Seeking the Ideal Privacy Protection: Strengths and Limitations of Differential Privacy

Kazutoshi Kan

In modern society where personal information has high industrial value, privacy protection is a mandatory prerequisite for utilizing the personal information. Differential privacy enables to achieve moderate privacy through quantifying the effectiveness of privacy-enhancing technologies. Many researchers have adopted differential privacy as a common and useful criterion in academic literatures regarding the privacy evaluation. This paper gives an overview of principles, laws, regulations, IT systems management, business practices, and privacy-enhancing technologies including ones based on differential privacy. It also explains the theory behind differential privacy and its application studies, and discusses the desirable privacy protection considering the strengths and limitations of the differential privacy. In particular, mathematical methodologies including ones based on differential privacy cannot solely suffice social demands for privacy protection, especially for the control over personal information about oneself. Desirable privacy protection for resolving the social issue should adopt a comprehensive approach that includes laws, regulations, IT systems management, business practices, as well as mathematical methodologies and information security.

Keywords: Differential privacy; Privacy protection; Control over personal information about oneself; Anonymization; Ethical, legal and social issues
JEL Classification: K22, Z00

Director, Institute for Monetary and Economic Studies, Bank of Japan
(E-mail: kazutoshi.kan@boj.or.jp)

The author would like to thank Hiroaki Kikuchi (Meiji University) and Keiko Yamamoto (Bank of Japan) for their useful comments. The views expressed in this paper are those of the author and do not necessarily reflect the official views of the Bank of Japan.

I. Introduction

In modern society, personal information[1] is being continuously and automatically collected at an unprecedented level of granularity through the widespread use of smartphones and electronic commerce. The high industrial value of personal information has led to its expanding use, e.g., the development of databases of personal information, automation of making decisions regarding personal loans and membership registration through machine learning (artificial intelligence, AI), utilization of microdata, and cross-border transfer of personal data.

When utilizing personal information, there must be a reasonable balance between social benefits and individual rights. The social benefits include emerging new industries, preventing disaster and crime, and improving the efficiency of official procedures and their relevant services. One example is the 12-digit identification number assigned to every resident in Japan called the Individual Number (also known as *My Number*). However, the expanded use of personal information increases the risk of privacy violations. For example, the leakage of small bits of personal data may expose the entirety of the corresponding personal profile when various personal information is collected into a database.[2] Databases of personal information can be combined with AI for sophisticated profiling of individuals. As Kukita (2020) points out, when profiling is applied to insurance, human resource management, law enforcement, and court decisions, it may lead to serious problems such as discriminatory biases.[3] These can be viewed as ethical, legal, and social issues (commonly referred to as ELSI) posed by AI. ELSI is often discussed in the field of AI ethics from the perspective of fairness and privacy.

Differential privacy (Dwork *et al.* [2006]) is useful in achieving a reasonable balance between the utilization of personal information and the privacy protection under their trade-off relationship. Differential privacy is a standard quantitative criterion in academia for evaluating the strength of the privacy protection of privacy-enhancing technologies (PETs). The concept of the evaluation is based on unconditional security, or information-theoretic security, which does not assume any attack models. Due to this advantageous property, privacy guarantees based on differential privacy have been considered more desirable for the following reasons. First, the information and computational resources available to attackers have been rapidly increasing. This makes it more difficult to ensure privacy under a specific and predetermined assumption of attackers' knowledge (the victim's personal information except for the target information to be exposed or stolen).[4] As a result, information-theoretic approaches have become of in-

[1] In this paper, personal information refers to information about an individual. Unless otherwise noted, this definition differs from that of the Act on the Protection of Personal Information. This paper also takes the point of view that the protection of personal information is included in privacy protection, defined as the control over information about oneself, which will be discussed later.

[2] In the United States, private credit bureaus collect personal information to calculate credit scores for individuals. A typical example is the FICO credit score provided by Fair Isaac Corporation. The credit scores have significant impacts on consumers' financial activities such as borrowing loans, purchasing insurance, and receiving other financial services. See Hayashi (2022) for a discussion of the legal issues surrounding credit scores in Japan, with reference to credit scores and the laws and regulations governing them in the United States.

[3] See Benjamin (2019) and O'Neil (2016) for further discussion of the discriminatory biases.

[4] Regarding privacy violations, Sweeney (2002) pointed out that personal data from health insurance in Massachusetts could be linked with those from the electoral rolls in Cambridge, Massachusetts using gender, zip

creasing importance. Second, as the threat of privacy violations increases, more proactive measures to prevent personal information leakage will be required. Information-theoretic security based on differential privacy is advantageous in the meaning that the privacy protection would be still effective even when new attack methods are discovered in the future.

The United States Census Bureau (2019) adopted privacy protection methods based on differential privacy starting with the 2020 Census, replacing *ad-hoc* methods such as cell suppression[5] and data swapping.[6] The decision stemmed from the increasing threat of database reconstruction attacks, which attempt to recover part or all of a database of personal information from a combination of statistic data generated from the database. The threat is no longer only theoretical but imminent and requires practical countermeasures (Garfinkel, Abowd, and Martindale [2019], Census Scientific Advisory Committee [2021]). Major IT companies, such as Google and Uber, have also introduced privacy-preserving methods based on differential privacy into their personal data collection.

Differential privacy, however, is not a panacea (see Section IV for details). There is no consensus, either theoretically or practically, on how to set parameter ϵ (privacy budget) that determines the strength of privacy protection. The Laplace mechanism, a simple and commonly used method based on differential privacy, adds large noises to raw data. As a result, the modified data is often too inaccurate to be used to generate statistical data. Designing a mechanism to eliminate this drawback is difficult because it requires a high level of expertise in practical data analysis and theoretical statistics. Assumptions of a database of personal information to guarantee a certain level of privacy protection by using differential privacy are not always satisfied in practical data. These challenges must be overcome for each application. Furthermore, PETs (privacy enhancing technologies) based on differential privacy are applicable to databases of personal information which accept and respond to arbitrary statistical queries. Thus, differential privacy is not a promising option in all situations. When adopting differential privacy as a practical criterion, both its strengths and limitations should be considered.

The social aspect of privacy protection cannot be addressed solely by mathematical engineering including differential privacy. The growing concern that private companies or governments who collect and handle huge personal information has led to an international trend toward adopting the concept of the *right to control over personal information about oneself.*[7] This extends the definition of privacy as *protection of sensitive information about an individual.*[8] Updating the goal of privacy protection is a

code, and date of birth as identifiers, even though information on names was not available. For other examples of privacy violations, see Sakuma (2016).

[5] Cell suppression is a method that conceals a part of tabular data in accordance with certain criteria. For example, a cell is hidden if the contribution of an individual for its value exceeds a predetermined threshold level.

[6] Data swapping (Dalenius and Reiss [1982], Willenborg and Waal [2001]) is a method that swaps attribute values between individual record data. It was adopted in the 2010 U.S. Census.

[7] For more information on the right to control personal information about oneself, see Nakagawa (2016) and Sogabe, Hayashi, and Kurita (2019).

[8] In this paper, the term “sensitive information” refers to information about an individual that he/she does not want others to know. Note that the definition differs from that in the Guidelines for the Protection of Personal

social challenge beyond the scope of engineering. Thus, it is necessary to take into account not only technological elements such as mathematical methodologies and information security but also legal and regulatory systems, IT systems management, and business practices.

The structure of this paper is as follows. Section II provides an overview of the elements of privacy protection, such as principles, laws, regulations, IT systems management, business practices, and related technologies including PETs. A group of these elements is referred to as a *privacy protection framework* in this paper. Section III reviews mathematical methodologies, including differential privacy. Section IV introduces the theory behind differential privacy. Section V presents the main research results of applying a concept of differential privacy. Section VI discusses the obstacles to the ideal privacy protection, considering the strengths and limitations of differential privacy.

II. Concept of Privacy and Protection Framework

A privacy protection framework aims to maximize the utility of personal information under constraints of social acceptance. This section supposes that the framework consists of three categories: *principles*, *rules*, and *methodologies* (see Figure 1). The principles define the concept of privacy and what privacy protection should accomplish. The rules include domestic laws, international arrangements, and industry self-regulations. The methodologies include technologies and mechanisms implemented for privacy protection. The methodologies can be divided into two classes: those to ensure information security, and those to provide functionalities aside from information security.

These categories work independently but complementary to each other to achieve privacy protection for the entire lifetime of personal information. In particular, information technology and mathematical technology play an important role in privacy protection in modern society, as personal information is accumulated massively, stored in databases, and processed on IT systems.

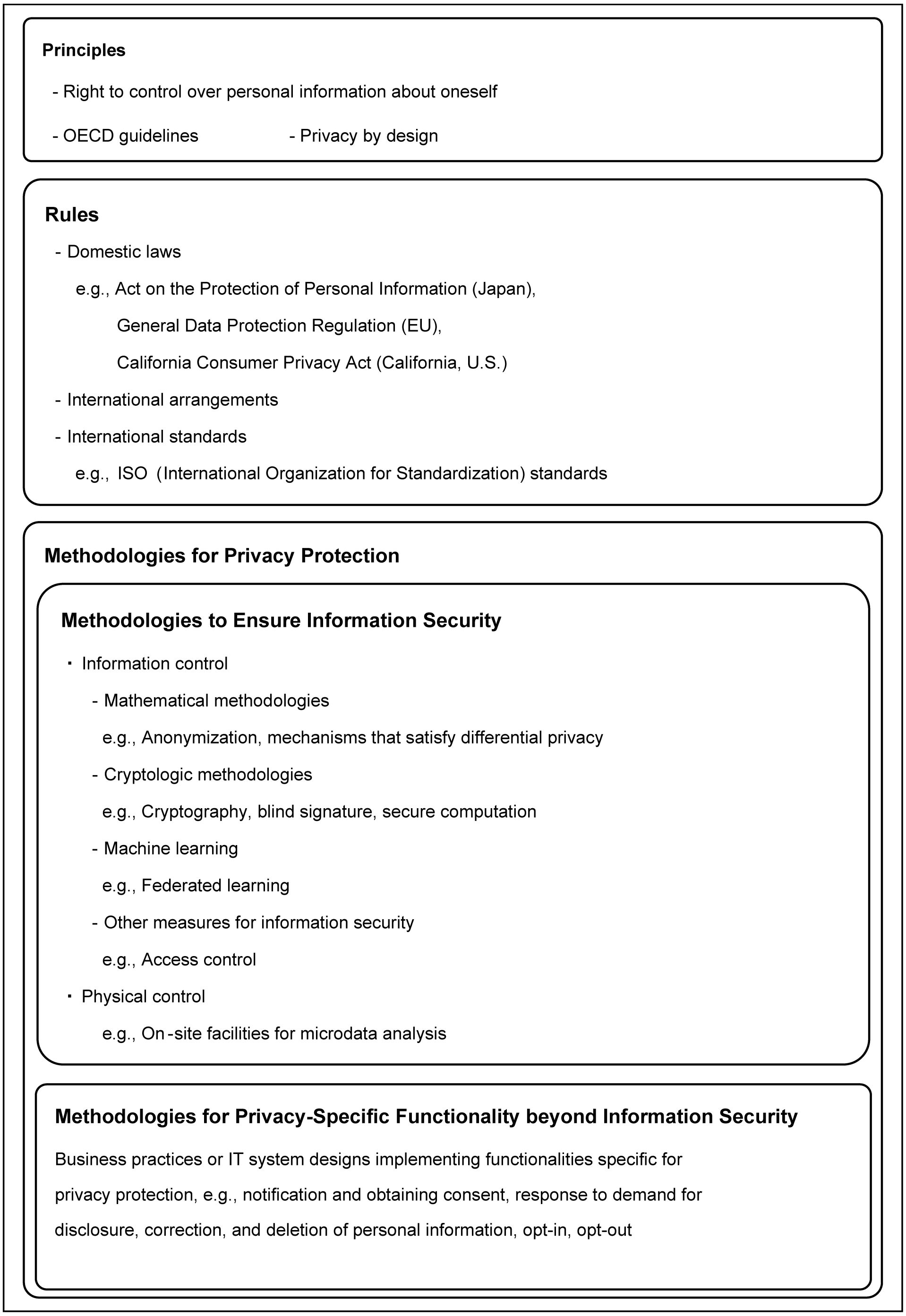
This section describes each category and social demand for privacy.

