Carnegie Mellon University, MSIT-Privacy Engineering Capstone Project, November 2018

Evaluating Privacy Enhancing Technologies for Organizations

Ao Chen, Jeremy Thomas
Advisor: Nicolas Christin

1. Executive Summary

Given the changing perspective on privacy in today's society, organizations must adapt to a growing set of requirements and constraints on practices that impact privacy. Individual choices about privacy, exposure of broken privacy practices, and expanding regulations force organizations towards expanding the influence of accepted principles of privacy throughout business operations. All of these factors encourage the use of technologies to provide stronger privacy guarantees and to directly address the privacy challenges these requirements create in an effective and efficient manner.

To better understand the use of privacy enhancing technologies (PETs) by organizations, we interviewed a set of privacy experts working across various industries and in multiple disciplines. We conducted 20 interviews of privacy experts from September 2018 to November 2018. We interviewed most participants over the phone or through video conferencing applications, and typically the interviews lasted 30 to 60 minutes.

To compare each technology's effectiveness and efficiency at addressing the privacy challenges of organizations, we first categorized privacy technologies based on their contributions towards the three primary objectives of privacy engineering as defined by The National Institute of Standards and Technology (NIST): predictability, manageability, and disassociability. Within each of these objectives, we determined the effectiveness of each technology in addressing privacy harms. For example, de-identification techniques increase the disassociability of a data subject within a dataset, and their effectiveness derives from the strength of de-identification (e.g. the value of *'k*' in *k*-anonymity). This reduces privacy risks for the individual such as those from undesirable disclosures of the dataset. The efficiency of any technology contains both the costs of implementation and the reduction, if any, in utility of the related practices to the organization. Continuing with the example of de-identification techniques, these techniques preserve the privacy of an individual within a dataset, but they potentially reduce the utility that an organization can extract from the dataset. In this way, we explored several privacy enhancing technologies that we identified through our interview process. Some key technologies include:

- Differential Privacy implementations at Microsoft, Google, Apple, and Uber
- Private Set Intersection as implemented by Google and Mastercard
- User access and control platforms such as the Data Transfer Project

Many of these technologies remain emerging technologies with most innovation occurring through academic and industry research such as differential privacy and private set intersection. Technology leaders (Google, Apple, Microsoft), represent the early adopters of many of these technologies. However, in many cases, these technologies expose the lack of maturity in the market for privacy enhancing technologies. More mature, PETs exist (access controls, audit logs, data tagging and mapping, etc.) and align with current regulations. Therefore, organizations must first apply mature PETs to safeguards on compliant data practices, and then layer emerging privacy technologies within products and services to protect higher order privacy requirements such as ethical and responsible data uses.

1. Executive Summary **2**

2. Abstract **5**

3. Introduction **5**

4. Background **6**

4.1. What are Privacy Enhancing Technologies? 6

4.2. Types of PETs 6

4.3. Requirements of Privacy Enhancing Technologies 8

4.4. Assessment of PETs 9

5. Methodology **11**

5.1. Expert Interviews 11

5.2. Publicly Available Sources 12

5.3. Components of Analysis 12

6. Results **13**

6.1. Privacy Preserving Approaches to Data Collection 13

6.1.1. Direct Identifiers 14

6.1.2. Indirect Identifiers 15

6.1.3. Differential Privacy for Telemetry 17

6.1.3.1. Apple Differential Privacy 17

6.1.3.2. Google Chrome RAPPOR 18

6.1.3.3. Uber Elastic Sensitivity 19

6.1.4. Differential Privacy Summary 20

6.2. Privacy Preserving Approaches to Data Use 20

6.2.1. Federated Learning 21

6.2.2. PATE 22

6.2.3. Private Set Intersection 23

6.2.4. Privacy Preserving Data Usage Summary 23

6.3. Safeguards on Data Practices 24

6.4. Accommodating Users 25

6.4.1. User Access and Control 25

6.4.2. Informed Consent 26

7. Discussion **27**

7.1. When Disassociability Makes Sense 27

7.2. A Predictable and Manageable Architecture Enforces Privacy Practices 28

7.3. Technology Necessary, Not Sufficient 29

8. Conclusion **29**

9. References **30**

10. Appendix I **38**

10.1. Interview Recruitment Text 38

10.2. Interview Script 38

10.2.1. Introduction 38

10.2.2. Questions 38

11. Appendix II **39**

11.1. Non-organizational Technologies 39

11.2. Research in Stronger Privacy Guarantees 40

11.3. Security Safeguards 41

2. Abstract

Organizations need to address a growing set of privacy challenges and privacy enhancing technologies (PETs) represent viable solutions for many of these issues. Additionally, privacy laws motivate organizations towards improved data practices and governance, often supported by these technologies. Most technologies provide solutions to known privacy harms by supporting commonly accepted principles of privacy. To better understand the use of privacy enhancing technologies by organizations, we interviewed a set of privacy experts working across various industries and in multiple disciplines. Based on these interviews, we analyzed a subset of the PETs used by organizations to determine their significance within the overall privacy objectives of organizations. The basic set of privacy technologies provide access controls, audit logs, and platforms for facilitating notice, choice, access, and portability requirements. Meanwhile, technologies for enabling data analysis while preserving privacy continue to evolve and remain limited in availability and adoption, yet they continue to address a growing intersection of privacy constraints and business objectives. Applying policies for ethical or reasonable uses of data in a generic and consistent approach also remains a difficult goal that technologies cannot single-handedly address.

3. Introduction

Given the changing perspective on privacy in today's society, organizations must adapt to a growing set of requirements and constraints on practices that impact privacy. Individual choices about privacy, exposure of broken privacy practices, and expanding regulations force organizations towards expanding the influence of accepted principles of privacy throughout business operations [84, 85, 86]. This encourages the use of technologies to provide stronger privacy guarantees and to directly address the privacy challenges these requirements create in an effective and efficient manner.

One primary driving force, which we include throughout our analysis, appears in the enactment of the General Data Protection Regulation (GDPR) within the European Union (EU) [29]. The GDPR is a comprehensive data privacy regulation with significant enforcement mechanisms (e.g. fines of up to four percent of global revenue) [29]. The regulation includes many requirements for organizations collecting and processing personal data and these requirements match many of the notice and choice-based privacy principles and frameworks. This law is globally significant due to the importance of the consumer market in the EU (an area with a population of approximately 500 million representing 16% of the world economy [89]) and the application of the law to any organization operating within the EU, thus encompassing nearly all global organizations. This regulation encouraged other governments towards considering similar legal frameworks as well as continuing to significantly shape the market for privacy technologies [86].

To better understand the use of privacy enhancing technologies by organizations, we interviewed a set of privacy experts working across various industries and in multiple disciplines.

We augmented these interviews with a multitude of information from sources such as academic conference proceedings, academic journals, open source projects, blogs, news articles, and other publicly available sources of information, this rich interview data allows us to identify, analyze, and assess the effectiveness and efficiency of these technologies towards their privacy objectives and the overall business goals of the organizations employing them.

4. Background

4.1. What are Privacy Enhancing Technologies?

A simple definition for privacy enhancing technologies (PETs) is “the broader range of technologies that are designed for supporting privacy and data protection” [12]. PETs use technical capabilities to protect and enforce the privacy choices of individuals and groups of individuals. In practice, the term is used broadly. For our purposes, we will work with this general definition of PETs without critically assessing the definition or attempting to define privacy ourselves.

4.2. Types of PETs

To ensure we cover the broad range of PETs, we require a method of categorizing each technology. Prior work provides some taxonomy for PETs. The Office of the Privacy Commissioner of Canada provides one such taxonomy, listed in Table 1 [33].

Category	Examples
Informed Consent	Data tagging with policies
Data Minimization	DuckDuckGo, Disconnect, Private Browsing
Data Tracking	Google Account Dashboard
Anonymity	Tor
Control	Attribute based credentials
Negotiation Terms and Conditions	P3P
Technical Enforcement	Ad-blocking software, transparency tools
Remote Audit of Enforcement	Automated tools for auditing practices
Use of Legal Rights	Tools to exercise rights to data control (opt-outs)

Table 1: A Taxonomy of Privacy Enhancing Technologies [33]

However, lists such as this seem limited to the time of development as well as the purpose of the list. Another approach attempts to first identify the motivation for privacy technologies [16]. This provides a hierarchy of privacy constraints with user preferences and privacy laws as the two top-level categories. For our analysis, we apply a similar hierarchy, seen in Figure 1 where we focus on the source of motivation for privacy technologies. Motivations include individual preferences, legal requirements, industry standards, and self-regulation. This allows us to first divide PETs into those employed by individuals and organizations. For example, browser-based ad-blocking software is included as a technology used by individuals, while de-identification algorithms fall within the domain of organizations. While some technologies overlap, this provides a useful distinction. This distinction exists due to the purpose of our analysis, identifying the technologies of interest to organizations. This purpose requires us to consider both technologies that the organizations could adopt as well as those technologies that an organization's customers, or other third parties, could independently use that would impact the organization.

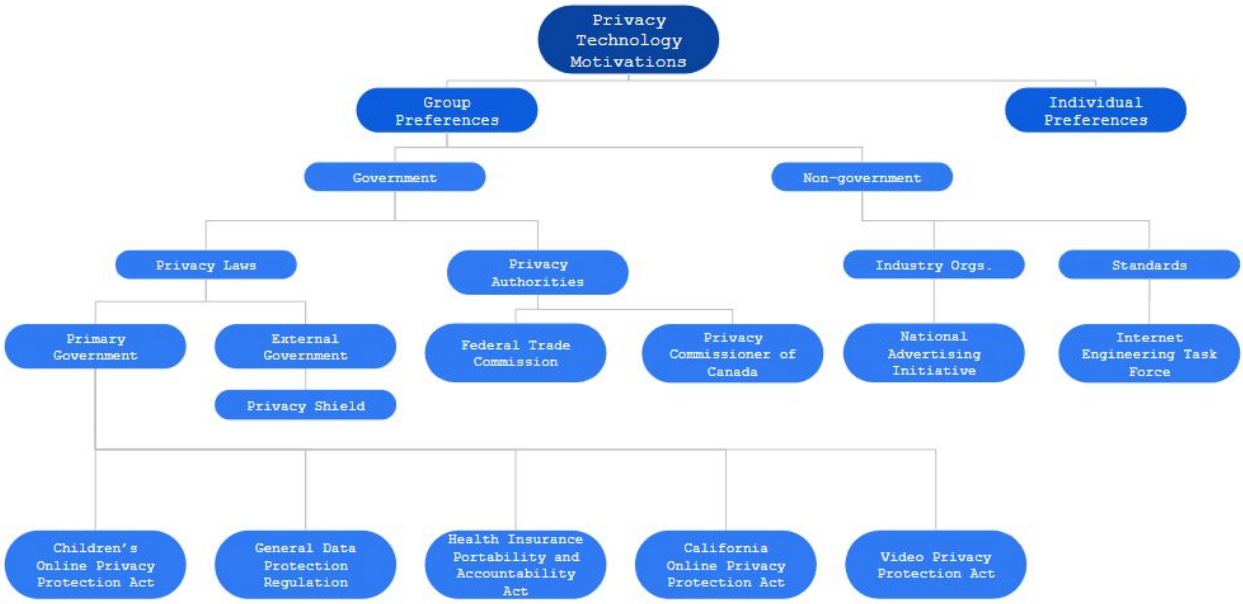


Figure 1: Privacy technology motivations

In addition to the motivation for PETs, similar to Wang et. al [16], we consider the privacy principles or properties provided or augmented by the technology. To build our set of privacy principles, we first look towards existing policies and frameworks such as the Organization for Economic Cooperation and Development's (OECD) *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data* from 1980 [32] and the U.S. Federal Trade Commission's Fair Information Practices [36].

1. **Collection Limitation Principle** - *There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.*

2. **Data Quality Principle** - *Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date*.
3. **Purpose Specification Principle** - *The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfilment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.*
4. **Use Limitation Principle** - *Personal data should not be disclosed, made available or otherwise used for purposes other than those specified in accordance with Paragraph 9 except: a) with the consent of the data subject; or b) by the authority of law.*
5. **Security Safeguards Principle** - *Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorised access, destruction, use, modification or disclosure of data.*
6. **Openness Principle** - *There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller.*
7. **Individual Participation Principle** - *An individual should have the right: a) to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to him; b) to have communicated to him, data relating to him i) within a reasonable time; ii) at a charge, if any, that is not excessive; iii) in a reasonable manner; and iv) in a form that is readily intelligible to him; c) to be given reasons if a request made under subparagraphs (a) and (b) is denied, and to be able to challenge such denial; and d) to challenge data relating to him and, if the challenge is successful to have the data erased, rectified, completed or amended.*
8. **Accountability Principle** - *A data controller should be accountable for complying with measures which give effect to the principles stated above.*

Not surprisingly, many of the PETs we analyzed align with one or more of these principles. By considering the motivation and privacy principles involved with a technology, we can begin to organize these technologies. Once categorized, we use the categorizations to provide assessments of each class of technology (e.g. algorithms for de-identification) as well as a comparative of technologies within each class (e.g. differential privacy).

