klaus Scheuerman, Jialun Aaron Jiang, Casey Fiesler, and Jed R Brubaker. A framework of severity for harmful content online. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2):1–33, October 2021.

[111] Theodor Schnitzler, Shujaat Mirza, Markus Dürmuth, and Christina Pöpper. SoK: Managing longitudinal privacy of publicly shared personal online data. *Proceedings on Privacy Enhancing Technologies*, 2021.

[112] Daniel J Solove. A taxonomy of privacy. *Univ. PA Law Rev.*, 154(3):477–564, 2006.

[113] Daniel J Solove. Understanding privacy. May 2008.

[114] Spyros Kokolakis. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Comput. Secur.*, 64:122–134, January 2017.

[115] Latanya Sweeney. k-ANONYMITY: A MODEL FOR PROTECTING PRIVACY. *Int. J. Uncertainty Fuzziness Knowledge Based Syst.*, 10(05):557–570, October 2002.

[116] Tewari, Shreya and Walter-Johnson, Fikayo. New records detail DHS purchase and use of vast quantities of cell phone location data. https://www.aclu.org/news/privacy-technology/new-records-detail-dhs-purchase-and-use-of-vast-quantities-of-cell-phone-location-data, July 2022. Accessed: 2022-10-10.

[117] Wiebke Thode, Joachim Griesbaum, and Thomas Mandl. "i would have never allowed it": User perception of third-party tracking and implications for display advertising. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1008.601&rep=rep1&type=pdf, 2015. Accessed: 2022-7-11.

[118] Kurt Thomas, Devdatta Akhawe, Michael Bailey, Dan Boneh, Elie Bursztein, Sunny Consolvo, Nicola Dell, Zakir Durumeric, Patrick Gage Kelley, Deepak Kumar, Damon McCoy, Sarah Meiklejohn, Thomas Ristenpart, and Gianluca Stringhini. SoK: Hate, harassment, and the changing landscape of online abuse. In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 247–267, May 2021.

[119] Patrick P Tsang, Man Ho Au, Apu Kapadia, and Sean W Smith. Blacklistable anonymous credentials: blocking misbehaving users without ttps. In *Proceedings of the 14th ACM conference on Computer and communications security*, CCS '07, pages 72–81, New York, NY, USA, October 2007. Association for Computing Machinery.

[120] unknown. U. s. DEPT. OF JUSTICE v. REPORTERS COMMITTEE, 1989.

[121] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, number Article 4 in SOUPS '12, pages 1–15, New York, NY, USA, July 2012. Association for Computing Machinery.

[122] US Census Bureau. Understanding differential privacy. https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/differential-privacy.html. Accessed: 2023-2-3.

[123] Paul Voigt and Axel von dem Bussche. *The EU General Data Protection Regulation (GDPR)*. Springer International Publishing.

[124] Isabel Wagner and David Eckhoff. Technical privacy metrics: A systematic survey. *ACM Comput. Surv.*, 51(3):1–38, June 2018.

[125] Noel Warford, Tara Matthews, Kaitlyn Yang, Omer Akgul, Sunny Consolvo, Patrick Gage Kelley, Nathan Malkin, Michelle L Mazurek, Manya Sleeper, and Kurt Thomas. SoK: A framework for unifying At-Risk user research. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 2344–2360, May 2022.

[126] Samuel D Warren and Louis D Brandeis. The right to privacy. *Harv. Law Rev.*, 4(5):193–220, 1890.

[127] Rick Wash. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, number Article 11 in SOUPS '10, pages 1–16, New York, NY, USA, July 2010. Association for Computing Machinery.

[128] Alan Westin. Privacy and freedom, 1967.

[129] Daniel W Woods and Rainer Böhme. SoK: Quantifying cyber risk. In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 211–228, May 2021.

[130] Wu, Yuxi and Edwards, W. Keith and Das, Sauvik. SoK: Social cybersecurity. In *2022 IEEE Symposium on Security and Privacy (SP)*, volume 0, pages 1863–1879, May 2022.

[131] G Wunsch. Theories, models, and data. *Demografie*, 36(1):20–29, 1994.

[132] Eric Wustrow, Scott Wolchok, Ian Goldberg, and J Alex Halderman. Telex: Anticensorship in the network infrastructure. In *USENIX Security Symposium*, page 45, 2011.

[133] Kathryn Ybarra, michele and Price-Feeney, Myeshia and Lenhart, Amanda and Zickuhr. INTIMATE PARTNER DIGITAL ABUSE, 2017.

[134] Zakrzewski, Cat and Verma, Pranshu and Parker, Claire. Texts, web searches about abortion have been used to prosecute women. *The Washington Post*, July 2022.