A. Principles of Privacy Protection

Traditionally, the main purpose of privacy has been to *prevent the leakage of personal information*. Since the 1960s, it has been said that the concept of privacy should include *the right to control over personal information about oneself*. This new right allows individuals to demand disclosure, correction, or deletion of their personal information. Rapid advances in information technology enable governments and private companies to develop massive databases of personal information by making use of advanced information technology. In the modern age of the Internet, the concept of privacy has

Information in the Financial Sector.

Figure 1 Overview of Privacy Protection Framework



Principles

- Right to control over personal information about oneself

- OECD guidelines - Privacy by design

Rules

- Domestic laws
  e.g., Act on the Protection of Personal Information (Japan),
  General Data Protection Regulation (EU),
  California Consumer Privacy Act (California, U.S.)
- International arrangements
- International standards
  e.g., ISO (International Organization for Standardization) standards

Methodologies for Privacy Protection

Methodologies to Ensure Information Security

・ Information control
- Mathematical methodologies
  e.g., Anonymization, mechanisms that satisfy differential privacy
- Cryptologic methodologies
  e.g., Cryptography, blind signature, secure computation
- Machine learning
  e.g., Federated learning
- Other measures for information security
  e.g., Access control

・ Physical control
  e.g., On-site facilities for microdata analysis

Methodologies for Privacy-Specific Functionality beyond Information Security

Business practices or IT system designs implementing functionalities specific for privacy protection, e.g., notification and obtaining consent, response to demand for disclosure, correction, and deletion of personal information, opt-in, opt-out expanded to include *the right to be forgotten*[9] or *the right not to be tracked.*[10, 11] The *personal data ecosystem* was proposed by Cavoukian (2012, 2013). In the ecosystem, privacy is defined as the right to know the purpose for which personal data is used, as well as the right to permit or deny the use of information about oneself on the basis of that purpose.

1. The OECD guidelines

The Organisation for Economic Co-operation and Development (OECD) Privacy Guidelines, established in 1980, inherently incorporate the concept of the right to control over personal information about oneself. The guidelines provide common principles that serve as the basis of domestic or regional privacy laws in Japan and other OECD member countries. OECD revised the guidelines in 2013 in response to the growing Internet economy and the increased cross-border flows of personal information (OECD [2013]).

The OECD guidelines define various general principles for privacy protection. For example, the Collection Limitation Principle limits collection of personal data and requires data controllers to obtain consent from data subjects (Paragraph 7). The Purpose Specification Principle requires them to specify the purpose of data collection and data usage in advance (Paragraph 9). The Data Quality Principle requires them to keep collected data accurate, complete, and up-to-date (Paragraph 8). The Security Safeguards Principle requires personal data to be protected with reasonable security measures (Paragraph 11). The Individual Participation Principle grants individuals the right to request data controllers to disclose, correct, or delete personal data (Paragraph 13). In addition, the OECD guidelines state that the amount of information collected and held should be the minimum necessary for achieving a specific purpose. This is called the *data minimization* principle. The 2013 amendments call for the establishment of privacy enforcement authorities who enforce privacy policies through international cooperation and develop national privacy strategies that reflect a coordinated approach.

2. Privacy by design

Privacy by design is a concept proposed by Cavoukian (2011) out of the concern that PETs and regulations by laws alone are not sufficient for privacy protection. This stems from idea that individuals have little control over their personal information stored in a huge database. Without appropriate privacy safeguards in place, an incentive for an individual to provide personal information would be lost, and the society would not gain the benefits of utilizing personal information. The concept of privacy by design underlies privacy protection systems in the EU and the United States.

Privacy by design consists of the following seven principles. Principle 1 is to take proactive rather than reactive measures for privacy protection. Principle 2 is to make privacy protection a default requirement of IT systems that process personal information. Principle 3 is to embed privacy protection measures into the architecture of IT systems and business practices by design. Principle 4 is to provide privacy protection

[9] The right to be forgotten is the right to have private information be removed from Internet searches.

[10] The right not to be tracked is the right to prevent third-parties from tracking a person by continuously collecting information about oneself.

[11] See Ishii (2014) for a discussion on the relationship with the legal systems in the European Union (EU) and the United States.

that benefits both service providers and users. Principle 5 is to ensure privacy protection throughout the lifecycle of personal information. Principle 6 is to keep privacy protection mechanisms visible and transparent. Principle 7 is to keep the mechanisms user-centric.

B. Rules

The rules of privacy protection include domestic laws, international arrangements, international standards, and self-regulation by industry organizations. This section focuses on the main domestic laws.

In the EU, the General Data Protection Regulation (GDPR) was legislated in accordance with the OECD guidelines. The right to privacy is recognized as one of fundamental human rights.

The California Consumer Privacy Act was enacted in the United States. Its basic concept is to ensure the consumers' right to privacy as stated in the Constitution.

In Japan, laws related to privacy protection include the Civil Code and the Act on the Protection of Personal Information.[12] The right to privacy is protected as a moral right, based on the right to the pursuit of happiness as stated in Article 13 of the Constitution. It is commonly recognized that the right to privacy corresponds to control over personal information about oneself (Sogabe, Hayashi, and Kurita [2019]).

The Act on the Protection of Personal Information was enacted mostly in accordance with the OECD guidelines. A purpose of the Act is to protect the rights and interests of individuals (Article 1), including both personal ones such as privacy and proprietary ones. However, the right to privacy and the right to control over personal information about oneself are not stated in the Act. One of reasons of this treatment is a lack of consensus regarding concepts of the rights, according to Sogabe, Hayashi, and Kurita (2019). In 2017, *Anonymously Processed Information* was introduced as a new data category. Data of this category are generated from personal information in such a way to prevent attackers from re-identifying the corresponding individual or restoring personal information. The attackers are assumed to use common methods for re-identification or restoration rather than advanced ones.[13] Requirements for third-party distribution of anonymously processed information have been eased more than those for personal information.

C. Methodologies for Privacy Protection

1. Information security

In the narrow context of information security,[14] privacy can be defined as the prevention of alteration or leakage of sensitive personal information. In general, methodolo-

[12] In addition, the Penal Code details penalties for defamation (Article 230) and insult (Article 231).

[13] Re-identification using advanced attack methods cannot be ruled out in terms of the anonymously processed information. Thus, privacy may be invaded if such information is exposed publicly. In the context of machine learning, personal information as training data can be inferred from the parameters of machine learning models, as described in Section V. D. Trained machine learning models are not regarded as personal information under the Act on the Protection of Personal Information.

[14] The three principles of information security are to ensure the integrity, confidentiality, and availability of information. Preventing the leakage of sensitive personal information corresponds to confidentiality. Ensuring the accuracy and consistency of sensitive personal information corresponds to integrity.

gies for ensuring information security consist of information and physical controls (see Figure 1).

a. Information control

This paper divides information control into mathematical, cryptologic, and machine learning methodologies.

Mathematical methodologies include mechanisms that satisfy differential privacy and statistical disclosure control. These mechanisms are mainly used in the fields of statistics and data mining. Due to the nature of statistical and probabilistic techniques, they mostly involve information loss for privacy protection. Mathematical methodologies will be discussed in detail in Section III.

Cryptologic methodologies provide anonymity and confidentiality of information through encryption.[15] Information is not degraded and completely preserved because it can be recovered by transforming encrypted data with the corresponding decryption key. This category includes cryptographic methodologies for secure communication protocols over the Internet, blind signatures, and zero-knowledge proofs such as the zero-knowledge succinct non-interactive argument of knowledge, or zk-SNARKs (Gennaro *et al.* [2013]). It also includes secure computation[16] and secret sharing.[17] Many research studies (Ramacher, Slamanig, and Weninger [2021], Lian *et al.* [2021]) have considered privacy protection as an additional goal. Modifications to Transport Layer Security (TLS), a standard secure protocol commonly used on the Internet, have been proposed. For instance, TLS Encrypted Client Hello[18] and Encrypted Server Name Indication[19] conceal a part of the transmission data (Client Hello and Server Name) to prevent entities involved in a TLS session from being identified. Thus, the TLS standardization takes into account confidentiality, traceability, and linkability[20] as additional features for privacy protection.

Machine learning methodologies combine training of machine learning models with privacy protection. For instance, federated learning (McMahan [2017]) enables multiple companies to build a single model using their databases without disclosing

[15] Even though personal information is encrypted, its status would be the same as that of personal information under the Act on the Protection of Personal Information.

[16] Secure computation is a set of techniques for processing data without disclosing them, e.g., by using advanced encryption schemes. For instance, secure computation enables users to encrypt their own confidential data separately, share the encrypted data with each other, and obtain statistics without disclosing confidential information. The core mathematical trick is homomorphic encryption, in which an outcome of arithmetic operations of plain data is equal to a decrypted outcome of the corresponding operations of the encrypted ones.

[17] Secret sharing enables storing a single personal-information database on multiple separate servers in a dispersed and encrypted manner. Recovering the database requires the cooperation of multiple servers whose number is equal to or more than a threshold number. Any one of the servers cannot recover the database by itself. Secret sharing can relax the assumption of trust in the central database holder in security analysis.