4.3. Requirements of Privacy Enhancing Technologies

Within these motivations and privacy principles three conceptual objectives related to privacy engineering and technologies emerge: predictability, manageability, and disassociability [76]. Predictability relates to the reliability, transparency, and accountability of a data system to users and owners. Manageability covers the controls and administration of data within a system by both users and owners for privacy-related operations such alteration of data, deletion of data, and appropriate access controls on data. Disassociability encapsulates the objective of limiting the exposure of an identifiable individual during the processing of data. The National Institute of Standards and Technology (NIST) consider these three areas the primary objectives of privacy

engineering which they further define as an engineering discipline aimed at “achieving freedom from conditions that can create problems for individuals with unacceptable consequences that arise from the system as it processes PII” [76]. Privacy engineering and privacy enhancing technologies are complementary areas of study as privacy engineering involves the utilization of various privacy enhancing technologies for the purpose of designing systems to respect privacy principles. Therefore, we use these three objectives of privacy engineering as the overarching requirements or functionality of PETs, also included in Figure 2.

Prior attempts to enumerate PETs highlighted the challenges of this expanding and amorphous space. For example, industry driven reports of privacy technology focus exclusively on products or services provided by software and technology vendors [77]. This type of market analysis succeeds in providing a timely and accurate list of technologies. However, they fail to capture the vast array of internal technologies used to address privacy at large organizations and, more importantly, the impact to privacy of each technology within the overall space. For example, a robust vendor-provided consent management platform provides a simplified and efficient approach for organizations to manage consent requirements. However, for organizations that maintain their own consent management platform or that do not engage directly with end users, this type of technology is not impactful. Another approach towards identifying PETs is to focus on the academic research produced at conferences, journals, and research-focused organizations. However, these sources fail to provide a reasonable measure of the impact of the work in a timely manner, (i.e., we likely cannot know today what recent paper or work will prove extremely valuable years from now).

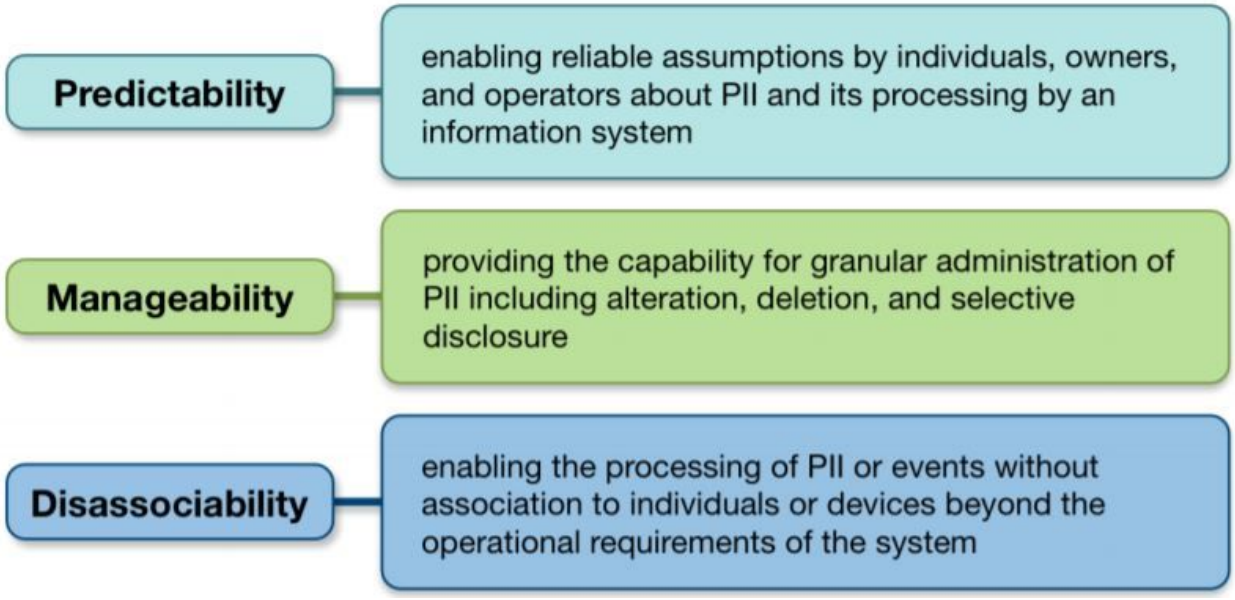


Figure 2: Privacy engineering objectives [76]

4.4. Assessment of PETs

Assessment of any technology presents a wealth of challenges. Fields with a larger history of assessing technology such as health or energy provide some useful insights into evaluation

methods. The U.S. Department of Energy, faced with the task of supporting research into energy efficiency and renewable energy, reported their technology evaluation methodologies to include peer review, case studies, commercialization tracking, surveys, bibliometrics, econometrics, spillover analysis, and benchmarking [78]. Throughout any of these evaluation methodologies, the constraining factor in the quality of the assessment is data. In this regard, privacy technology does not fit well into these models due to the challenges in acquiring the required data. For example, peer review of privacy enhancing technologies through academic journals such as the *Proceedings on Privacy Enhancing Technologies* provides useful data on research, but often says little of a technology's impact across industry. Likewise, commercialization tracking and econometric methods typically require analyzing a technology's performance as a product within a market, which is not the case for many PETs. Often PETs are not products that are sold or do not provide accepted metrics of success and impact. Case study methods seem the most promising for PETs and our methodology reflects this. For further guidance, we look to the cybersecurity field which encounters similar issues and has moved towards an evaluation approach that focuses on risk management [87]. Risk management addresses uncertainty through consideration of the likelihood of adverse events and estimates of impact. This fits well for evaluating privacy technology due to the significant unknown variables throughout privacy such as the cost to an individual of disclosure of personal information. A risk-based approach allows us to address this uncertainty and variability in our analysis. NIST provides a risk-based approach towards analyzing privacy engineering in systems that we will apply towards the PETs in our analysis [76].

To apply this risk model, we must consider the problematic data actions that may occur to the users of the system, as outlined in Figure 3. Primarily, these actions include any "data action that causes an adverse effect, or problem, for individuals" [76]. For many organizations, the adverse effect is an externality that does not impact them, instead individuals bear the majority of the impact. However, we must determine the impact on both the individual user and the ways in which that impact translates to the organization. These impacts must be considered within the context of the practices of the organization handling the data or interacting with the individuals. For example, within the context of processing payment information in an e-commerce transaction an adverse data action would be the user of the associated personal information for additional purposes such as targeted marketing. In this case, the impact to the user is that they are burdened with unwanted exposure to marketing that is perhaps too highly personalized. This impact is then translated by the user into a negative impression of the organization (e.g., annoying or "creepy" advertisements). We model privacy harms within the context of privacy technologies that attempt to mitigate them.

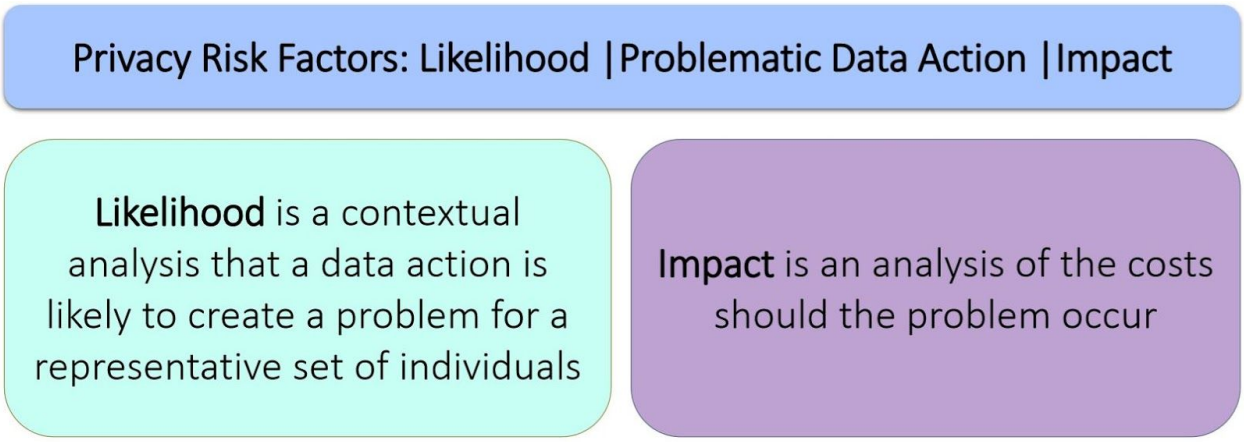


Figure 3: Privacy Risk Factors from NIST report on Privacy Engineering [76]

5. Methodology

5.1. Expert Interviews

To supplement the prior work in this area, we approached the issue by focusing on the privacy experts involved in the field today. This provides the opportunity to gather information on the current technologies across industries as well as the overarching trends within privacy technologies. We conducted 20 interviews of privacy experts from September 2018 to November 2018. Most interviews were conducted over the phone or through video conferencing applications, and typically the interviews lasted 30 to 60 minutes. We did not offer any direct compensation to participants, as we felt professional interest would influence our desired participants more than monetary compensation. Therefore, we ensured our participants the results of the process would be made publicly available once the project is completed. We promised and maintained strict confidentiality of our participants, their responses, and their associated organizations.

We recruited interview participants through contacts within our department (faculty) as well through various privacy profession groups. The participants included engineers, lawyers, policy analysts, chief privacy officers, and a variety of other professions related to privacy within organizations. The organizations represented included a wide array of large and small technology companies, law firms, consulting firms, and other large corporations. Due to the challenges of recruiting participants in a short period of time, our interviews encompass only a fraction, and likely a skewed fraction, of privacy professionals. However, in conducting the interviews, we primarily sought to identify technologies, and did not attempt to provide robust, qualitative analysis of the content or the sample of the population interviewed. This was mostly a fact-finding endeavor.

With this goal in mind, we conducted the interviews in a semi-structured format and included only a small number of open-ended questions which are listed in Appendix I. This format was chosen to avoid focusing the interviews on any one technology and to gather the

most information about those technologies and privacy challenges of significance to the participants. We did provide follow-up questioning on particular topics or technologies throughout the interviews, mostly to ensure our understanding of the topics discussed and to extract any further related technologies or topics of interest.

5.2. Publicly Available Sources

In addition to expert interviews, we conducted extensive research of privacy technologies across a variety of sources including academic research, industry reports, open source technologies, technology and organizational blogs, news articles, and other similar content available across the Web. For vendor technologies, we focused on industry reports such as those created professional organizations (e.g. International Association of Privacy Professionals). To cover research, we focused on extracting privacy-related research from specific academic journals (e.g. Proceedings on Privacy Enhancing Technologies), public-private research collaborations (e.g. Microsoft Research), and any further work the researchers from these sources contributed towards (open-source projects, personal blogs, etc.). The interview participants provided information on their experiences with different technologies which we used to drive further research into these sources. Additionally, interview participants provided specific resources and suggestions that helped to drive this aspect of the project.

5.3. Components of Analysis

To assess each privacy enhancing technology, we focus primarily on analyzing the effectiveness and efficiency as highlighted in Figure 4. For effectiveness, we aim to determine how well the technology reduces the targeted privacy concerns by supporting any of the three objectives of privacy engineering previously discussed (predictability, manageability, and disassociability). For example, de-identification techniques can reduce the identifiability of a data subject within a dataset, and their effectiveness is based on the implementation of the de-identification (e.g. *k*-anonymity, *l*-diversity, *t*-closeness, tokenization). This area would clearly fall under the scope of disassociability. Furthermore, our measures of effectiveness consider the risk-based methods previously described. Within this example, we would consider the reduction in risks due to de-identification of personal data.

The efficiency of any technology is largely determined by both the costs of implementation and the reduction, if any, in utility of the related practices to the organization. While de-identification better preserves the privacy of an individual within a dataset, it may reduce the utility that an organization can extract from the dataset. Additionally, the complexity and counterintuitiveness of de-identified data may incur indirect costs on the organization such as training data analysts to understand the implications of de-identification on their analysis processes. Finding optimal technology solutions to privacy challenges requires consideration and balancing of these metrics, as well as planning for the objectives and future path of the overall organization.

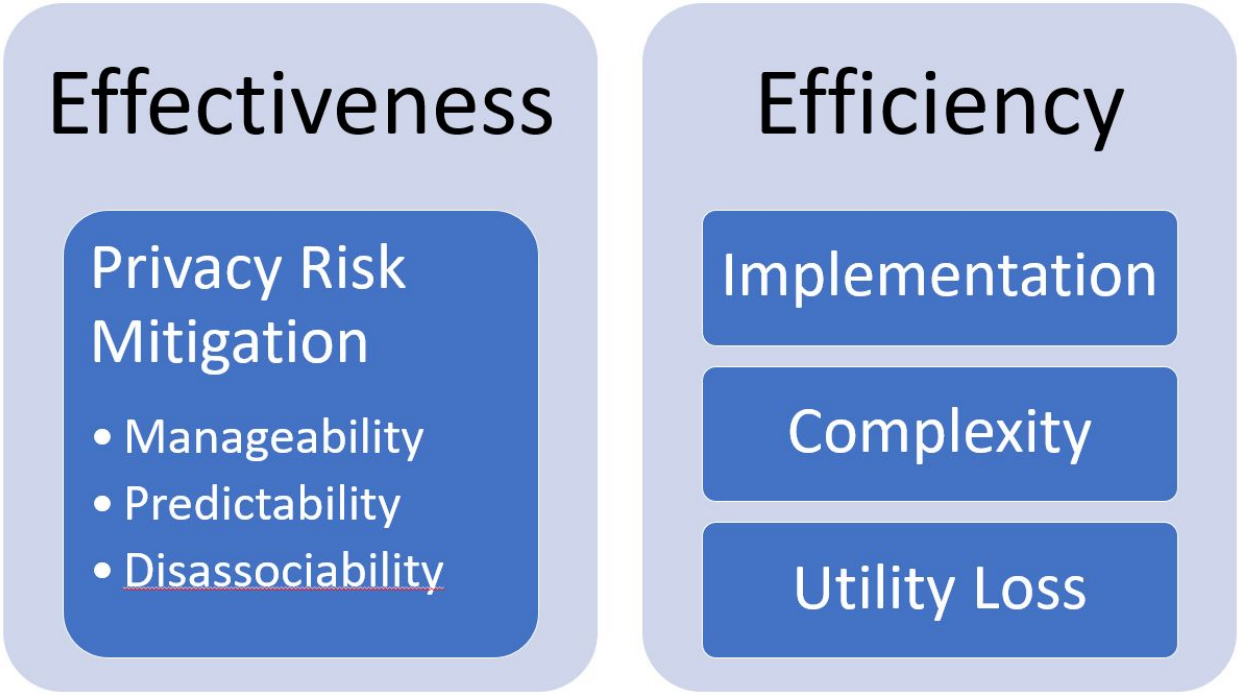


Figure 4: Evaluation factors for privacy enhancing technologies

For each class of technology (e.g. de-identification algorithms), we highlight the privacy risk factors involved as each technology is designed and applied towards reducing privacy risks. In this respect, we identify the adverse actions, the impact of those actions on both individuals and the organization, and the likelihood of those actions creating an adverse outcome. This provides an overall assessment of the value of the technology towards the privacy goals of an organization. Additionally, to provide a practical aspect to our analysis, we include details of how the technologies relate to current regulations such as GDPR. This assists with connecting the technological functionality to the motivation for adoption of technologies by use of actual legal requirements. From the use-cases provided by the source material, we extract the limitations, challenges of implementation, and any reductions in utility caused by the technology. This informs our analysis on the efficiency of each class of technology.

Within each class of technology, we consider the factors that differentiate each instance of the technology (e.g. differential privacy at Apple, Google, Uber). This provides a comparative analysis, highlighting the costs and benefits of each instance of a technology in comparison to similar technologies.

6. Results

6.1. Privacy Preserving Approaches to Data Collection

Techniques to preserve individual privacy within large datasets represent a growing area of technology advancement. These technologies provide privacy properties to the individuals whose data is included in large datasets maintained by organizations. Primarily, these

technologies attempt to reduce the risks individuals assume by their inclusion in these datasets such as from harmful disclosures or costly uses against their best interests. While unintended disclosures from security failures are a persistent risk, they remain outside of the scope of our analysis. Abuse or misuse of an individual's data represents an adverse action that some privacy technologies attempt to directly reduce. Examples of these privacy harms are provided in Figure 5.

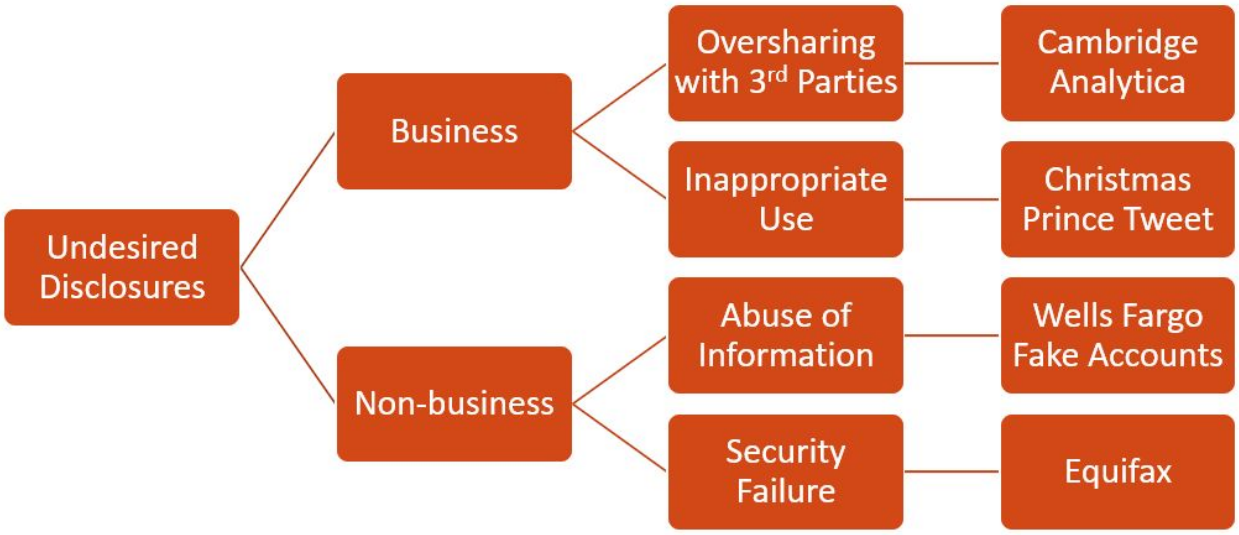


Figure 5: Examples of disclosure-based negative outcomes involving large datasets [85, 90, 91, 92]

One approach to limit the likelihood of a negative outcomes for individuals is to remove the links between the individual and their associated data. If the data cannot reasonably be attributed to an individual, then the individual is protected from many adverse outcomes from disclosure of the dataset. Reducing identifiability attempts to remove the possibility for direct consequences to the individual that may derive from their inclusion in a dataset. Overall, this concept fits perfectly into the objective of disassociability.

This approach aligns well with the GDPR section on pseudonymization. GDPR defines "personal data" as "information relating to an identified or identifiable natural person data subject" [29]. For processing personal data, the GDPR defines "pseudonymization" as transforming personal data to a form that is neither anonymous nor directly identifying [29]. This transformation protects the individuals while also freeing the data controller to safely process the data for their intended purposes. For example, in case of a data breach, pseudonymization reduces the likelihood of identifying data subjects from a leaked dataset. Therefore, GDPR encourages pseudonymization of data to address the risks of potentially revealing identifiable individuals within datasets [29]. There are many techniques that are helpful to pseudonymize, some of which are included in here.

6.1.1. Direct Identifiers

Reducing the identifiability of a dataset requires first removing any direct identifiers (e.g. name, social security number, email address, IP address, etc.). Technology can solve this by

applying algorithms to locate and redact, mask, or tokenize these identifiers within a dataset. These approaches provide varying levels of reduction in identifiability, and can be implemented to align with the organization's desire for potential re-identification. Google Data Loss Prevention (DLP) API provides a robust set of operations for providing these types of reduction in identifiability, as seen in Table 2. In general, these techniques involve removing data or reducing the precision of data while still maintaining some key properties (equivalence, ordering, etc.) of the dataset. Additionally, with the possibility of re-identification, additional use-cases for the data remain possible. For example, an organization tracking user behavior across devices and locations for the purpose of identifying anomalous activity indicative of a compromised account requires re-identification of the individual's data to apply remedial action. In cases such as this, tokenization facilitates a limited and controlled process for handling this need for re-identification.

Google's DLP technology effectively addresses privacy risks from direct identifiers, given the data resides within Google's Cloud. This however remains a barrier or cost for efficient use of the technology. While the technology may be configure to function beyond Google's cloud, the value is maximized by the organization including all of their data on Google's cloud.

Technique	Description	Example Before	Example After
Date Shifting	Maintain the chronological order of records while removing precise dates	2009-06-09, 2009-06-11	2009-06-01, 2009-06-02
Bucketing	Combining precise values into general buckets	Senior engineer, junior engineer, principal engineer	Engineer
Redaction	Removal of sensitive data	"SSN is 123-456-7890"	"SSN is ___"
Tokenization	Sensitive data replaced with tokens in a consistent manner, allows for re-identification	Employee ID = 1	Employee ID = a3Rf

Table 2: Example of de-identification algorithms on Google Cloud's DLP product

6.1.2. Indirect Identifiers

A further challenge for reducing identifiability involves addressing the indirect identifiers present within a dataset. An indirect identifier uses the uniqueness of a individuals data to distinguish them from other individuals within the dataset. These unique properties then allow for the identification of individuals within external datasets such as through intersection attacks.

The uncertainty of identifiability of indirect identifiers present a form of moving target for system engineers. For example, 87% of Americans can be uniquely identified by only their zip code, gender, and date of birth [42]. This variance within the identifiability of data points allows for unique problems based on both the data types and the populations included in the dataset. The previously mentioned Google DLP technology provides a risk analysis for data sets based on identifiability within the dataset using k-anonymity, l-diversity, k-map, and delta-presence. These aspects of the technology are explained in Table 3.

Algorithm	Description
k-anonymity	The essential arrangement of *k*-anonymity is to shield a dataset against re-identified by summing up the characteristics that may be used in a linkage attacks (semi identifiers). A dataset is considered *k*-anonymous if each data thing can't be recognized from at least *k-1* elective information things [74].
l-diversity	The l-diversity model is an extension of the k-anonymity model which reduces the granularity of data representation using techniques including generalization and suppression such that any given record maps onto at least *k-1* other records in the data.
k-map	In DLP, a *k*-map estimate is computed through a statistical model to estimate re-identification using auxiliary datasets.
delta-presence	*Delta*-presence means that a recipient of an anonymized database should not be able to identify any individual as being in that database with certainty greater than δ [73].

Table 3: Google DLP algorithms [74]

Another approach towards this set of risks is differential privacy. Differential privacy "addresses the paradox of learning nothing about an individual while learning useful information about a population" [11]. A number of organizations have recently implemented differential privacy-based technologies into their data collection and analysis pipeline [8]. These algorithms are designed to meet the requirement that individuals "will not be affected, adversely or otherwise, by allowing data to be used in any study or analysis, no matter what other studies, data sets, or information sources, are available" [11]. In many cases, the differential privacy guarantee is accomplished by adding random noise to the various data points returned from queries or individuals. This noise reduces the information available about an individual within a

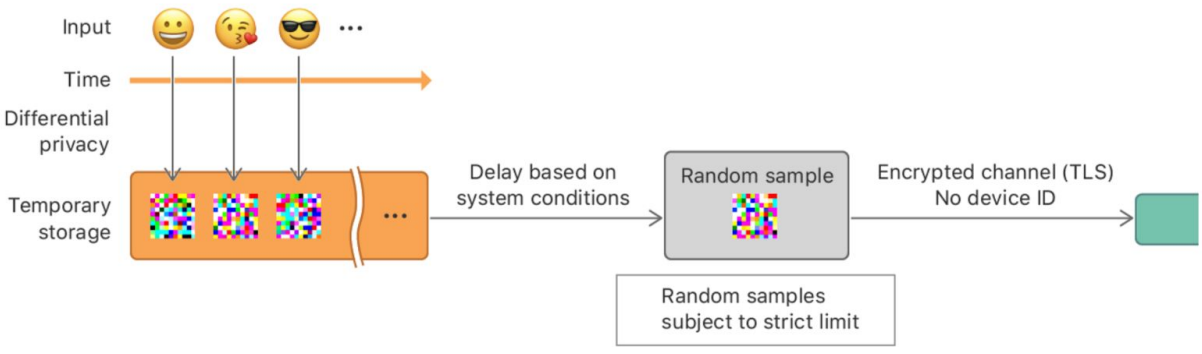
dataset by addressing the hypothetical case of an attacker comparing adjacent databases for differences based on the exclusion of a single individual from one of the databases.