[18] The IETF draft for TLS Encrypted Client Hello is available at https://www.ietf.org/archive/id/draft-ietf-tls-esni-14.txt (expired on August 17, 2022).

[19] The IETF draft for Encrypted Server Name Indication for TLS 1.3 is available at https://www.ietf.org/archive/id/draft-rescorla-tls-esni-00.txt (expired on January 3, 2019).

[20] The confidentiality of personal information can be conceptually divided into the following two: the confidentiality of its contents and that of the corresponding subject. Linkability is the property in which a specific (but anonymous) subject cannot be identified even when the corresponding contents are known. Traceability is defined as the property in which multiple anonymized records can be associated to a specific (but anonymous) person. If a certain database meets the linkability, it also meets the traceability.

confidential information to each other. The main feature of this procedure is its decentralized manner. Namely, a model is trained without consolidating the databases into a single one. However, there is still a risk of information leakage from update information of machine learning model (e.g., gradients). The update information is transmitted from each database. To mitigate this risk, a method to prevent unauthorized entities from eavesdropping on the information has been proposed by Bonawitz *et al.* (2017).

Information control measures also include fixing vulnerable source codes and assigning appropriate permission to access databases of personal information. Such ordinary measures are always required whenever IT systems are operated. When only a limited number of users are supposed to access a database, such as for internal use within a company, access permission should only be granted to the users.

b. Physical control

Physical control regulates the environment in which personal information is handled. For example, *on-site facilities*[21] for official statistics in Japan allow researchers access to individual data (microdata) within a physically isolated space and under supervision through cameras. Purposes of the use of the database is also limited to public interests such as academic research. Under such strict physical control, systems for secondary use of official statistical databases have also been established.[22]

2. Privacy-specific functionality beyond information security

When privacy is defined as the right to control over personal information about oneself, business practices and IT system designs are also essential to implement such control. This is because information security measures cannot cover several objectives relating to control over personal information about oneself. For example, such objectives include preserving the right to demand disclosure, correction, and deletion or addition of personal information. The opt-out or opt-in scheme is another objective for achieving privacy protection. These should be systematically embedded into IT systems and business practices in accordance with the principle of privacy by design.

D. Social Demand for Privacy

Along with the development of the privacy protection framework, the use of personal data is expanding as well, as described in Section I. This leads to the more significant threat of privacy invasion. The social demand for privacy and security has also increased, particularly for precautionary measures. It is nearly impossible to compensate for damages caused by privacy invasion. Moreover, the malicious use of leaked information is difficult to recognize.

The social demand for privacy protection cannot be met solely by technology due to its ambiguous nature, as privacy cannot be recognized until it is invaded. The damage from privacy invasion has subjective aspects that differ from person to person, and the

[21] On-site facilities are located in government agencies and research institutions including universities. See below for a list of facilities. https://www.e-stat.go.jp/microdata/data-use/on-site-facilities (in Japanese)

[22] In Japan, the authority of official statistics aims to promote the utilization of microdata obtained from statistical surveys. Systems for secondary use of personal data have been developed. The secondary uses are for tailor-made tabulation, anonymized data, and microdata. The advanced use of microdata opens the door to the analysis of consolidated multiple microdata. The analysis must be conducted in a physically isolated location called an on-site facility under camera supervision.

impacts of the damage on an individual are also situation dependent.[23]

Defining privacy and setting the goals of its protection should be done through a process of forming social consensus, overcoming the diversity of individual subjectivities and values. The privacy protection framework shown in Figure 1 is expected to be updated to reflect the deeper use of personal information and the social consensus of the times. Privacy protection as a social issue can also be discussed in the context of ELSI brought about by new technologies, as described in Section VI. C.

III. Mathematical Methodologies for Privacy Protection

Privacy protection has been discussed across multiple disciplines including statistics, database, cryptology, and information theory. Many studies refer to an associated set of techniques collectively as privacy preserving data mining. Several methodologies (Agrawal and Srikant [2000]) have been studied to protect the privacy of information processed with advanced methods such as trained machine-learning models. This section briefly describes mathematical methodologies, as categorized in Section II. These methodologies protect privacy by reducing or degrading (the contribution of) personal information.

A. Classification of Mathematical Methodologies

Formal mathematical methodologies can be classified from three perspectives: (1) applicable stages, (2) scope of privacy protection, and (3) principles of protection (Figure 2 and Table 1).

These methodologies are applicable in three stages. Figure 2 shows a flow of analysis of personal information: (a) Personal data about individual x is gleaned and (b) stored in database D. A user sends a query to D and (c) D outputs statistics in response to the query. Approaches that process x at stage (a) do not need to assume that an owner of D is trusted. Approaches that process D itself at stage (b) are useful when D is distributed to or shared with a third party. There are also approaches that process outcomes from D at stage (c).

B. Introduction of Mathematical Methodologies

This section presents each mathematical methodology roughly in the order shown in Table 1.

1. Methods satisfying local differential privacy

Methods based on *local differential privacy* make it difficult to infer personal information from the corresponding personal data by randomizing the information before sending it to D. Local differential privacy shares the same core property with differential privacy: it guarantees privacy protection against any attacker with arbitrary background knowledge. In addition, it does not need to assume that an owner of database is trusted. For more details on differential privacy, see B. 5 and Section IV, and for details on local differential privacy, see Section IV. E.

[23] For example, information that a person has borrowed money from illegal finance for gambling purposes may cause serious harm when he or she is seeking a job.

Figure 2 Flow of Personal Data Analysis

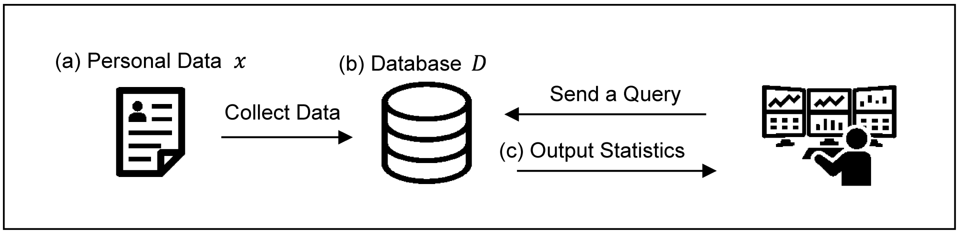


Table 1 Mathematical Methodologies and Principles of Protection

	Applicable Stage	Scope of Protection	Principles of Protection
Methods Satisfying Local Differential Privacy	(a)	Owner of Database, User of Database	Personal information gained from outcomes is limited (probabilistic)
Anonymization/ Pseudonymization	(b)	User of Database	Impossibility of identification from pseudonym (deterministic)
k-Anonymization	(b)	User of Database	Indistinguishability of k individuals who shares an identical ID
Synthetic Data	(b)	User of Database	Difficult to restore the original data from synthetic data generated from probability distribution.
Statistical Disclosure Control	(c)	User of Database	Suppression/Concealment of cells (deterministic)
Methods Satisfying Differential Privacy	(c)	User of Database	Personal information gained from outcomes is limited (probabilistic)
Random Sampling	(c)	User of Database	Impossibility of identifying which data are used for calculating outcomes (probabilistic)

Note: Scope of Protection indicates the range in which people cannot gain the accurate personal data. User of Database indicates people who can access to the outcomes generated from databases.

2. Anonymization and pseudonymization

Anonymization is a procedure to remove information that directly identifies an individual (personal ID, quasi-identifier) from personal data. A *personal ID* denotes an attribute that identifies an individual without any additional information. For example, the Individual Number corresponds to his/her personal ID. A *quasi-identifier* denotes a tuple of attributes that identifies an individual.[24]

Pseudonymization is a procedure to replace a personal ID with a pseudorandom number (pseudonym). A pseudonym is generated by using a random number generator or hashing the personal ID. In cases where an individual contributes to multiple data records in a database, anonymity can be improved by assigning different pseudonyms for each record (multiple pseudonymization).

3. *k*-anonymization

k-anonymization (Sweeney [2002]) is a methodology to transform quasi-identifiers such that any individual is hidden among k-1 people whose quasi-identifiers are identi-

[24] For example, the combination of four basic personal attributes (name, gender, date of birth, and address) can identify an individual with a high degree of accuracy.

cal.[25] The property that k or more people are assigned the same quasi-identifier is called *k-anonymity*. This property has a risk of exposing attributes. To illustrate this risk with an example of a medical database, suppose that there are k or more individuals with the same quasi-identifier, and all of them turn out to have digestive diseases. Even if each of them cannot be distinguished by using the quasi-identifier, such an attribute is exposed.

l-diversity (Machanavajjhala *et al.* [2007]) is a property that is designed to eliminate this risk by ensuring that k individuals who share the same quasi-identifier have l or more different attribute values. Returning to the previous example, quasi-identifiers are transformed such that there are l or more types of diseases for the k individuals. This additional constraint makes it infeasible to narrow down the diseases of the individuals.

Even if l-diversity is satisfied, the risk of exposing attributes may remain in cases where each distribution of attribute values is similar.[26] k-anonymization based on *t-closeness* (Li, Li, and Venkatasubramanian [2007]) is a refinement of l-diversity that considers the proximity of distributions of attribute values.

While k-anonymization is easy to implement, the database becomes less useful when attribute information is eliminated. The method of evaluating the effectiveness of k-anonymization depends on attack models. Furthermore, finding optimal k-anonymization is known to be an NP-hard problem (a class of computational problems that are believed to be difficult to efficiently find solutions). Optimal k-anonymization means that the number of transformed quasi-identifiers is minimized (Meyerson and Williams [2004]).

4. Statistical disclosure control

Statistical disclosure control (Hundepool *et al.* [2010, 2012]) generally denotes methodologies that transform aggregate tables or microdata in a way that avoids disclosing personal information. In the history of research on privacy protection, statistical disclosure control has been studied in the fields of statistics and databases since around the 1980s.

Statistical disclosure control conceals a part of tabulated data before publication. It has been widely adopted in domestic and international official statistics in order to prevent privacy exposure. For example, a n–k dominance rule of cell suppression is one of widely used methods based on statistical disclosure control. The rule conceals sensitive aggregate values in which n individuals contribute k% or more of the total value. However, as with k-anonymity, it is necessary to assume a specific attack model in order to evaluate the effectiveness of statistical disclosure control.

[25] A quasi-identifier can be transformed by generalization or record deletion. Generalization reduces information in the quasi-identifier. For example, extracting a birth year (1990) from a birth date (March 14, 1990) belongs to such a method. Record deletion is to delete the data record of an individual. This is done when generalization alone cannot generate quasi-identifiers that satisfy k-anonymity.