6.1.3. Differential Privacy for Telemetry

Local differential privacy (LDP) algorithms have recently emerged as a tool that allows data collectors to estimate various population statics while preserving privacy. Microsoft developed LDP mechanisms geared towards repeated collection of counter data (e.g. number of times an application is run by a user), with formal privacy guarantees even after being measured over an arbitrarily long period of time [88]. This technology is currently used in Windows 10 to collect users' telemetry data such as daily application or system usage statistics [88].

6.1.3.1. Apple Differential Privacy

Apple uses differential privacy for aggregation of a variety of data points from their user's devices such as typing suggestions, emoji suggestions, and Safari crashing domains [71]. The technique adopted by Apple is a form of local differential privacy, similar to Microsoft's implementation. This design allows Apple to transform the information of individual users before it leaves user's device. First, the information is privatized by adding some amount of noise to the data and removing any direct identifiers such as IP addresses. Then, aggregation occurs when the privatized records are shared with Apple's centralized database and made accessible to the various consumers of the data (e.g. Apple product teams working with these metrics). Figure 6 provides a visualization of the steps as well as a sample of a privatized record. Within this implementation, Apple provided new methods for applying differential privacy to frequency estimators for dictionaries of known and unknown values. For example, identifying common, new, non-dictionary words typed by users [71]. This represents an important design challenge for differential privacy implementations, the variations in algorithms necessary for various data types. For example, location-based data remains a seemingly unsolved challenge for differential privacy.



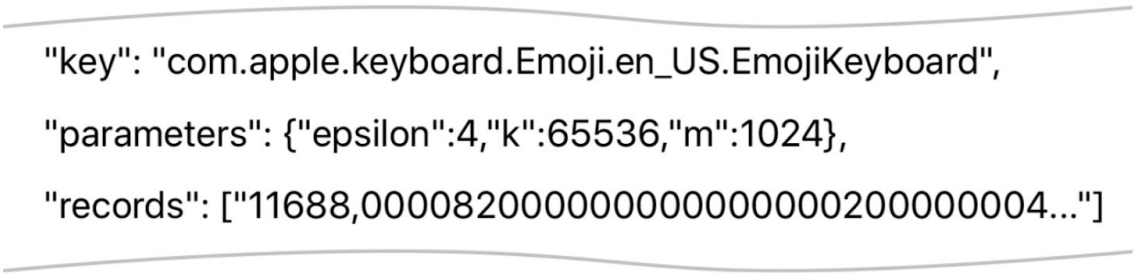


Figure 6: Visualizations of Apple's differential privacy implementation [71].

As with all implementations of differential privacy, the parameters chosen greatly impact the privacy guarantees of the system and provide a tradeoff with utility. Due to the localized nature of this implementation, Apple further parameterized the system to consider both device bandwidth and server computational costs. This full space of parameters, utility, privacy, server computation, and device bandwidth, must be balanced in the deployment of the technology. In Apple's case, the utility is clearly proven by their stated results from these systems as they succeed in identifying popular and trending new words across various user languages as well as identifying popular emojis [70].

6.1.3.2. Google Chrome RAPPOR

Google implemented differential privacy throughout the telemetry data collection system used within the Google Chrome browser. This technology, Randomized Aggregatable Privacy-Preserving Ordinal Response, or RAPPOR, allows for crowdsourcing statistics from end-user client software, anonymously, with strong privacy guarantees [67]. This implementation also is a form of local differential privacy and further addresses the challenge of privacy degrading over multiple observations of the user's private information. RAPPOR provides more provably-strong privacy guarantees while eliminate the privacy externalities from individual data collections.

Chrome is an open-source project and the implementation and use of this differential privacy mechanism can be easily observed. This proves very valuable for analysis, since most other implementations of differential privacy include only an academic paper, blog, or maybe some code samples of the privacy relevant portions of the system. The fully open-source nature of Chrome allows us to extract both the implementation of differential privacy as well as the integration of it into the overall software. From the source code, we can see Chrome uses differential privacy for telemetry data points such as user permission choices, web page load times, process performance measures, and many other browser-specific metrics.

Again the tradeoff between privacy guarantees and utility extracted from the data is highlighted in Google's results. Figure 7 highlights the number of learnable values from a population of the indicated size. In some cases, this may be impactful, however the use-case highlighted show that these limitations are acceptable for a large number of cases. For example, when observing the most common homepage setting (the first page that opens when a user runs the Chrome browser) for Chrome users with a sample size of approximately 14 million, most users have one of approximately 31 web pages set as their homepage. With the differential privacy implementation used by the authors, only 8,616 total different values could

possibly have been observed,. However, due to the distribution's concentration within those 31 web pages made this limitation acceptable. This represents a challenge for most differential privacy implementations, the value of the data is greatly impacted by the application and the overall size of the sample available.

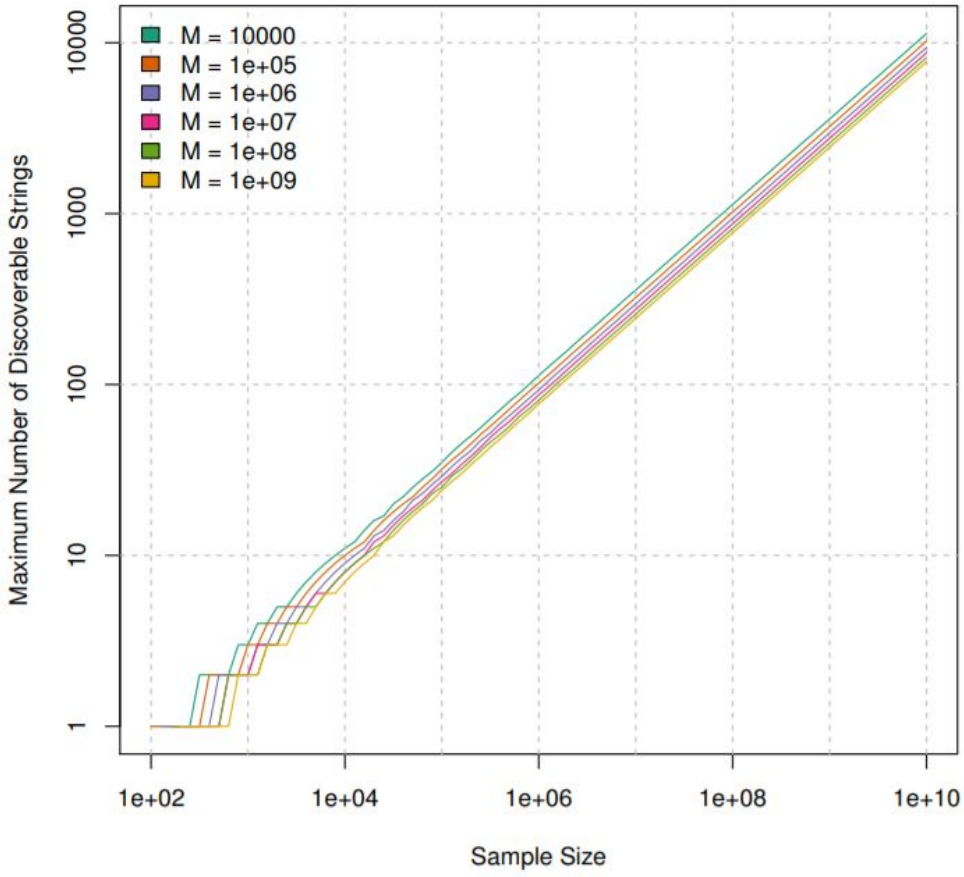


Figure 7: Impact of sample size on RAPPOR results [67]

6.1.3.3. Uber Elastic Sensitivity

Uber developed a query analysis and rewriting framework to enforce differential privacy for general-purpose database query languages (e.g. SQL, structured query language). This framework transforms database queries to allow for preserving privacy. The transformed query enforces differential privacy on its results. The framework is built on an approach called *elastic sensitivity*, a method for approximating the local sensitivity of queries that include the joining of data to related data [68]. SQL is an extremely common query language, and this technology allows for adding differential privacy constraints to SQL-accessible datasets in a generic approach. Figure 8, from this work, highlights one of the overarching considerations with differential privacy, the reduced utility extracted from the dataset. Of significant importance in this use-case is that the utility of the results decreases based on both the complexity of the query (based on joins) and the size of the resulting dataset or volume of data included. In this case, the utility is measured by the error rate observed within the results due to the added noise.

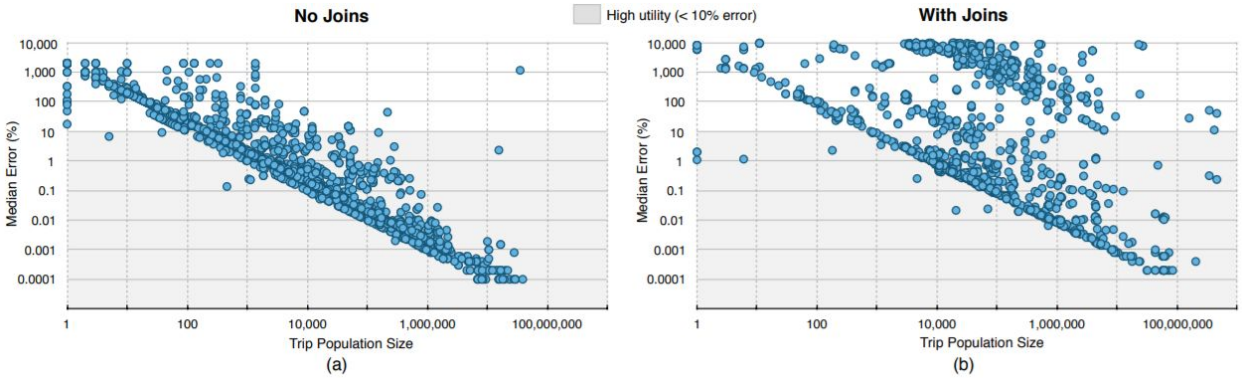


Figure 8: Elastic sensitivity utility variations by results set size and joins [68]

6.1.4. Differential Privacy Summary

Differential privacy implementations currently face several barriers to wider adoption. Primarily, most of the existing implementations take full advantage of large scale user systems. This allows organizations to reduce the information learned from each individual to a very small value while still extracting valuable information from their large population of users. Similar approaches with smaller datasets can have a significant reduction in value extracted from the data,as seen in the elastic sensitivity case. This represents a technical challenge that advancing research will continue to address. However, the strongest support for differential privacy is that these decisions become quantifiable and parameterized. Parameterization of privacy allows organizations to tune their policies based on appetite for risk and the balance of privacy and utility. A further challenge identified in these case studies is the open issue of applying these methods to various data types as seen in Apple's implementation addressing frequency estimations for known and unknown dictionaries. However, approaches for preserving privacy in various data types remain incremental with specific techniques developed for different data representations (e.g. sets, dictionaries, location data, graph data).

Furthermore, the cost of implementing differential privacy may be significant. As there are limited public resources available, most of the implementations included our analysis came from academic partnerships with private industry or academic experts hired by large technologies companies. This approach is unlikely infeasible for most organizations. Local differential privacy approaches, like those at Microsoft, Google, and Apple, also cannot be applied to data already collected or many types of existing data collection systems.

6.2. Privacy Preserving Approaches to Data Use

A key aspect of preserving privacy in data analysis is balancing the needs of the organization or the value of the data with the privacy constraints. In this area, we see several advancements in technology intended to preserve privacy while allowing organizations to extract utility from the data. Specifically, we focus on machine learning-based uses of datasets and data sharing. These represent two complex and fundamental aspects of data usage today.

6.2.1. Federated Learning

Federated learning is an alternative approach to training machine learning models. The initial use-case for federated learning is Google's Gboard, the keyboard used across Android devices [69]. The typical workflow for training a machine learning model involves collecting a large training dataset, labelling that training set with the desired parameter, training a model based on that dataset, and applying the model to future datasets to arrive at a prediction for the labelled parameter, Figure 9 provides an example from GBoard.

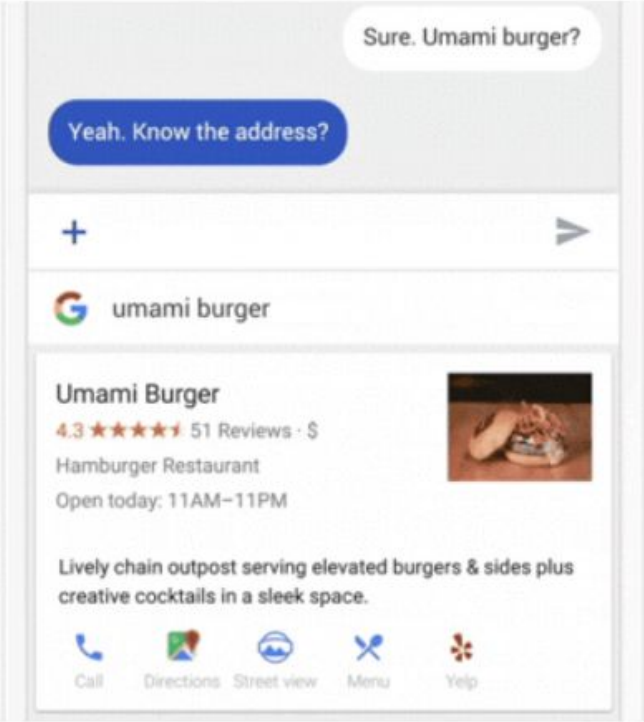


Figure 9: An example suggestion within Google's GBoard. The user typed in "Umami Burger" and the model provided suggestions based on the user's query. In this case, the user's choice will be used in the training process [69].

When Gboard shows a suggestion, the application locally stores information about the current context and the user's response to the suggestion [69]. With federated learning, the device receives a training model, updates the model based on the user's behavior, and then propagates only the updated model to a centralized server as shown in Figure 10. This process takes advantage of a federated averaging algorithm which balances the processing resource constraints (bandwidth and computation) involved to minimise the impact of the operations on user experience. For a subset of use-cases, federated learning provides a powerful approach towards balancing privacy while maintaining valuable utility of data. It enables client devices, such as mobile phones, to collaboratively learn a shared prediction model while avoiding a large, central database of user data [69]. This enhances privacy in a very effective way by eliminating the risks of large, central datasets as well as limiting the information extracted on each user to only those details required by the machine learning model. One limiting factor in

this technology is the generalizability to other machine learning applications. With this model, the data remains indirectly labeled, as it is dependent on user actions. Many applications of machine learning cannot fit this model where the user essentially provides the labeling of the training dataset.

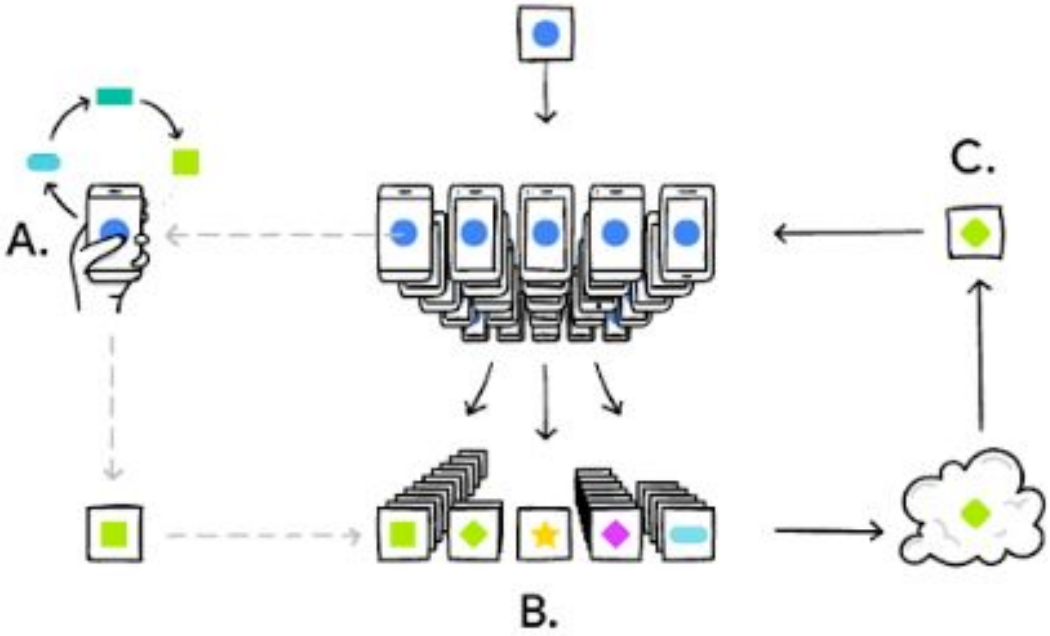


Figure 10. Process of federated learning [69]

6.2.2. PATE

For general machine learning algorithm, there is no guarantee that the output model generalizes away the specifics of any individual user. Several proposed attacks exploit this implicit memorization in machine learning to demonstrate that private, sensitive training data can be recovered from models. To address this problem, Private Aggregation of Teacher Ensembles (PATE) introduces a new strategy that involves a teacher-student model which effectively hides the details of sensitive data [72].

Figure 11 shows the process of PATE. In this strategy, first, an ensemble of teacher models is trained on different, non-overlapping sets of labeled sensitive data. Then a student model is trained on the aggregate output of the teacher models. However the student model only learns a specific result if it can be derived from multiple teachers, and includes a noise element to provide greater privacy guarantees. This strategy ensures that the students does not depend on the details of any single sensitive training data point. This reduces the risk of an individual's private information being included in a trained model [66]. Compared to general machine learning algorithms, PATE sacrifices some accuracy, achieving around 90% accuracy of the original algorithm, to ensure the strong privacy guarantee.

This technology is effective at addressing privacy risks for individuals related to the inferences learned from large datasets. However, the likelihood of this kind of risk is difficult to assess, and therefore limits the potential impact of this technology. Considering the costs to

accuracy and complexity of implementing this technology in an existing machine learning pipeline, it likely is very difficult to apply this beyond the narrow cases described by the authors of the work.

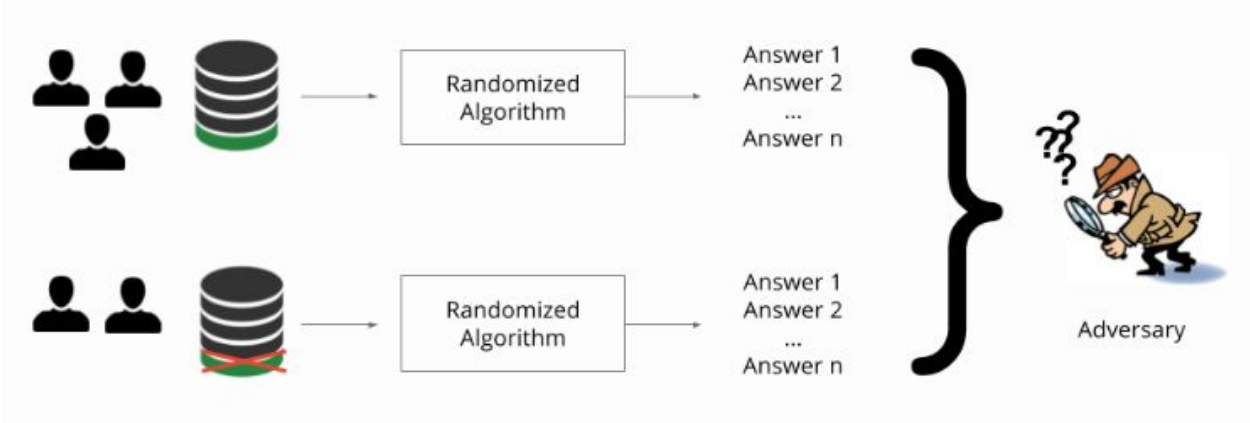


Figure 11. The process of PATE [66]

6.2.3. Private Set Intersection

Private set intersection is an emerging area of technology in the context of sharing information between owners of large dataset. In this technology, two parties wish to share the overlapping data points in their datasets without revealing any other information from their dataset. One practical example is the case of private contact discovery with messaging applications such as Signal [75]. Private contact discovery attempts to answer the question “how do we determine which contacts are registered with a service, without revealing the contacts to the service?” [75]. This arises when an individual joins a new service and wishes to discover which of their existing contacts already have joined the service. The non-private approach simply involves asking the service which user already has an account, based on a common identifier such as an email address or phone number. This process leaks the individual’s contacts to the service. This privacy challenge appears across a large variety of use-cases. Google and Mastercard recently reached an agreement to employ one form of technology that allows for private set intersection for the purposes of attribution of offline purchases to digital ads [79]. In this scenario, Google compares the individuals shown ads on their network against Mastercard’s list of users completing transactions offline at retail stores. If one of Google’s ad viewers appear in Mastercard’s list and the advertisement shown to the user matches based on time and content to the location of the purchase, then in some cases Google can make some assumptions about attribution.

More details on research in this area can be found within the appendix.

6.2.4. Privacy Preserving Data Usage Summary

Overall, we see several technologies focused on reducing the information collected from individuals while still maintaining reasonable levels of information extraction from large

populations. Use-cases such as machine learning can likely adapt to privacy preserving mechanism such as federated learning, and privacy challenges specific to learning-based models can be addressed with techniques such as PATE.

Private set intersection represents the rare case where a privacy-preserving technology can directly enable business goals such as as data sharing. Significant research and maturity for these technologies will be needed before widespread adoption can occur.

6.3. Safeguards on Data Practices

Another area of technological solutions revolves around the safeguarding of practicing to ensure privacy requirements are met and enforced. These types of technologies overlap significantly with the security realm, and thus are much more common and mature. We list some of these technologies in Figure 12. Increased adoption of these technologies seems partially in response to regulations. GDPR created new requirements for handling security such as requirements for breach notifications [50]. These requirements force organizations to formalize some of their security procedures especially when dealing with responding to security incidents or data breaches. Security is a requirement for an organization to meet their privacy objectives. Furthermore, Article 32 of GDPR specifies data controllers and processors must “implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk” [50]. This section mandates appropriate security measures as a requirement to meeting the privacy regulation. While security requirements significantly impacts organizations, for the purposes of this report we will not focus on purely information security technologies as PETs, only those that augment privacy in some unique and significant form. This limitation keeps the scope of the project with reason, as security technology is far too large a field to consider in addition to privacy technology.

Figure 12: Technologies to safeguard privacy practices

Each of these technologies allows for improved management of privacy requirements throughout organizations. Access controls within systems provide a limitation on which employees throughout the organization can access what data and under specific circumstances. The audit logs supports this functionality by allow organizations and third parties, such as customers, to review accesses to data for appropriateness. Data discovery further expands the data practices to identify all datasets under the purview of the organization. A common struggle for large decentralized organizations is identifying the data relevant to privacy constraints. Confidentiality tools and techniques, primarily various forms of encryption (at-rest, in transit, key management, etc.) assist with enforcing these adjacent technologies by rendering the data useless without the proper authorization. Similarly, platforms for notice and consent ensure the privacy practices of the organization are properly available to users and that the consent for those data practices is gathered when necessary. These technologies represent some of the most common across our interviews, however they are defined more by the properties they provide than any one product, service, or vendor.

6.4. Accommodating Users

6.4.1. User Access and Control

GDPR stipulates many user access requirements for data controllers. This allows individuals to request the deletion of personal data, and, in cases where the controller has publicized the data, to require other controllers to also comply with the request. Additionally, data access and portability rights require controllers to provide personal data to the data subject in a commonly used format and to transfer that data to another controller [28]. Some existing rights, like the

data subject's right to object to processing, also expanded. Openness and data quality are also addressed. Data subjects are entitled to receive notice about processing activities, gain access to the information involved in processing, and force the controller to resolve inaccuracies. For example, data controllers must provide the functionality to remove user data when "a data subject has withdrawn his or her consent" [50]. To adhere to these requirements, software vendors developed platforms for managing user consent agreements and enforcing rules related to user opt-outs such as in the case of email marketing [5]. The purpose of these software technologies is to automate and enforce the workflows required by regulations.

The Data Transfer Project represents another technology approach towards addressing user requests for their data [2]. This project, supported by companies such as Facebook, Google, Microsoft, and Twitter, aims to provide data portability to users across web services. This project is similarly motivated by empowering users to "be in control of their data on the web, part of this is the ability to move their data" [2].