[26] Suppose that the annual income ranges of k individuals who share the same transformed identifier are either USD 10,000–11,000 or USD 11,000–12,000. In this case, the annual income of any individual will be distributed in a band of USD 10,000–12,000. However, if the ranges are either USD 10,000–11,000 or USD 60,000–61,000, the band becomes much wider, i.e., USD 10,000–61,000, due to the divergence of the annual income ranges.

5. Methods satisfying differential privacy

Differential privacy provides quantitative criteria for evaluating privacy-preserving techniques on the basis of information-theoretic security. The concept of differential privacy was formulated with reference to cryptologic methodologies (Dwork *et al.* [2006]). In related research, differential privacy is positioned as the development of *inference control* combined with information theory (see Section IV. B for the relationship to cryptology). Inference control aims to prevent the inference of confidential information from responses to arbitrary queries (Iwamura and Nishijima [1991], Denning and Denning [1979], Denning [1980, 1982], Beck [1980]).[27]

Suppose a probabilistic method outputs a random variable from a database. If the method satisfies a property based on differential privacy, the corresponding privacy can be achieved against any attackers who attempt to infer personal information from outputs of the database by using arbitrary background knowledge. Section IV discusses differential privacy in detail.

6. Random sampling

Random sampling (see Adam and Wortmann [1989] for example) is a traditional approach that randomly selects and extracts records from a database of personal information. Only sampled records are used for calculating statistics. In this approach, an attacker cannot ascertain the use of a particular individual's records in the process of generating outcomes from the database.

7. Synthetic data

Synthetic data is an approach to generate a new database that preserves the statistical properties of the original one. The owner of the original database publishes the synthetic data instead of the original one. The advantage of this approach is that a user of the database can apply methodologies for the original data directly to the synthetic data. In general, synthetic data is generated by using the estimated parameters of a probability distribution that reflects the statistical characteristics of the original data. The effectiveness of synthetic data can be measured with the difficulty of inferring the original data from the synthetic data. The risk of information leakage via parameters is likely to be low; however, extreme values in the original database may be inferred from the synthetic one. Thus, synthetic data satisfying differential privacy has been actively studied. In addition, it is difficult to capture statistical properties of individual data for diversified applications because they are inherently multifaceted.

8. Combination of methodologies

The methodologies previously introduced in this sections can be combined. Li, Qardaji, and Su (2012) extended the definition of differential privacy and proved that *k*-anonymization combined with random sampling satisfies the extended differential privacy.

Another approach combines anonymization with federated learning to assign a group ID to each group of individuals. Google researchers proposed a method called *federated learning of cohorts* (FLoC), which utilizes federated learning (see Section II. C. 1) to group individuals with similar attributes and to assign a group ID for each

[27] See Igarashi and Takahashi (2012) for the history of differential privacy in the field of privacy-preserving data mining.

group.[28] FLoC was developed as a possible alternative to third-party cookies that collect web browsing history for targeted advertising. There has been a growing movement to restrict the use of third-party cookies on the grounds that excessive user tracking for targeted advertising may violate their privacy.

IV. Theory of Differential Privacy

This section introduces the theoretical foundations of differential privacy. For more details, see Dwork and Roth (2014), Dwork (2008), and Sakuma (2016).

A. Importance of Privacy Protection Based on Information-Theoretic Security

Differential privacy guarantees privacy protection based on information-theoretic security, which is becoming increasingly important due to the following.

First, attackers have progressively been able to use more and more personal information. Even if each piece of personal information in public circulation is of little use, such information can be comprehensively compiled into a single database through deduplication. The availability of the deduplicated database to an attacker is a threat in itself. Nevertheless, privacy-preserving technologies, such as anonymization which does not guarantee privacy protection based on information-theoretic security, may not be sufficient to protect a database of personal information against re-identification and exposure of confidential values.[29] There has also been an increase in the amount of external information (background knowledge) and published statistics from personal information databases. Furthermore, as relational databases that respond to arbitrary SQL (Structured Query Language) queries become more common, the advanced use of databases, such as tailored aggregation to meet the statistical demand, has expanded. Even if each statistic or return value from the databases is individually harmless, their combination may enable an attacker to expose a part of the corresponding personal information. There are too many possible combinations to evaluate the risk for all cases.

Second, attackers' computational power has been increasing. A database reconstruction attack[30] had been considered difficult to execute in reality due to the limitation of computational power. However, as computational power has developed, the attack has shifted from a theoretical risk to a practical issue which requires countermeasures. This motivated the implementation of privacy protection based on differential privacy in the U.S. Census Bureau.[31]

Third, it is difficult to anticipate and prepare for all potential threats. An official bureau of statistics usually publishes outcomes from personal information databases in a tabular format. Even if each table is secure in terms of privacy, the combination of multiple tables can increase the risk of personal information exposure. Experts have

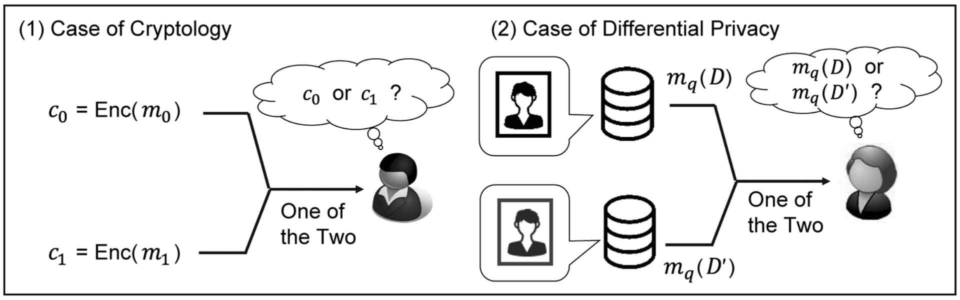
[28] For details, see the project page. https://github.com/WICG/floc

[29] For example, suppose that an attacker has access to an anonymized database that records the time and location of individuals. If the attacker gains new information that a target individual was witnessed at a particular location and time, the attacker may be able to find his/her record in the database.

[30] This attack is an attempt to reveal confidential microdata in a personal information database using statistical tabulations as constraints, similar to arithmetical restorations or *Sudoku* puzzles.

[31] The U.S. Bureau of the Census released a leaflet, "A History of Census Privacy Protections," in October 2019, which states that the 2010 Census was the "last census to use ad-hoc privacy protections."

Figure 3 Indistinguishability in Cryptology and Differential Privacy



carefully and manually controlled this risk through statistical disclosure control by considering as many information exposure attacks as possible (Hundepool *et al*. [2010]). In principle, however, this approach cannot prepare for unknown attacks at the time of publication.

In general, methodologies such as pseudonymization and anonymization cannot guarantee privacy protection because it is difficult to incorporate all background knowledge into analyses. In fact, k-anonymization does not satisfy differential privacy. The security of each methodology is dependent on the attacker's specific model.

In contrast, the protection from methodologies based on differential privacy is effective against attackers with any background knowledge. These methodologies can address the database reconstruction attack which is of concern to the U.S. Census Bureau, as well as unknown attacks that may be discovered in the future. Thus, differential privacy is ideal for precautionary measures. However, it is not suitable for all situations; its advantages are demonstrated typically in situations where a database responds to arbitrary statistical queries. For example, differential privacy is not necessary when only statistical tables are regularly published in a fixed format.

B. Definition of Differential Privacy

Differential privacy borrows the concept of indistinguishability in cryptology and is considered a formal definition of a degree of privacy protection in academia. In cryptology, given encryption function Enc and two plaintexts m_0 and m_1, if it is (computationally) difficult for an attacker to distinguish their ciphertexts $c_0 = \text{Enc}(m_0)$ and $c_1 = \text{Enc}(m_1)$, the corresponding cryptosystem is considered secure.[32]

Differential privacy is defined as indistinguishability between two probability distributions (see Figure 3). Suppose we have two databases D and D' that differ only in a database record about individual x. If it is difficult to distinguish probability distributions $m_q(D)$ and $m_q(D')$ of output values obtained from D and D', respectively, information about x obtained from these values is negligible. Namely, it is difficult to extract information about x from the difference between D and D'.

The term *differential* in differential privacy originates from the idea that one database contains data of an individual and the other does not. The exact definition is as follows. Assume that database D consists of records. Each record corresponds to

[32] More specifically, the security is mathematically formulated as an interactive game between an adversary and a challenger. A typical example is indistinguishability under chosen-ciphertext attack (IND-CCA).

data for one individual. The theory starts with the definition of the adjacency in D.

Definition: Two databases of personal information D and D' are *adjacent* if they differ by at most one individual's data. The relation of adjacency is expressed as $D \sim D'$.

Query q requires statistical values from D. Statistical values $q(D)$ represent a response to q from D. Map m_q is a randomization mechanism (hereafter, simply *mechanism*) that outputs stochastic values depending on q. Typically, outcomes of m_q, denoted as $m_q(D)$, are defined as stochastic values obtained by adding stochastic noises to exact statistical values $q(D)$.

Definition: Given q and a positive constant ϵ, mechanism $m_q \colon \mathcal{D}^n \rightarrow \mathcal{R}$ satisfies *ϵ-differential privacy* if inequality

$$\Pr[m_q(D) \in S] \leq \exp(\epsilon) \times \Pr[m_q(D') \in S],$$

is satisfied for all pairs of adjacent databases $D \sim D'$ and all sets of statistics $S \subseteq \mathcal{R}$. $\mathcal{R}$ denotes the range of mechanism.

The above definition is a strict formulation of the indistinguishability (or ϵ-indistinguishability) between probability distributions $m_q(D)$ and $m_q(D')$. The interpretation of the inequality by ϵ is as follows:

- When $\epsilon = 0$, $\exp(\epsilon) = 1$ holds in the right hand side and the probability distributions of $m_q(D)$ and $m_q(D')$ are perfectly identical. The statistics do not depend on D, hence privacy is fully protected (perfect confidentiality), but the output value is statistically meaningless.
- When $\epsilon = +\infty$, $\exp(\epsilon) = +\infty$ holds, and this enables the probability distribution of the statistics to change infinitely. Although the utility of the statistics is maximized, privacy protection is not guaranteed at all, i.e., partial data about an individual can be definitively inferred from the statistics.
- When $0 < \epsilon < +\infty$, a moderate level of privacy protection is achieved; a smaller value of ϵ provides a higher level of privacy protection and decreases the utility of the statistics.

Instead of ϵ-differential privacy, (ϵ, δ)-differential privacy is used as a relaxed version of the original differential privacy (Dwork *et al.* [2006]).

$$\Pr\left[m_q(D) \in S\right] \leq \exp(\epsilon) \times \Pr\left[m_q(D') \in S\right] + \delta.$$

This definition roughly means that ϵ-differential privacy can be violated with probability δ.