6.4.2. Informed Consent

Informed consent remains a challenging intersection of technology and users. Notice and choice-based regulations require consent agreements between individuals and controllers. For example, GDPR requires the data subject to signal agreement by "a statement or a clear affirmative action". [27] Generally, GDPR changes the consent requirement from three perspectives:

1. Individuals must consent through a "clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement" [50]
2. Explicit consent required for special categories of personal data
3. Parental consent required for processing children's personal data

Under GDPR, harsh penalties exist for violations of the consent requirements. Therefore, using PETs to prevent violations of consent is greatly beneficial. Organizations are adopting tools to help to gather consent from data subjects on a contextual basis, considering the features of the system that users are accessing while maintaining records for future auditing by regulators. These types of tools allow companies to manage the consent process as they would any other business process [6]. Often, these are vendor provided platforms and technologies and they are typically implemented by large, diverse organizations whose business lies outside of technology.

A related area of academic research appears in the form of privacy policies used by organizations to provide information on their data practices. Privacy policies provide information on the data practices of an organization. These policies are used by regulators, such as the FTC, to hold organizations accountable [35]. Therefore, organizations have two primary challenges in regards to privacy policies. First, they must ensure that their practices comply with their stated privacy policy. Second, to allow for informed consent, they must make the policies accessible to their data subjects. Currently, most privacy policies are written in natural language or, in some special cases such as the financial industry, in a shorter format [22]. One aspect of this challenge is designing a privacy policy language for expressing this information. In the past,

access control languages and frameworks have been considered as potential solutions such XACML [1] and similar technology has been applied to privacy such as P3P [51]. Sen et al. designed Legalease as a “usable, expressive, and enforceable privacy policy language” [65]. Their work further applied this language to the privacy policies of Google and Bing, successfully capturing the constraints of the policy on data practices. While this work allows for organizations and regulators to validate that data practices comply with a stated policy, a further challenge arises in expressing these practices to the users or data subjects. Researchers investigated standardized privacy notice formats and found them to improve “accuracy and speed of information finding” [24]. Technology provides some opportunities for improving notices by expanding the design space from a simple document to include more informative mechanisms. Shaub et al. provide an explanation of this design space and the use-cases of the various mechanisms (timing, channel, modality, and control) available when integrating a privacy notice with technology such as a mobile application [39]. Alternatively, Harkous et al. expanded on existing research to provide Polisis, a deep learning based automated framework for querying nature language privacy policies [17]. This research built on prior work where natural language privacy policies were extracted from across the web [46] and analyzed using crowdsourcing [47]. Automated policy analysis technology can provide enhanced auditing technologies for regulators “to automatically, and continuously perform such checks at scale” and improve privacy relevant information available to consumers [17]. With techniques for machines to interpret privacy policies, technology can further apply user decisions across multiple services with different policies, easing the policy related work of the user. This concept appears in work related to personalized privacy assistants [26], which have been successful in both learning a user’s privacy preferences and apply those preferences to a broad range of applications and services.

7. Discussion

Our analysis included a broad set of privacy technologies that address a wide range of requirements. An important aspect when considering these technologies is understanding the overall process and requirements that they contribute towards. In our case, this includes the privacy requirements of the organization. These requirements include legal requirements, customer requirements, industry standards, best practices, and in some cases a desire for moral responsibility.

7.1. When Disassociability Makes Sense

Disassociability, or reducing identifiability, is a valuable technique for reducing privacy risks, both in likelihood and impact. These technologies assume that a large dataset must be maintained to extract value for the organization. Deterministic approaches, such as tokenization, redaction, and blurring, provide a high level of guaranteed reduction in risk. Once all direct identifiers are removed or pseudonymized, a large set of threats to the data are reduced in likelihood. In many cases, this prevents the unsophisticated risks (e.g. an employee looking up

their neighbor in a dataset). These approaches can be added throughout a data pipeline, and reduce the exposure of direct identifying data to a more controlled dataset (such as the secure storage of tokenization or re-identifying material). Most of these technologies are both mature and widely available. Additionally, this converts the challenges of working with personally identifiable information to an access control problem. Therefore, the only utility loss is a minor increase in costs to re-identify individuals in cases where that is a business requirement. Many organizations already employ technologies to address direct identifiers.

Once direct identifiers are removed the risk of re-identification becomes clear. This is the risk of circumvention of the removal of direct identifiers. Again, these technologies assume that the data must be persisted in a database. Differential privacy, as exemplified in previous sections, provides a statistical method of reducing the risks of re-identification, primarily through the addition of noise to the data for analysis.The effectiveness of this approach is parameterized and therefore configurable on a per-case or even per-query basis. As discussed the complexity of the implementation is dependent on whether a local or centralized model is applied. Additional complexity arises from the lack of standardized approaches or tools, despite the presence of several open-source implementations. Furthermore, the current implementations require different approaches towards handling different types of data. The utility of the data is likely reduced, however the parameterized nature of these methods allows the system designers to ensure the business requirements of the system are still met despite the data transformations. As can be seen by the use-cases, this technology seems to fit well for monitoring population statistics or aggregate data such as in the case of telemetry datasets. In these use-cases, individual or small sets of data points are of less interest than the larger segments of the population. Therefore, the differential privacy implementations are configured to reflect these usages and provide enhanced privacy for the individuals.

7.2. A Predictable and Manageable Architecture Enforces Privacy Practices

An alternative to collecting large amounts of user data is to extract the utility from the data without building a central database. In federated learning, the system is designed to extract the utility from the data, training a machine learning model, without collecting the training data into a central database. This approach effectively limits the risks to the user to only the inferences learned and included in the model. This mitigates most future risks from data breaches, re-identification attacks, or uses that do not adhere to organizational or legal policies. The user's data is only used for exactly the purposes built into the system itself. This model is as efficient as it requires an expensive implementation and future machine learning algorithms may not be compatible or prohibitively costly to implement in a federated manner. However, for mature use-cases that experience limited volatility in results or techniques, this technology allows for maintaining the full utility of the data to both the user and the organization without exposing the user to unnecessary risks.

Technologies that safeguard data practices provide a baseline for addressing privacy concerns. Primarily, the data practices of the organization must be enforced throughout the data

systems of the organization. This requires applying access controls, audit logs, consent management, data mapping, and a wide variety of tools that support this goal. Each organizations data and technology pipeline will dictate the necessary technologies. Third party platforms require organizations to address data practices in contracts and vendor agreements, technology is limited in addressing these challenges.

7.3. Technology Necessary, Not Sufficient

PETs alone do not solve the privacy challenges organizations face today. Privacy will likely never be the primary objective of an organization, however it does require the attention of overall business to be successful. A robust combination technologies can enable privacy preserving data practices to eliminate a wide array of privacy risks. For example, access controls, audit logs, and proper data confidentiality constraints can avoid some of the most common and detrimental forms of privacy harms such as unauthorized disclosures. As these overlap with many security requirements, we see a widespread, but not universal or comprehensive, adoption of these technologies. For a determined organization, most of these technologies are used along significant business processes and privacy management teams to ensure adherence. Today's organizations with robust approaches towards privacy appear successful in meeting their stated practices and the requirements of the law. Some proceed further than is necessary in addressing privacy concerns.

Determining an appropriate or reasonable use of data remains an open issue. Fortunately, a significant number of organizations, public and private, are beginning to consider this issue within the context of ethics and artificial intelligence. A significant portion of modern artificial intelligence is machine learning and one of the key components of machine learning is data. Therefore, the use of artificial intelligence and data are naturally intertwined. We see ethics of artificial intelligence appearing throughout industry groups and individual organizations [93].

8. Conclusion

As previously discussed, many methods of evaluating technologies exist. In our analysis, we focused on interviews as our primary source of information, and further supplemented this with publicly available sources of information. By comparison, the International Association of Privacy Professionals (IAPP) conducted a survey of privacy technology vendors to provide a summary of the technologies available in the market [77]. Both of these methods succeed in identifying different, and in some cases overlapping, sets of technologies. After identifying the technologies in use, the natural next step is to assess them. We provided a comparative assessment of a variety of technologies and the implementations we encountered. Success of each technology is dependent on both the product and the organization. For example, large technology companies use differential privacy successfully for telemetry data and population statistics. However, many of these technologies remain largely academic, and not widely available. This is no failure of the researchers, engineers, or practitioners, but more of the lack

of maturity in the market for these technologies. More mature PETs exist and most of these directly address privacy challenges that are highlighted by regulations such as GDPR. Therefore, organizations must adopt the mature technologies to assist with addressing compliance requirements, and structure their organization technologies (products and services relying on data collection and processing) for future adoption of emerging PETs.

Privacy Engineering Objective	Challenge	Technologies	Use-case / examples
Disassociability	Privacy Preserving Data Analysis	Discovery, redaction, tokenization, generalization	Google DLP API, Amazon Macie, Microsoft Office 365
Disassociability	Privacy Preserving Data Analysis	Local Differential Privacy	Google Chrome, Apple
Disassociability	Privacy Preserving Data Analysis	Central Differential Privacy	Elastic sensitivity
Disassociability	Privacy Preserving Data Analysis	Federated Learning	Google GBoard
Disassociability	Privacy Preserving Data Analysis	Learning model privacy	PATE
Disassociability, manageability	Secure Computation	Private set intersection, homomorphic encryption	Online-to-offline ad conversions, private contact discovery
Predictability, Manageability	User Controls	Universal consent platforms	OneTrust, Evidon
Predictability, manageability	User Controls	Data portability	Data Transfer Project
Manageability	User Controls	Data access	Google Takeout, Apple Data & Privacy

Table 4: Summary of technologies

9. References

[1] “A Brief Introduction to XACML.” OASIS | Advancing Open Standards for the Information Society, www.oasis-open.org/committees/download.php/2713/Brief_Introduction_to_XACML.htm.

[2] datatransferproject.dev, datatransferproject.dev/. Accessed 26 Sept. 2018.

[3] “Adblock Plus | The World's # 1 Free Ad Blocker.” Adblock Plus | The World's # 1 Free Ad Blocker, Eyeo GmbH, adblockplus.org/. Accessed 26 Sept. 2018.
[4] “Brave Features.” Brave Browser, Brave Software Inc., brave.com/features/. Accessed 26 Sept. 2018.
[5] Chiavetta, Ryan. “Revamped Solution Helps Marketing Teams Comply with Privacy Laws.” Inside the EPrivacy Regulation's Furious Lobbying War, IAPP, 11 Sept. 2018, iapp.org/news/a/revamped-solution-helps-marketing-teams-comply-with-privacy-laws/. Accessed 26 Sept. 2018.
[6] Chiavetta, Ryan. “New Solution Tackles Gathering Mobile, IoT Consent.” IAPP, IAPP, 26 July 2018, iapp.org/news/a/new-solution-tackles-gathering-mobile-iot-consent/. Accessed 26 Sept. 2018.
[7] De Luca, Alexander, et al. “Expert and Non-Expert Attitudes towards (Secure) Instant Messaging.” Symposium on Usable Privacy and Security, vol. 12, 22 June 2016, pp. 147–157., www.usenix.org/system/files/conference/soups2016/soups2016-paper-de-luca.pdf.
[8] “Differential Privacy.” Apple, Apple, www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf. Accessed 26 Sept. 2018.
[9] Disconnect. “Take Back Your Privacy.” Disconnect, Disconnect, disconnect.me/. Accessed 26 Sept. 2018.
[10] “Do Not Track.” Electronic Frontier Foundation, Electronic Frontier Foundation, www.eff.org/issues/do-not-track. Accessed 26 Sept. 2018.
[11] Dwork, Cynthia, and Aaron Roth. The Algorithmic Foundations of Differential Privacy. Now Publ., 2014.
[12] European Union Agency for Network and Information Security. “Privacy Enhancing Technologies - ENISA.” Risk Management & Information Security Management Systems - ENISA, European Union Agency for Network and Information Security, 1 Apr. 2016, www.enisa.europa.eu/topics/data-protection/privacy-enhancing-technologies. Accessed 26 Sept. 2018.
[13] Ewerlf, Alex. “GDPR Pseudonymization Techniques – Alex Ewerlöf – Medium.” Medium, Medium, 31 May 2018, medium.com/@alexewerlof/gdpr-pseudonymization-techniques-62f7b3b46a56. Accessed 26 Sept. 2018.
[14] “Firefox Focus.” Products | Mozilla Support, Mozilla, support.mozilla.org/en-US/kb/focus. Accessed 26 Sept. 2018.
[15] “Ghostery Makes the Web Cleaner, Faster and Safer!” Ghostery, Cliqz, www.ghostery.com/. Accessed 26 Sept. 2018.
[16] Wang, Yang, et al. “Privacy-Enhancing Technologies.” Handbook of Research on Social and Organizational Liabilities in Information Security, Information Science Reference, Hershey, PA, 2009, pp. 203–227.
[17] Harkous, Hamza, et al. “Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning.” 27th USENIX Security Symposium, vol. 27, 15 Aug.

2018, pp. 531–548., www.usenix.org/system/files/conference/usenixsecurity18/sec18-harkous.pdf.
[18] Heimes, Rita. “Top 10 Operational Impacts of the GDPR: Part 1 – Data Security and Breach Notification.” Top 10 Operational Impacts of the GDPR: Part 1 – Data Security and Breach Notification, IAPP, iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-1-data-security-and-breach-notification/. Accessed 26 Sept. 2018.
[19] Heimes, Rita. “Top 10 Operational Impacts of the GDPR: Part 5 - Profiling.” IAPP, IAPP, 20 Jan. 2016, iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-5-profiling/. Accessed 26 Sept. 2018.
[20] Heimes, Rita. “Top 10 Operational Impacts of the GDPR: Part 9 - Codes of Conduct and Certifications.” IAPP, IAPP, 24 Feb. 2016, iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-9-codes-of-conduct-and-certifications/. Accessed 26 Sept. 2018.
[21] “Home.” Basic Attention Token, basicattentiontoken.org/. Accessed 26 Sept. 2018.
[22] “How To Comply with the Privacy of Consumer Financial Information Rule of the Gramm-Leach-Bliley Act.” Federal Trade Commission, 7 Jan. 2015, www.ftc.gov/tips-advice/business-center/guidance/how-comply-privacy-consumer-financial-information-rule-gramm. Accessed 26 Sept. 2018.
[23] “HTTPS Everywhere.” Electronic Frontier Foundation, Electronic Frontier Foundation, 27 Mar. 2018, www.eff.org/https-everywhere. Accessed 26 Sept. 2018.
[24] Kelley, Patrick Gage, et al. “Standardizing Privacy Notices.” Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI 10, 2010, doi:10.1145/1753326.1753561.
[25] Li, Ninghui, et al. “t-Closeness: Privacy Beyond k-Anonymity and l-Diversity.” 2007 IEEE 23rd International Conference on Data Engineering, 2007, doi:10.1109/icde.2007.367856.
[26] Liu, Bin, et al. “Follow My Recommendations: A Personalized Privacy Assistant for Mobile App Permissions.” Twelfth Symposium on Usable Privacy and Security, vol. 12, 22 June 2016, www.usenix.org/system/files/conference/soups2016/soups2016-paper-liu.pdf.
[27] Maldoff, Gabe. “Top 10 Operational Impacts of the GDPR: Part 3 – Consent.” Top 10 Operational Impacts of the GDPR: Part 3 – Consent, IAPP, 12 Jan. 2016, iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-3-consent/. Accessed 26 Sept. 2018.
[28] Maldoff, Gabe. “Top 10 Operational Impacts of the GDPR: Part 6 - RTBF and Data Portability.” IAPP, IAPP, 25 Jan. 2016, iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-6-rtbf-and-data-portability/. Accessed 26 Sept. 2018.
[29] Maldoff, Gabe. “Top 10 Operational Impacts of the GDPR: Part 8 - Pseudonymization.” IAPP, IAPP, 12 Feb. 2016,

iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-8-pseudonymization/. Accessed 26 Sept. 2018.

[30] Myers, Anna. “Top 10 Operational Impacts of the GDPR: Part 4 - Cross-Border Data Transfers.” Inside the EPrivacy Regulation's Furious Lobbying War, IAPP, 19 Jan. 2016, iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-4-cross-border-data-transfers/. Accessed 26 Sept. 2018.

[31] Myers, Anne. “Top 10 Operational Impacts of the GDPR: Part 7 - Vendor Management.” IAPP, IAPP, 4 Feb. 2016, iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-7-vendor-management/. Accessed 26 Sept. 2018.

[32] “OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data.” OECD, Organization for Economic Cooperation and Development, 23 Sept. 1980, www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm. Accessed 26 Sept. 2018.

[33] Office of the Privacy Commissioner of Canada. “Privacy Enhancing Technologies – A Review of Tools and Techniques.” Office of the Privacy Commissioner of Canada, Office of the Privacy Commissioner of Canada, 15 Nov. 2017, www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2017/pet_201711/. Accessed 26 Sept. 2018.

[34] “Pi-Hole®: A Black Hole for Internet Advertisements.” Pi-Hole®: A Black Hole for Internet Advertisements, Pi-Hole, LLC, pi-hole.net/. Accessed 26 Sept. 2018.

[35] “Privacy and Security Enforcement.” Federal Trade Commission, Federal Trade Commission, 23 Aug. 2018, www.ftc.gov/news-events/media-resources/protecting-consumer-privacy/privacy-security-enforcement. Accessed 26 Sept. 2018.

[36] United States, Congress, “Privacy Online: Fair Information Practices in the Electronic Marketplace: a Report to Congress.” Privacy Online: Fair Information Practices in the Electronic Marketplace: a Report to Congress, Federal Trade Commission, 2000.

[37] Redmon, Gant. “Incident Response Under GDPR: Before, During and After a Data Breach.” Security Intelligence, SecurityIntelligence, 27 July 2018, securityintelligence.com/incident-response-under-gdpr-what-to-do-before-during-and-after-a-data-breach/. Accessed 26 Sept. 2018.

[38] Reyes, Irwin, et al. “‘Won’t Somebody Think of the Children?’ Examining COPPA Compliance at Scale.” Proceedings on Privacy Enhancing Technologies, vol. 2018, no. 3, Jan. 2018, pp. 63–83., doi:10.1515/popets-2018-0021.

[39] Schaub, Florian, et al. “A Design Space for Effective Privacy Notices*.” The Cambridge Handbook of Consumer Privacy, pp. 365–393., doi:10.1017/9781316831960.021.

[40] “SELF-REGULATORY PRINCIPLES.” Digital Advertising Alliance (DAA) Self-Regulatory Program, www.aboutads.info/principles/. Accessed 26 Sept. 2018.

[41] “Signal >> Home.” Signal Messenger, Signal Messenger, signal.org/. Accessed 26 Sept. 2018.

[42] Sweeney, Latanya. “Simple Demographics Often Identify People Uniquely.” Data Privacy Working Paper 3, 2000, dataprivacylab.org/projects/identifiability/paper1.pdf.
[43] “TigerText Essentials | Products.” TigerConnect, www.tigerconnect.com/products/tigertext-essentials/#parentHorizontalTab1. Accessed 26 Sept. 2018.
[44] United States, Congress, Wannisky, Kathleen E. “Department of Health and Human Services, Office of the Secretary: Health Insurance Reform: Security Standards.” Department of Health and Human Services, Office of the Secretary: Health Insurance Reform: Security Standards, U.S. General Accounting Office, 2003.
[45] Whitten, Alma, and J.D. Tygar. “Why Johnny Can't Encrypt: a Usability Evaluation of PGP 5.0.” SSYM'99 Proceedings of the 8th Conference on USENIX Security Symposium , vol. 8, 23 Aug. 1999, p. 14.
[46] Wilson, Shomir, et al. “The Creation and Analysis of a Website Privacy Policy Corpus.” Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2016, doi:10.18653/v1/p16-1126.
[47] Wilson, Shomir, et al. “Demystifying Privacy Policies Using Language Technologies: Progress and Challenges.” TA-COS ’16: LREC Workshop on Text Analytics for Cybersecurity and Online Safety, May 2016, usableprivacy.org/static/files/swilson_ta-cos_2016.pdf.
[48] “YourAdChoices Gives You Control.” YourAdChoices.com, Digital Advertising Alliance, www.youradchoices.com/. Accessed 26 Sept. 2018.
[49] HHS Office of the Secretary,Office for Civil Rights, and OCR. “Combined Text of All Rules.” HHS.gov, US Department of Health and Human Services, 12 May 2017, www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/combined-regulation-text/index.html.
[50] “General Data Protection Regulation (GDPR) – Final Text Neatly Arranged.” General Data Protection Regulation (GDPR), gdpr-info.eu/.
[51] “Platform for Privacy Preferences (P3P) Project.” Same Origin Policy - Web Security, www.w3.org/P3P/.
[52] “Tor.” Tor Project: Overview, Tor Project, www.torproject.org/about/overview.html.en.
[53] “Monero: Home.” Getmonero.org, The Monero Project, getmonero.org/. Accessed 26 Sept. 2018.
[54] “Recent Zcash Blog Posts.” Zcash - How Zk-SNARKs Work in Zcash, z.cash/. Accessed 26 Sept. 2018.
[55] Bocovich, Cecylia, and Ian Goldberg. "Secure asymmetry and deployability for decoy routing systems." Proceedings on Privacy Enhancing Technologies 2018.3 (2018): 43-62.
[56] Barradas, Diogo, Nuno Santos, and Luís Rodrigues. "DeltaShaper: Enabling unobservable censorship-resistant TCP tunneling over videoconferencing streams." Proceedings on Privacy Enhancing Technologies 2017.4 (2017): 5-22.
[57] “Secure Multi-Party Computation.” Wikipedia, Wikimedia Foundation, 23 Aug. 2018, en.wikipedia.org/wiki/Secure_multi-party_computation.

[58] Hesamifard, Ehsan, et al. "Privacy-preserving Machine Learning as a Service." Proceedings on Privacy Enhancing Technologies 2018.3 (2018): 123-142.
[59] Gascón, Adrià, et al. "Privacy-preserving distributed linear regression on high-dimensional data." Proceedings on Privacy Enhancing Technologies 2017.4 (2017): 345-364.
[60] Freyberger, Michael, et al. "Cracking ShadowCrypt: Exploring the Limitations of Secure I/O Systems in Internet Browsers." Proceedings on Privacy Enhancing Technologies 2018.2 (2018): 47-63.
[61] Rogaway, Phillip, and Yusi Zhang. "Onion-AE: foundations of nested encryption." Proceedings on Privacy Enhancing Technologies 2018.2 (2018): 85-104.
[62] Unger, Nik, and Ian Goldberg. "Improved Strongly Deniable Authenticated Key Exchanges for Secure Messaging." Proceedings on Privacy Enhancing Technologies 2018.1 (2018): 21-66.
[63] Bild, Raffael, Klaus A. Kuhn, and Fabian Prasser. "SafePub: A Truthful Data Anonymization Algorithm With Strong Privacy Guarantees." Proceedings on Privacy Enhancing Technologies 2018.1 (2018): 67-87.
[64] Murakami, Takao, Hideitsu Hino, and Jun Sakuma. "Toward Distribution Estimation under Local Differential Privacy with Small Samples." Proceedings on Privacy Enhancing Technologies 2018.3 (2018): 84-104.
[65] Sen, Shayak, et al. “Bootstrapping Privacy Compliance in Big Data Systems.” 2014 IEEE Symposium on Security and Privacy, 2014, doi:10.1109/sp.2014.28.
[66] Anon. 2018. Privacy and machine learning: two unexpected allies? (April 2018). Retrieved November 13, 2018 from http://www.cleverhans.io/privacy/2018/04/29/privacy-and-machine-learning.html
[67] Erlingsson, Úlfar, Vasyl Pihur, and Aleksandra Korolova. "Rappor: Randomized aggregatable privacy-preserving ordinal response." Proceedings of the 2014 ACM SIGSAC conference on computer and communications security. ACM, 2014.
[68] Johnson, Noah, Joseph P. Near, and Dawn Song. "Towards practical differential privacy for SQL queries." Proceedings of the VLDB Endowment 11.5 (2018): 526-539.
[69] Anon. 2017. Federated Learning: Collaborative Machine Learning without Centralized Training Data. (April 2017). Retrieved November 13, 2018 from https://ai.googleblog.com/2017/04/federated-learning-collaborative.html?m=1
[70] Anon. Learning with Privacy at Scale - Apple. Retrieved November 13, 2018 from https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html
[71] Anon. Differential Privacy Overview - Apple. “Differential Privacy” Retrieved November 13, 2018 from https://www.apple.com/privacy/
[72] Papernot, Nicolas, et al. "Semi-supervised knowledge transfer for deep learning from private training data." arXiv preprint arXiv:1610.05755 (2016).
[73] Nergiz, M. Ercan, and Chris Clifton. "S-Presence Without Complete World Knowledge." (2008).
[74] Tiwari, Anisha, and Minu Choudhary. "A Review on K-Anonymization Techniques."

[75] Anon. The Difficulty Of Private Contact Discovery. Retrieved November 26, 2018 from https://signal.org/blog/contact-discovery/

[76] Brooks, Sean, et al. An introduction to privacy engineering and risk management in federal systems. US Department of Commerce, National Institute of Standards and Technology, 2017.