Differential privacy is a conservative and strong standard based on the assessment of the worst-case scenario, i.e., the largest amount of information leakage among all changes in an arbitrary single record. The assessment does not assume any attack models. One of the key characteristics of differential privacy appears in the post-processing theorem, which claims that the privacy protection level does not deteriorate by computing any function values from the output of a mechanism that satisfies differential privacy. Note that the function does not contain any additional information about the original database.

Post-processing theorem (Dwork and Roth [2014]): If mechanism $m_q : \mathcal{D}^n \rightarrow \mathcal{R}$ satisfies (ϵ, δ)-differential privacy, then for any map $f : \mathcal{R} \rightarrow \mathcal{R}'$, composite map $f \circ m_q$ also satisfies (ϵ, δ)-differential privacy.

Moreover, $u\epsilon$-differential privacy is guaranteed for a group of u distinct records if ϵ-differential privacy is satisfied. In other words, differential privacy is guaranteed for a set of data with size u. This property is called *group privacy*. This implies that the guaranteed level of privacy protection becomes weaker when database records are more strongly correlated. A theoretical proof of differential privacy generally assumes that each database record is independently generated. However, actual record-generating processes and probability distributions followed by the records are not always known. If the actual distributions are different from those assumed in advance, the level of privacy protection may lower than theoretically expected. In practical databases, the same individual can contribute to multiple records, which results in a strong correlation between the records.[33] Attention should be paid to the risk that actual privacy protection may be weaker than the proven one.

C. Appropriate Level of ϵ

The constant ϵ indicates the strength of privacy protection, where a smaller ϵ means stronger privacy protection. The value of ϵ must be exogenously determined by a database provider. In practice, ϵ is typically set to about 0.1 to a single digit number. However, there is no clear consensus on an appropriate level of ϵ for practical privacy protection thus far, so it is ultimately the policy maker's choice.[34]

The privacy protection guaranteed by differential privacy depends on the definition of database adjacency. Metaphorically speaking, differential privacy provides a ruler that measures the strength of privacy protection, and its scale is determined by the adjacency. The value of ϵ indicates the length measured by the ruler. Thus, even if ϵ is identical, the sense of privacy protection may differ depending on the adjacency. Given $D \sim D'$ that differ only by records $x \in D$ and $x' \in D'$, the sense of privacy protection depends on how x differs from x' (even if ϵ is identical). Dwork *et al.* (2006) defined adjacency as the difference between D and D' such that D contains x whereas D' does not. In this case, adjacency means the membership of an individual in a database. However, membership exposure (whether or not the individual's data are included) may not pose a risk in certain cases, depending on the nature of databases of personal information. In such cases, concealing the membership can be an excessive goal of privacy protection. Adjacency should be defined appropriately for the purpose of applications that utilize databases as the meaning of ϵ varies with the definition of the adjacency.

The composition theorem states that any combination of mechanisms that satisfy differential privacy also satisfy differential privacy.[35]

Composition theorem: Suppose that mechanisms $m_1, m_2, \ldots, m_k$ satisfy $\epsilon_1, \epsilon_2, \ldots,$

[33] For example, when search history data of the same person are collected separately from a tablet and a PC, these sets of data are likely to strongly correlate.

[34] In economics, a study has suggested an approach for determining an appropriate level of ϵ through economic optimization. For details, see Abowd and Schmutte (2019).

[35] For instance, a pair of mechanisms can be regarded as a single mechanism that outputs a pair of statistics. If each mechanism in the pair satisfies differential privacy, the combined mechanism also satisfies differential privacy.

ϵ_k-differential privacy, respectively. Then, the combination of these mechanisms also satisfies $(\sum_{1\leq i\leq k}\epsilon_i)$- differential privacy.

This theorem also holds for (ϵ, δ)-differential privacy. This theorem indicates that, as the number of queries increases, ϵ is accumulated by each query, which means that privacy protection weakens. Due to this property, differential privacy requires setting an upper limit on the number of queries in accordance with ϵ, which is determined uniquely for a database. Thus, ϵ is also called the *privacy budget*. This is the acceptable amount of information leaked from outputs of queries, which is allocated to each query similarly to budget control.

D. Mechanism Design

This section describes how to design m such that it satisfies differential privacy, given ϵ and q.

An effective mechanism yields more useful statistics while providing stronger protection. The usefulness of a mechanism can be measured as the similarity between $m_q(D)$ and $q(D)$ in the form of a probability bound as follows,

$$\Pr\left[\left\|q(D) - m_q(D)\right\| > g(n)\right] \leq \beta.$$

Function $g(n)$ denotes how fast an output of the mechanism converges to an exact statistic as database size n increases. The similarity above is referred to as *utility*. Noticeably, the concept of utility depends on q. In general, there is a trade-off between the usefulness of the statistics and the strength of privacy protection. Designing a more useful mechanism with higher utility can take into account the specific nature of each q, but such a design is not versatile.

A modular approach is widely adopted to address this problem, in which a mechanism for advanced statistical analysis is created by combining ones for simple analysis. The composition theorem described above justifies this approach which reduces the design of a set of mechanisms for complicated analysis to that for a single query. In the following, we discuss the mechanism design for a single query.

1. Mechanisms based on global sensitivity

a. Definition of global sensitivity

Generic methodologies with global sensitivity make it possible to design differentially private mechanisms with arbitrary statistical queries. Global sensitivity is defined as the maximum (worst case) change of $q(D)$ in response to a change in a single record.

Definition: *Global sensitivity* of query $q\colon \mathcal{D}^n \to \mathbb{R}^d$ is

$$\mathrm{GS}_q = \max_{D, D'\colon D\sim D'} \left\|q(D) - q(D')\right\|.$$

Where d denotes a dimension of an outcome of q. A norm represents ℓ_1 and ℓ_2 norms in the Laplace mechanism (described in the following) and Gaussian mechanisms, respectively. In the following, the definition of norms is omitted for simplicity.

Global sensitivity is independent of the values in database records. Intuitively, from the perspective of privacy protection, a larger noise should be added to an outcome with a larger global sensitivity. For example, given a database in which the value range of each data is normalized to $[0, 1]$, the global sensitivity of a query that returns the average is $1/N$. Here, N denotes the number of database records. The global sensitivity of a query that returns the maximal value is 1. Thus, a mechanism should add a larger noise to the maximal value than to the average value. This magnitude of noises is consistent with the intuition that the exact maximal value of the records is more likely than the average value to expose personal information. Global sensitivity can be regarded as a measure of how cautious one should be about disclosing outcomes of a query.

b. Mechanism design with global sensitivity

Typically, a mechanism designed with global sensitivity adds stochastic noises from a specific probability distribution.

The *Laplace mechanism* (Dwork and Roth [2014]) adds noises following the Laplace distribution[36] to $q(D)$: $m_q(D) = q(D) + \text{Laplace}(\text{GS}_q/\epsilon)$. The magnitude (standard deviation) of noises is proportional to the global sensitivity. The Laplace mechanism has been proven to satisfy differential privacy.

Theorem: Given query q, the Laplace mechanism that adds noises of which the probability distribution follows $\text{Laplace}(\text{GS}_\text{q}/\epsilon)$ satisfies ϵ-differential privacy.

The Laplace mechanism is easy to implement and widely used today. The *Gaussian mechanism*, where noises follow a Gaussian distribution, also satisfies differential privacy. Specifically, a mechanism $m_q(D) = q(D) + N(0, \text{GS}_q^2 \cdot \sigma^2)$ satisfies (ϵ, δ)-differential privacy under certain parameter conditions (Dwork and Roth [2014]). The *exponential mechanism* (McSherry and Talwar [2007]) probabilistically returns an element that maximizes utility score $q\colon \mathcal{D}^n \times \mathcal{A} \to \mathbb{R}$.[37] The probability that element $a \in \mathcal{A}$ is chosen depends on the global sensitivity of q. Specifically, $\Pr[a] \sim \exp(\epsilon \cdot q(D, a)/2 \cdot \text{GS}_q)$, where ~ denotes proportionality. This mechanism satisfies differential privacy. Mechanisms designed with global sensitivity tend to have a drawback that excessively large noises are added to statistics because the magnitude of noises is determined based on the worst-case change in the statistics in response to a change in a single record.

2. Definition of local sensitivity

Approaches for designing mechanisms with local sensitivity (Nissim, Raskhodnikova, and Smith [2007], Dwork and Lei [2009]) make it possible to add smaller noises than mechanisms based on global sensitivity do.

Definition (local sensitivity): given database $D \in \mathcal{D}^n$, *local sensitivity* of query $q\colon \mathcal{D}^n \to \mathbb{R}^d$ at D is

$$\text{LS}_q(D) = \max_{D'\colon D \sim D'} \left\| q(D) - q(D') \right\|.$$

[36] A probability density function of random variable $x \sim \text{Laplace(b)}$ is expressed as $f(x; b) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right)$. The variance is $2b^2$.

[37] The exponential mechanism is useful when the utility varies significantly in response to the fluctuations of data values. For example, given a set of bid prices {USD 10, USD 20, USD 50} that determines a demand curve, consider determining an optimal price that maximizes profit (= utility) while keeping a bid price secret for privacy protection. Consider that a noise is added to a price itself. Setting a price to USD 50 results in sales of USD 50, while setting it to USD 51 results in sales of USD 0. A significant decrease in profit occurs due to the tiny variation. By contrast, the exponential mechanism selects a price from a set *independent* of the bid price. In this case, privacy can be protected while maximizing profit with high probability.

By definition, local sensitivity is smaller than global sensitivity: $\mathrm{LS}_q(D) \leq \mathrm{GS}_q$ holds for any $D \in \mathcal{D}^n$. Because local sensitivity depends on D, there is a risk that information on D is leaked from the outcomes of a mechanism. In general, a mechanism that adds noises whose magnitude is proportional to local sensitivity does not satisfy differential privacy. Thus, designing a mechanism with local sensitivity requires preventing the leakage of information about D so that the mechanism satisfies differential privacy.

Johnson, Near, and Song (2018) proposed a mechanism named FLEX that defines a variable (*elastic sensitivity*) as an upper bound of local sensitivity and adds noises proportional to the elastic sensitivity to responses to practical SQL queries. This mechanism design can be applied to a wide range of SQL queries by reducing the computation of the elastic sensitivity for complicated queries to that for simpler ones.

This approach modifies the statistics gained from a database by adding noises but does not modify the query itself. Due to this design, it cannot be used for a query with clipping which requests a mean of data except for outliers. CHORUS (Johnson *et al.* [2020]) is an extension of the FLEX approach to overcome this limitation.

Nissim, Raskhodnikova, and Smith (2007) proposed a method that defines a variable (*smooth sensitivity*) as approximation of an upper bound of local sensitivity. The mechanism based on this method adds noises proportional to smooth sensitivity. They showed that smooth sensitivity can be computed quickly for some queries, such as one for the median.