[77] Anon. 2018 Privacy Tech Vendor Report. Retrieved November 26, 2018 from https://iapp.org/resources/article/2018-privacy-tech-vendor-report/

[78] Ruegg, Rosalie, and Gretchen Jordan. "Overview of evaluation methods for R&D programs." A directory of evaluation methods relevant to technology development programs, prepared for US Department of Energy, Office of Energy Efficiency and Renewable Energy (2007).

[79] Mark Bergen and Jennifer Surane. Google and Mastercard Cut a Secret Ad Deal to Track Retail Sales. Retrieved November 26, 2018 from https://www.bloomberg.com/news/articles/2018-08-30/google-and-mastercard-cut-a-secret-ad-deal-to-track-retail-sales

[80] Anon. No Knowledge, Secure-by-Default Products. Retrieved November 26, 2018 from https://spideroak.com/no-knowledge/

[81] Removed

[82] Removed

[83] Removed

[84] Amir Nasr. 2017. Poll: Little Trust That Tech Giants Will Keep Personal Data Private. (April 2017). Retrieved November 26, 2018 from https://morningconsult.com/2017/04/10/poll-little-trust-tech-giants-will-keep-personal-data-private/

[85] Kevin Granville. 2018. Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens. (March 2018). Retrieved November 26, 2018 from https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html

[86] Robin Kurzer. 2018. The United States finally starts to talk about data privacy legislation. (August 2018). Retrieved November 26, 2018 from https://martechtoday.com/the-united-states-finally-starts-to-talk-about-data-privacy-legislation-219299

[87] Cybersecurity, Critical Infrastructure. "Framework for Improving Critical Infrastructure Cybersecurity." Framework 1 (2014): 11.

[88] Ding, Bolin, Janardhan Kulkarni, and Sergey Yekhanin. "Collecting telemetry data privately." Advances in Neural Information Processing Systems. 2017.

[89] Kimberly Amadeo. World's Largest Economies. Retrieved November 26, 2018 from https://www.thebalance.com/world-s-largest-economy-3306044

[90] Investors are pushing proposals to rein in Wells Fargo (WFC). Wells Fargo timeline: Bank's 20-month nightmare. Retrieved November 26, 2018 from https://money.cnn.com/2018/04/24/news/companies/wells-fargo-timeline-shareholders/index.html

[91] Zoe Kleinman. 2017. Netflix defends A Christmas Prince tweet. (December 2017). Retrieved November 26, 2018 from https://www.bbc.com/news/technology-42323366

[92] Matt Weinberger. 2018. The Equifax breach resulted in the leak of 56,200 drivers' licenses, passports, and other forms of ID. (May 2018). Retrieved November 26, 2018 from https://www.businessinsider.com/equifax-breach-check-details-update-2018-5

[93] Anon. Partnership on AI. Retrieved November 26, 2018 from https://www.partnershiponai.org/partners/

10. Appendix I

10.1. Interview Recruitment Text

A group of Carnegie Mellon University Privacy Engineering students is seeking experts in privacy technology for a brief 15-30 minute interview as part of a course project. They are interested in learning about recent experiences and opinions of privacy technology (especially any experiences with cloud technology) throughout industry or academia.

All information provided will be kept confidential and only available to the individuals conducting the study. The study will not disclose or share your name, organization, contact information, or participation in this study. The findings will be used for a report, which will be available on the department's website in December 2018.

If you are interested and available either in person, over the phone, or through video-conferencing, please complete this form: https://goo.gl/forms/MQx6F0ChcceDPvMj1 or email: thomasjm@andrew.cmu.edu

This project is supervised by Professor Lorrie Cranor and Professor Nicolas Christin

10.2. Interview Script

10.2.1. Introduction

The purpose of this study is to learn about the current landscape of privacy technologies. We are students at Carnegie Mellon University studying privacy engineering.

All information provided as part of the interview will be kept completely confidential. Additionally, we will not disclose or share your name, organization, contact information, or participation in this study. At most, we would use a distinction of your choice such as "privacy expert working in the financial industry". Otherwise we will refer to you as privacy expert #_____ Is there any description we can use?

10.2.2. Questions

- Can you describe your role and responsibilities in this position?
- Can you describe the information privacy concerns of your organization? (regulations, users, consent, marketing, data processing, employees, etc.)
- Do you currently use any technologies to address privacy? Or does anyone you work with use these technologies? (Describe a use-case of this technology)

- Any specific laws you must address? Does your organization operate in Europe?
- Does your organization work with large sets of user data? Any specialized datasets (e.g. genomics)? Privacy concerns?
- Does your organization take advantage of cloud platforms (AWS, Azure, GCP)? Privacy concerns there?
- Are there any technologies your company has not adopted but very interested in?
- Any privacy technology you think is very compelling or potentially impactful?

11. Appendix II

11.1. Non-organizational Technologies

Individual users employ a variety of PETs. These are excluded from our report due to our focus on technologies that organizations can adopt. However, these can be influential technologies due to their indirect impact on organizations. For example, individuals who take advantage of these technologies could impact organizations significantly. Here, we focus on those technologies used to address information privacy concerns. These technologies mostly provide different functionalities related to allowing the individual to control the flow of their data into the world.

Encryption of data-in-transit such as using Transport Layer Security or HTTPS provides users with a higher assurance of confidentiality for the data they send while online. By itself, this encryption is not a privacy technology of interest, however tools such as HTTPS Everywhere allow users to choose to use encryption as the default when browsing [23]. Applications such as this fulfill the user's preference for communications secure from network surveillance actors such as governments, ISPs, or malicious actors on local networks. Secure communication technologies follow a similar user choice for safeguarding private communications from surveillance or service providers. For example, Signal provides an end-to-end encryption technology for messaging [41]. Secure messaging applications provide a high assurance of confidentiality from even more sophisticated threats such as government surveillance as well as potential abuses by messaging service providers.

Another aspect of user's controlling their information is restricting the flow of information to undesired third parties. The primary example of this appears in the advertisement ecosystem across the Web. This system employs a variety of user tracking technologies (cookies, pixels, fingerprinting techniques, etc.) to collect information on Web users for the purposes of the tracking organizations. Many users employ tracker and advertisement blocking software technologies, across mobile and desktop devices, to apply controls over this flow of their information, mainly to restrict or block. Browser extension solutions such as Ghostery, Ad Block Plus, and Disconnect supplement the user's browser to block advertisements and trackers [15, 3, 9]. Other solutions replace the user's default browser with a specialized, ad-blocking browser such as Firefox Focus on mobile devices [14]. Furthermore, network layer technologies are emerging that users add to their home network to block ads and trackers across all devices in

their home such as the open-source project Pi-hole which uses Raspberry Pi devices on home networks [34]. One approach towards ad-blocking appears with the Brave web browser [4]. Currently, the browser provides an ad-free experience, however future goals of the project focus on integrating the Basic Attention Token into the ad ecosystem [21]. This represents an attempt to use blockchain and cryptocurrency to transform the current ad ecosystem by "[paying] publishers for their content and users for their attention, while providing advertisers with more in return for their ads" [21]. The cryptocurrency aspects of the project allow for a more transparent and decentralized approach for ad-based revenue across the Web.

Another element of addressing Web privacy concerns related to advertisements and trackers involves privacy policy implementations such as the Do Not Track header [10]. Do Not Track is an attempt at universal web policy for opting out of tracking. AdChoices represents a similar advertising technology created by the Digital Advertising Alliance [40]. However, AdChoices provides only restrictions the use of data for interest-based advertising, without limiting collection [40].

Anonymity systems present another subset of user technology that limits data collection. An anonymity system is designed to ensure one set of information flow is indistinguishable and unattributable. One common anonymity technology is the Tor Project [52]. Tor provides a network level of anonymity to users. This prevents network observers, such as organizations performing network traffic analysis, from associating network traffic with a specific sender [52]. For a user, Tor functions similar to a Virtual Private Network technology, allowing the user to connect to sites across the web while sending their traffic through the Tor network instead of the VPN. VPNs and network proxies provide a similar functionality, however they rely on trusting the service provider to maintain the user's anonymity. Anonymity appear in areas other than network communication, such as for protecting transaction in cryptocurrencies like Monero and Zcash [53, 54].

Anti-censorship technologies circumvent censorship systems. While not user or general organizational technologies, these technologies could also be impactful for organizations. One example is in the area of decoy routing systems. These offer a solution to censorship resistance that uses real connections to unblocked sites to deliver censored content to users [55]. Cecylia Bocovich and Ian Goldberg recently proposed a new technique that is more secure than previous proposals and provides an option for tiered deployment, allowing more deployments of lightweight, non-blocking relay stations that aid in defending against adversaries [55]. Further network-centric approaches towards circumventing censorship exist such as DeltaShaper, a censorship resistant research technology that offers a data-link interface and supports TCP/IP applications that tolerate low throughput/high latency links [56]. This technology attempts to support the use of technologies like videoconferencing applications (Skype) by providing a carrier for unobservable covert TCP/IP communications [56].

11.2. Research in Stronger Privacy Guarantees

A number of areas of active research involve expanding the strength of privacy guarantees available in data-driven systems. These represent approaches towards limiting data collection by organizations such as secure computation, a "subfield of cryptography with the goal of

creating methods for parties to jointly compute a function over their inputs while keeping those inputs private" [57]. This technology presents a significant opportunity to preserve privacy while allowing for data-driven machine learning algorithms across a large number of fields. Recent work includes implementing deep neural network algorithms in the encrypted domain, and developing techniques to adopt neural networks within the practical limitations of current homomorphic encryption schemes [58]. Some organizations have built products with advanced confidentiality guarantees enabled by these types of technologies [80].

Secure systems remain of great significance to all parties in information privacy. However, existing systems are not perfect, and improving their security properties remains an open challenge. Security properties such as confidentiality are provided through technologies such as encryption. These technologies, while mature, are not complete, and we see significant advances that improve or expose flaws in technologies like encryption. Freyberger et al. analyzed the limitations of recent attempts to secure user input within web browsers (ShadowCrypt IO system) [60]. Rogaway et al. recently provided a provable-security treatment for onion authenticated-encryption as according to their research, the encryption technique presently used in Tor does not meet the requirements of the application [61]. And there are also some technologies to enhance the secure systems. Unger proposed three new strongly deniable key exchange protocols, which are designed to be used in modern secure messaging applications while eliminating the weaknesses of previous approaches [62]. This type of research supports the many applications and systems built on top of these technologies.

Privacy preserving algorithms and differential privacy have become popular both amongst researchers and organizations. These technologies already play an important role in many different fields. Bild presented a data publishing algorithm that satisfies the differential privacy model. Insead of perturbing input data or generating synthetic output data, records are randomly drawn from the input dataset and the uniqueness of their features is reduced. This provides a generic approach that can be parameterized with different objective functions for different applications [63]. Murakami proposed an improvement in local differential privacy, a model of differential privacy that removes the risks of centralized database containing the data with potentials for leakage [64]. We see from the ongoing work that the privacy protections these technologies provide continue to strengthen.

11.3. Security Safeguards

While GDPR is a new and comprehensive regulation, organizations adapting to long existing regulations provide some insights into the direction of PETs. For example, the U.S. Health Insurance Portability and Accountability Act of 1996 (HIPAA) placed regulations on information privacy and security in the health sector [49]. One significant requirement of this regulation is the *Security Rule*. This rule requires covered-entities (usually hospitals) to provide a minimum level of security mechanisms for electronic health records [44]. This in turn resulted in technology designed to meet these requirements such as the TigerConnect secure messaging platform for clinical communications [43]. Traditionally, usability challenges plagued early implementations of encrypted or secure messaging and limited adoption to security experts [45].

Regulations, like HIPAA, provided additional motivation for addressing these challenges and eventually resulted in organizations adopting technology solutions.

GDPR also created new requirements for handling security breaches [50]. These requirements force organizations to formalize some of their security procedures especially when dealing with responding to security incidents or data breaches. Security is a requirement for an organization to meet their privacy objectives. Some organizations fulfill the breach requirements of GDPR by integrating security orchestration, automation, and response (SOAR) platforms into their security operations centers [37]. In these cases, the goal of the technology is to enforce and automate the procedures required by regulations in response to a breach. Furthermore, Article 32 of GDPR specifies data controllers and processors must “implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk” [50]. This section mandates appropriate security measures as a requirement to meeting the privacy regulation. This section significantly impacts organizations, however for the purposes of this review we will not focus on purely information security technologies as PETs, only those that augment privacy in some unique and significant form. This limitation keeps the scope of the project with reason, as security technology is far too large a field to consider in addition to privacy technology.