It is not easy to design mechanisms using local sensitivity that not only satisfies differential privacy but also operates at high speed for a wide range of statistical analyses.

E. Local Differential Privacy

Differential privacy focuses on privacy protection when publishing statistics and assumes that the owners of databases of personal information are trusted. In contrast, *local differential privacy* (LDP; Duchi, Jordan, and Wainwright [2013]) focuses on privacy protection when collecting user data and does not assume trust in such owners. The definition of LDP is as follows.

Definition: Mechanism $m_q : \mathcal{D} \to \mathcal{R}$ satisfies *ϵ-local differential privacy* if for any pair of user data $v, v' \in \mathcal{D}$, and for any $S \subseteq \mathcal{R}$, the following inequality holds:

$$\Pr[m_q(v) \in S] \leq \exp(\epsilon) \times \Pr[m_q(v') \in S].$$

Mechanisms that satisfy LDP typically randomize user data in two ways (Yang *et al.* [2020]). One is to fluctuate user data by adding noises. The other is to replace user data with false ones, based on the randomized response. The former usually applies in cases where user data take continuous values. The noises follow the Laplace or Gaussian distribution in many cases. The latter applies in cases where user data take discrete values.

Randomized response was proposed by Warner (1965) as a survey technique to remove biases that arise from answers to sensitive questions such as "*Have you ever committed a crime in the past?*" This method flips true and false values with a certain probability so that statistics can be estimated accurately while leaving respondents with plausible deniability for their answers.

Consider the following example of a mechanism. A respondent tosses a coin in secret. If the coin comes up heads, the respondent answers honestly; if tails, the respondent tosses the coin again. If heads, the respondent always answers *yes*, and if tails, the respondent answers *no*. In this case, a simple calculation shows that local differential privacy of $\epsilon = \ln(3)$ is satisfied.[38] As is trivial from the structure of the mechanism, even if information that an individual has answered *yes* is leaked, the individual still has room to deny the possibility that the answer reflects the truth by claiming that the answer is only based on outcomes of the second coin flipping. While protecting privacy in this way, the survey conductor is able to statistically estimate an accurate proportion of affirmative respondents (whose honest answers are *yes*) from the randomized responses. This ability poses an assumption that respondents had followed the prescribed protocol faithfully.

Holohan, Leith, and Mason (2017) developed an optimal randomized response in terms of minimizing the error in the maximum likelihood estimator when the response takes a binary value as described above. When the response takes three or more values, the general randomized response (Kairouz, Bonawitz, and Ramage [2016]) is applicable.

V. Research for the Applications of Differential Privacy

A number of mechanisms that satisfy differential privacy have been proposed in the growing literature of applied research. This section introduces some of the methods and implementations in public statistics and private business.

A. Application for Aggregate Tables

Population data from the national census and population and human flow data based on smartphone locations are aggregated in hierarchical geographic units.[39] The U.S. Census Bureau also provides a breakdown by race, gender, ethnicity, and other attributes for each geographic unit. Differential privacy can be applied to these aggregations by randomizing responses to counting queries (queries that require a database to return the number of records that meet certain criteria).

1. TopDown algorithm

A naive approach to aggregating population in hierarchical geographic units is to apply the Laplace mechanism to randomize the aggregate population of the finest units. Then this population is aggregated from the lower levels (narrower geographic units) to the upper levels (wider geographic units). However, this method has three drawbacks. First, the method may output a negative aggregate population, which should not be negative. Second, the accumulated noises degrade the accuracy of the aggregated population in wider geographical units. Third, the aggregated population at the national and state

[38] Consider conditional probability Pr [Answered value|True value]. Pr [Yes|Yes] $= 1/2 + 1/2 \times 1/2 = 3/4$, Pr [Yes|No] $= 1/2 \times 1/2 = 1/4$. The next follows: Pr [Yes|Yes] / Pr[Yes|No] = 3. Similarly, Pr [No|No] / Pr [No|Yes] = 3. These results prove that ln(3)-local differential privacy is satisfied.

[39] The U.S. Census provides aggregated populations in hierarchical geographic units that are defined from the nation, state and county to the most granular census block levels.

levels is inconsistent with the accurate figures published by the U.S. Census Bureau.

The U.S. Census Bureau adopted the TopDown algorithm (Abowd *et al*. [2019]) to resolve or mitigate these drawbacks. This algorithm recursively subdivides the aggregated population from the top level to the bottom. At each hierarchical level, the aggregate values that are randomized by the Laplace mechanism are modified by replacing them with the solution of an integer programming problem with constraints. The constraints includes consistency with publicly known facts (such as the exact population of each state and the entire nation), non-negative sign restriction of the population, the arithmetic equality of each elements and their totals, etc.

The aggregate tables as a whole satisfy ϵ/h-differential privacy by allocating the privacy budget of ϵ to each level of the h-tier hierarchical geographic units. This method is advantageous in cases where the accurate populations of certain areas such as each state or the entire nation are published and the census and the facts need to be consistent while balancing the accuracy of the statistics and privacy protection. However, the accuracy of this method can be improved by using the Privelet method described in Section V. A. 2.[40]

2. Privelet method

Xiao, Wang, and Gehrke (2010) proposed a privacy-preserving wavelet named *Privelet* that improves the accuracy of subtotals and totals of population by combining discrete wavelet transformation[41] with differential privacy.

Privelet first applies wavelet transformation to population tables. Then, the wavelet coefficients are randomized by the Laplace mechanism and transformed back with the inverse wavelet transformation. Compared with the methods that directly randomize the original raw data, Privelet satisfies differential privacy with smaller noises, making the output statistics more useful. The variance of the added noises is $\mathrm{O}((\log_2(V))^3/\epsilon^2)$ in Privelet for the aggregate value V and privacy budget ϵ, whereas it is $\mathrm{O}(V/\epsilon^2)$ for the direct method.

The advantage of Privelet is that it satisfies differential privacy and improves the accuracy of subtotals and totals. Conversely, its disadvantage is that the response to a counting query may still be negative, and the sparsity of data is lost.[42] The negative outcomes lead to a risk of reducing the usefulness and trustworthiness of the data in certain applications.[43] The loss of sparsity indicates that a large number of nonzero values appear in the randomized data. This increase in the density of data, for example, may cause a significant delay in a service that processes human flows in real-time.

Terada *et al.* (2015) proposed a method that improves Privelet to satisfy the non-negative constraints and maintain the data sparsity. The main feature of this method is the *top-down refinement* process, which satisfies the non-negative constraints when

[40] The accuracy of aggregated population generally degrades because of the noises that are accumulated from narrower areas to wider ones.

[41] Discrete wavelet transformation is a linear transformation applied primarily in the field of image processing. Conceptually similar to the Fourier transform, the transformation represents a continuous function by the superposition of wavelets that are local waves; Xiao, Wang, and Gehrke (2010) adopted the Haar wavelet transform (Stollnitz, DeRose, and Salesin [1996]) that uses the Haar basis.

[42] Loss of data sparsity stems from nonzero aggregate values due to added noises even though the most exact values take the zero because much of the original data is sparse (taking zero values).

[43] Simply correcting negative values to zero is undesirable because it creates a positive bias in the statistics.

applying the inverse of Haar wavelet transformation. Hongo *et al*. (2020) subsequently improved the computational efficiency by omitting a part of the top-down refinement process that follows the shape of a binary tree. Their method prunes (omits) processes after certain branch points on the tree.

B. Applications of Local Differential Privacy

1. RAPPOR

Randomized aggregatable privacy-preserving ordinal response (RAPPOR; Erlingsson, Pihur, and Korolova [2014]) combined two-stage randomized responses and a Bloom filter (Bloom [1970])[44] to collect personal data from users while satisfying local differential privacy. The Bloom filter converts the data format from string type to numerical type and compresses the data. This method is applicable to arbitrary string data as well as numerical data, and provides robust privacy protection against multiple queries.

RAPPOR first converts original data v into fixed-length data B of k bits through the Bloom filter. Next, B is randomized bit-by-bit with two-stage randomization responses. In the first stage, a *permanent* randomized response converts B to B'. This conversion is performed only once for each data. In the second stage, an *instantaneous* randomized response generates bit string S of length k from B'. S is sent to the central server as user data from the user's device and used for aggregation. The instantaneous randomized response is generated repeatedly for each query.

The permanent randomized response protects private information from an averaging attack in which the same query is repeatedly sent to find the true value of B. The permanent use of B' instead of B ensures that an attacker cannot definitively determine B from the outputs of RAPPOR. The instantaneous randomized response makes it difficult for an attacker to trace an identical user using B' as a clue. Thus, permanent and instantaneous randomized responses protect privacy from long-term and short-term risks, respectively.

Specifically, in the permanent randomized response, for the i-th bit $(0 \leq i \leq k)$ of B, the corresponding bit of B' is determined as follows:

$$B_i^{'} = \begin{cases} 1, & \text{with probability } \frac{1}{2}f \\ 0, & \text{with probability } \frac{1}{2}f \\ B_i, & \text{with probability } 1-f, \end{cases}$$

where B_i, B_i' represent the i-th bit values of B, B', respectively. The parameter f represents the probability of exchanging response values. In the instantaneous randomized response, all digits of S are initialized with zero, and some are flipped to one in accor-

[44] A Bloom filter is a fixed-length data structure that makes it possible to probabilistically and quickly determine whether an element is included in a set. The advantages are that it saves memory space and the computational complexity of the decision is as fast as O(1). The disadvantage is that data that does not exist in the set may be misjudged as existing.

dance with the following probabilities:

$$\Pr[S_i = 1] = \begin{cases} q, & \text{if } B'_i = 1 \\ p, & \text{if } B'_i = 0, \end{cases}$$

for each i-th bit ($1 \leq i < k$).

Demonstrations have shown the usefulness of RAPPOR for counting queries that count the occurrences of specific words in user data. RAPPOR has been implemented in the open source project Chromium.[45] It has also been incorporated into the Google Chrome browser to obtain user statistics on search engine use.

2. Contact trace application for infectious disease

Apple and Google (2020) proposed the Exposure Notification System as a mechanism for confirming close contacts on the basis of device locations while protecting users' privacy. This system was developed as a countermeasure against the spread of COVID-19. The application based on this mechanism enables local devices to receive notifications which indicate possible close contacts, while the administrator of the application (Apple or Google) cannot identify the devices on which the notifications appear or the locations of such devices. Only public health authorities can identify the devices on which the notifications are displayed, though they still cannot obtain the locations.

Apple and Google (2021) also proposed Exposure Notification Privacy-preserving Analytics (ENPA) as a method of generating statistics on notifications by combining privacy-preserving techniques such as local differential privacy, secret sharing, and zero-knowledge proof. According to Apple and Google (2021), the system structure of ENPA can be considered one structure consisting of three types of servers.

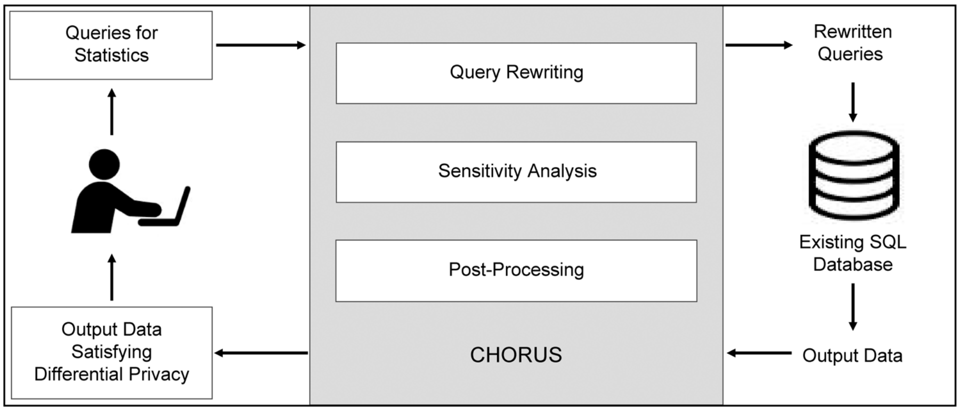
The high-level specification of ENPA can be described as follows. First, a server (denoted as server 1) converts multiple indicators[46] regarding close contacts detected by users' local devices into discrete data. The discrete data can be represented in the form of a binary vector (series of 0 or 1). Server 1 then randomizes the discretized data, for example, with the randomized response, so as to satisfy local differential privacy. Finally, server 1 and the other two servers (denoted as servers 2 and 3) cooperatively produce statistics from the randomized data through secret sharing and zero-knowledge proof (secret-shared non-interactive proof; Corrigan-Gibbs and Boneh [2017]). Note that server 1 is trustworthy, and servers 2 and 3 are assumed not to disclose their secrets to each other.

In the final step, server 1 fragments the randomized data and sends them as secrets to servers 2 and 3 through secret sharing. After receiving the secrets, servers 2 and 3 cooperate to provide the zero-knowledge proof that each pair of the fragmented data is genuine, i.e., the sum of each pair of fragmented data is equal to the (original) randomized data. With the validation of the proof, either server 2 or 3, which is in charge of

[45] For details, see the Chromium Projects design documents, "RAPPOR (randomized aggregatable privacy-preserving ordinal response)." https://sites.google.com/a/chromium.org/dev/developers/design-documents/rappor

[46] Indicators are continuous values. The continuous values are typically converted into integers that show the corresponding intervals in the histogram. Although the indicators are not specified in Apple and Google (2021), they may include, for example, the time and location of close contacts.

Figure 4 Overview of CHORUS



Source: Adapted from Johnson *et al.* (2020).

calculating statistics, can produce reliable statistics. Most importantly, neither server can know the original data.

C. General-Purpose Framework Highly Compatible with SQL Databases

Approaches that modify algorithms for calculating statistics to satisfy differential privacy require expertise in both differential privacy and statistical analysis. These approaches lack versatility and flexibility due to the difficulty of modifying a variety of statistical queries. Thus, research (Kotsogiannis *et al.* [2019], Bater *et al.* [2020], Wilson [2020]) is underway on a generic framework that enables statistical analysis with privacy protection for analysts who are not highly skilled in privacy-preserving techniques.

CHORUS (Johnson *et al.* [2020]) is a general-purpose framework enabling preprocessing that modifies queries before being sent to SQL databases, in addition to post-processing that modifies (adds noise to) values returned from the databases. The pre-processing provides support for queries with *clipping*; a query that enforces a bound on the maximum and/or minimal values of data and requires performing statistical computation on the clipped data. This approach is scalable because it satisfies differential privacy without making any changes to existing database systems.

The overall process consists of three stages: *query rewriting*, *sensitivity analysis*, and *post-processing* (Figure 4). For example, suppose that analysts intend to run a query with clipping to calculate the mean of data among values that fall within a certain range. In the first stage, when the data is named *distance*, the query is rewritten as follows:

Before rewriting: SELECT SUM (distance) FROM database,

After rewriting: SELECT SUM (max(0, min(100, distance))) AS SUM FROM database.

In the second stage, global sensitivity $\mathrm{GS} = (u - l) * s$ is calculated by sensitivity analysis. Then a noise that follows $\mathrm{Laplace}(\mathrm{GS}/\epsilon)$ is added to a return value of the query. Here, parameters u and l denote the upper (100) and lower (0) bounds of the clipping, respectively. s, or *stability*, denotes the upper bound of the number of records that are excluded by the clipping.

CHORUS is open source,[47] and Uber utilizes CHORUS for their internal research to comply with GDPR.

D. Application to Machine Learning

Machine learning models learn training data that sometimes contains private and confidential information. Such secret information is at risk of being leaked from the outputs of machine learning models.

Typical known attacks include the membership inference attack (Shokri *et al.* [2017]) and the model inversion attack (Fredrikson, Jha, and Ristenpart [2015]). Incorporating differential privacy into the training process of machine learning has been studied as a defense against these threats.

Abadi *et al.* (2016) proposed a method that applies differential privacy to deep learning. This method has been proven to be effective against strong attackers who have internal information of models. It is particularly effective when machine learning models are installed on users' devices and an attacker knows both their parameters and the learning algorithm. Experiments demonstrated that the method achieved high computational efficiency and accuracy within a modest privacy budget ϵ of a single digit number. Deep learning models with tens of thousands to millions of parameters were trained on TensorFlow.[48] Benchmark image classification tasks on the MNIST[49] and CIFAR-10[50] datasets yielded accuracies of 97% and 73%, respectively, while satisfying $(8, 10^{-5})$-differential privacy (see Section IV. B. for definition).

Arachchige *et al.* (2020) proposed a method that applies local differential privacy to deep learning. This method enables an individual to add a randomization layer, which adds random numbers to training data before the data leave the user's device and reach potentially untrusted machine learning services. Similar to the previously mentioned studies, experiments on the MNIST and CIFAR-10 datasets demonstrated high accuracy rates of 91–96%, even with a small privacy budget ($\epsilon = 0.5$).

VI. Discussion

Applications of differential privacy to corporate activities are expanding. This section discusses limitations and challenges of differential privacy and the desired privacy protection.

A. Challenges of Differential Privacy

Theoretical research on differential privacy has mostly matured. The future challenges are to disseminate it to social infrastructures. Differential privacy ensures privacy pro-

[47] For details, see the project homepage. https://github.com/uvm-plaid/chorus

[48] An open source platform for machine learning, released in 2015 by Google.

[49] The Modified National Institute of Standards and Technology (MNIST) dataset is a large database containing images of handwritten digits and their corresponding labels. The database is commonly used for training machine learning models as a performance measurement benchmark in the field of image recognition.

[50] The CIFAR-10 (Canadian Institute For Advanced Research) dataset is a database containing 60,000 color images of 32 × 32 pixel objects and the corresponding 10 label data that represents the classification of the objects.

tection independent of attack models. As personal information circulating in society increases, it will be more difficult to assume plausible background knowledge of attackers. Thus, differential privacy will become more desirable in the future due to its precautionary nature. Adoption of general-purpose frameworks can be a promising option due to the following two reasons. First, it is difficult to develop mechanisms that satisfy differential privacy in a flexible and agile manner. Second, expertise of privacy-preserving techniques and statistical analysis is required in order to make use of differential privacy.

However, differential privacy is not a panacea. The true probability distribution followed by each record of databases is not clear in many cases. Requirements for differential privacy, based on specific assumptions about the probability distribution, may not be satisfied in practice. For example, if records of a database are strongly correlated with each other, strong privacy protection cannot be ensured. In addition, the best practice of privacy protection depends on the environment in which the database is operated. The advantage of differential privacy is mostly demonstrated in situations where databases accept and respond to a variety of queries. Thus, differential privacy is not always the optimal choice. Appropriate access control may be more suitable for internal use in companies. Furthermore, at the time of this writing, there is no consensus on the level of privacy budget that is sufficient for protecting privacy in practice.

B. Required Comprehensive Privacy Protection Measures

Mathematical techniques and information technology alone cannot provide the functionalities that grant individuals control over the disclosure, correction, and deletion of their own information. To implement these functionalities, privacy protection measures should be integrated into laws, regulations, and IT systems in accordance with the concept of privacy-by-design. All of these measures empower individuals to control over their own information.

Privacy protection measures improve the trade-off between social benefits from data utilization and individuals' benefits from privacy protection. Thus, introducing privacy protection measures does not merely restrain the use of personal information but can expand it, i.e., the use of personal information can be socially acceptable only if appropriate privacy safeguards are in place. In particular, for companies that collect massive amounts of personal data, such as digital platform providers and financial institutions, it is essential for their business to gain social acceptance. Privacy protection is critical since it affects how strictly such companies should be regulated. Moreover, further exploitation of personal data will bring novel and unprecedented threats of privacy invasion. As technological innovations pose these threats in general, non-technical countermeasures will also need to be taken.

C. Technological Innovation and Desirable Privacy Protection

As the privacy protection framework becomes a growing priority in modern society, more attention should be directed to the purposes for which personal information is used and to the manner of how it is used. Personal information has been referred to as *oil* or *currency* in the Internet age. The utilization of personal information combined with AI has become deeply entrenched in social infrastructure in finance, insurance, human resource management, and judiciary, in order to maximize social benefits from personal information.

However, the pursuit of rational goals, such as maximizing benefits and minimizing risks and losses, does not necessarily bring happiness to society. The utilization of AI and personal information may lead to new ELSI related to fairness as well as privacy. Kukita (2021) stated that many people naively believe that technology is neutral in the sense that technology itself is neither good nor evil and the purpose of its use does matter, but such neutrality does not apply to AI as it can be easily abused. For example, a surveillance society where all kinds of personal information is collected and censored by the government is not desirable, but the pursuit of rational and virtuous objectives against crime and corruption could unexpectedly lead to such a dystopia.

In contrast, technological development will lead to more potential options for privacy protection implemented in society. Even though comprehensive surveillance is feasible, society can intentionally choose not to install it. Privacy protection measures can mitigate ELSI by improving the trade-off between social and individuals' benefits. In addition, new technologies, including AI, can change the values and norms in society. As social norms change, the desired privacy protection will also change. It is essential to continue to search for the ideal privacy protection while building social consensus.

References

Abadi, Martin, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang, “Deep Learning with Differential Privacy,” Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016, pp. 308–318.

Abowd, John, Daniel Kifer, Brett Moran, Robert Ashmead, Philip Leclerc, William Sexton, Simson Garfinkel, and Ashwin Machanavajjhala, “Census Topdown: Differentially Private Data, Incremental Schemas, and Consistency with Public Knowledge,” Technical Report, United States Census Bureau, 2019 (available at https://systems.cs.columbia.edu/private-systems-class/papers/Abowd2019Census.pdf).

———, and Ian M. Schmutte, “An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choice,” *American Economic Review*, 109(1), 2019, pp. 171–202.

Adam, Nabil R., and John C. Wortmann, “Security-Control Methods for Statistical Databases: A Comparative Study,” *ACM Computing Surveys*, 21(4), 1989, pp. 515–556.

Agrawal, Rakesh, and Ramakrishnan Srikant, “Privacy-Preserving Data Mining,” *ACM SIGMOD Record*, 29(2), 2000, pp. 439–450.

Apple, and Google, “Exposure Notifications: Using Technology to Help Public Health Authorities Fight COVID-19,” Google, 2020 (available at https://www.google.com/covid19/exposurenotifications/).

———, and ———, “Exposure Notification Privacy-Preserving Analytics (ENPA),” White Paper, Apple, 2021 (available at https://covid19-static.cdn-apple.com/applications/covid19/current/static/contact-tracing/pdf/ENPA_White_Paper.pdf).

Arachchige, Pathum Chamikara Mahawaga, Peter Bertok, Ibrahim Khalil, Dongxi Liu, Seyit Camtepe, and Mohammed Atiquzzaman, “Local Differential Privacy for Deep Learning,” *IEEE Internet of Things Journal*, 7(7), 2020, pp. 5827-5842.

Bater, Johes, Yongjoo Park, Xi He, Xiao Wang, and Jennie Rogers, “SAQE: Practical Privacy-Preserving Approximate Query Processing for Data Federations,” Proceedings of the VLDB Endowment, 13(12), Very Large Data Bases Endowment, 2020, pp. 2691–2705.

Beck, Leland L., “A Security Mechanism for Statistical Databases,” *ACM Transactions on Database Systems*, 5(3), 1980, pp. 316–338.

Benjamin, Ruha, *Race after Technology: Abolitionist Tools for the New Jim Code*, Polity Press, 2019.

Bloom, Burton H., “Space/Time Trade-Offs in Hash Coding with Allowable Errors,” *Communications of the ACM*, 13(7), 1970, pp. 422–426.

Bonawitz, Keith, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth, “Practical Secure Aggregation for Privacy-Preserving Machine Learning,” Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 1175–1191.

Cavoukian, Ann, “Privacy by Design: The 7 Foundational Principles,” Information and Privacy Commissioner of Ontario (January 2011, revised version), 2011 (available at https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf).

———, “Privacy by Design and the Emerging Personal Data Ecosystem,” Information and Privacy Commissioner of Ontario, 2012 (available at https://www.ipc.on.ca/wp-content/uploads/Resources/pbd-pde.pdf).

———, and Drummond Reed, *Big Privacy: Bridging Big Data and the Personal Data Ecosystem through Privacy by Design*, Information and Privacy Commissioner of Ontario Canada, 2013.

Census Scientific Advisory Committee, “Determining the Privacy-Loss Budget: Research into Alternatives to Differential Privacy,” United States Census Bureau, 2021 (available at https://www2.census.gov/about/partners/cac/sac/meetings/2021-05/presentation-research-on-alternatives-to-differential-privacy.pdf).

Corrigan-Gibbs, Henry, and Dan Boneh, “Prio: Private, Robust, and Scalable Computation of Aggregate Statistics,” Proceedings of the 14th USENIX Conference on Networked Systems Design and Implementation, USENIX Association, 2017, pp. 259–282.

Dalenius, Tore, and Steven P. Reiss, “Data-Swapping: A Technique for Disclosure Control,” *Journal* *of Statistical Planning and Inference*, 6(1), 1982, pp. 73–85.

Denning, E. Dorothy, “Secure Statistical Databases with Random Sample Queries,” *ACM Transactions on Database Systems*, 5(3), 1980, pp. 291–315.

———, *Cryptography and Data Security*, Addison-Wesley Longman Publishing, 1982.

———, and Peter J. Denning, “The Tracker: A Threat to Statistical Database Security,” *ACM Transactions on Database Systems*, 4(1), 1979, pp. 76–96.

Duchi, John C., Michael I. Jordan, and Martin J. Wainwright, “Local Privacy and Statistical Minimax Rates,” Proceedings of 2013 IEEE 54th Annual Symposium on Foundations of Computer Science, 2013, pp. 429–438.

Dwork, Cynthia, “Differential Privacy: A Survey of Results,” Proceedings of International Conference on Theory and Applications of Models of Computation 2008, Lecture Notes in Computer Science, 4978, Springer, 2008, pp. 1–19.

———, and Jing Lei, “Differential Privacy and Robust Statistics,” Proceedings of the 41st Annual ACM Symposium on Theory of Computing, 2009, pp. 371–380.

———, Frank McSherry, Kobbi Nissim, and Adam Smith, “Calibrating Noise to Sensitivity in Private Data Analysis,” Proceedings of Theory of Cryptography Conference 2006, Lecture Notes in Computer Science, 3876, Springer, 2006, pp. 265–284.

———, and Aaron Roth, “The Algorithmic Foundations of Differential Privacy,” *Foundations and Trends in Theoretical Computer Science*, 9(3–4), Now Publishers, 2014, pp. 211–407.

Erlingsson, Úlfar, Vasyl Pihur, and Aleksandra Korolova, “RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response,” Proceedings of the 2014 ACM SIGSAC Conference on Computer Communications Security, 2014, pp. 1054–1067.

Fredrikson, Matt, Somesh Jha, and Thomas Ristenpart, “Model Inversion Attacks That Exploit Confidence Information and Basic Countermeasures,” Proceedings of the 2015 ACM SIGSAC Conference on Computer and Communications Security, 2015, pp. 1322–1333.

Garfinkel, Simson, John M. Abowd, and Christian Martindale, “Understanding Database Reconstruction Attacks on Public Data,” *Communications of the ACM*, 62(3), 2019, pp. 46–53.

Gennaro, Rosario, Craig Gentry, Bryan Parno, and Mariana Raykova, “Quadratic Span Programs and Succinct NIZKs without PCPs,” Proceedings of EUROCRYPT 2013, Lecture Notes in Computer Science, 7881, Springer, 2013, pp. 626–645.

Hayashi, Mako, “Shinyo Sukoa Ni Kansuru Kiritsu No Arikata: Waga Kuni To Beikoku Ni Okeru Shinyo Joho No Toriatsukai Wo Fumaete (Discipline on Credit Scoring: Analysis Based on the Handling of Credit Information in Japan and the United States),” *Kin’yu Kenkyu* (Monetary and Economic Studies), 41(4), Institute for Monetary and Economic Studies, Bank of Japan, 2022, pp. 31–68 (in Japanese).

Holohan, Naoise, Douglas J. Leith, and Oliver Mason, “Optimal Differentially Private Mechanisms for Randomised Response,” *IEEE Transactions on Information Forensics and Security*, 12(11), 2017, pp. 2726–2735.

Hongo, Sadayuki, Masayuki Terada, Akihiro Suzuki, and Jun Inagaki, “Hifuseichika Wo Tomonau Puraiburettohou Ni Okeru Enzan Kouritsukashuhou No Seinou Kozyo (Improvement of the Computational Efficiency for Privelet with Non-Negative Refinement),” *Johoshorigakkai Ronbunshi* (Transactions of Information Processing Society of Japan), 61(9), Information Processing Society of Japan, 2020, pp. 1458–1471 (in Japanese).

Hundepool, Anco, Josep Domingo-Ferrer, Luisa Franconi, Sarah Giessing, Rainer Lenz, Jane Naylor, Eric Schulte Nordholt, Giovanni Seri, and Peter-Paul De Wolf, *Handbook on Statistical Disclosure Control*, The European Commission, 2010.

———, ———, ———, ———, Eric Schulte Nordholt, Keith Spicer, and Peter-Paul de Wolf, *Statistical Disclosure Control*, John Wiley & Sons, 2012.

Igarashi, Masaru, and Katsumi Takahashi, “Chumoku No Puraibashi Differential Privacy (Brief Introduction to Differential Privacy),” *Konpyuta Sofutouea* (Computer Software), 29(4), Japan Society for Software Science and Technology, 2012, pp. 40–49 (in Japanese).

Ishii, Kaori, “America No Puraibashi Hogo Ni Kansuru Doko (Recent Trends on Privacy Protection in the United States),” *Johoshori* (IPSJ Magazine), Information Processing Society of Japan, 55(12), 2014, pp. 1346–1352 (in Japanese).

Iwamura, Mitsuru, and Yuko Nishijima, "Tokeideta No Kohyokokai To Puraibashi No Hogo: Suironseigyo No Riron, Sono Syokai To Oyo (Microdata Disclosure and Privacy Protection: Theory of Inference Control and its Applications)," *Kin'yu Kenkyu* (Monetary and Economic Studies), 10(4), Institute for Monetary and Economic Studies, Bank of Japan, 1991, pp. 67–93 (in Japanese).

Johnson, Noah, Joseph P. Near, and Dawn Song, "Towards Practical Differential Privacy for SQL Queries," Proceedings of the VLDB Endowment, 11(5), Very Large Data Bases Endowment, 2018, pp. 526–539.

———, ———, Joseph M. Hellerstein, and Dawn Song, "CHORUS: A Programming Framework for Building Scalable Differential Privacy Mechanisms," Proceedings of 2020 IEEE European Symposium on Security and Privacy, 2020, pp. 535–551.

Kairouz, Peter, Keith Bonawitz, and Daniel Ramage, "Discrete Distribution Estimation under Local Privacy," Proceedings of the 33rd International Conference on Machine Learning, 48, ML Research Press, 2016, pp. 2436–2444.

Kotsogiannis, Ios, Yuchao Tao, Xi He, Maryam Fanaeepour, Ashwin Machanavajjhala, Michael Hay, and Gerome Miklau, "PrivateSQL: A Differentially Private SQL Query Engine," Proceedings of the VLDB Endowment, 12(11), Very Large Data Bases Endowment, 2019, pp. 1371–1384.

Kukita, Minao, "Jinkochino To Ningen No Yoriyoi Kyosei No Tameni (For Better Symbiosis of Artificial Intelligence and Humans)," *RAD-IT21 WEB magazine*, Nippon RAD, 2020 (available at https://rad-it21.com/ai/kukita-minao_20200317/, in Japanese).

———, "Jinkochino No Rinri To Sono Kyoiku (Ethics of Artificial Intelligence in Education)," Shingaku Giho (IEICE Technical Report), 121(119), The Institute of Electronics, Information and Communication Engineers, 2021, pp. 50–55 (in Japanese).

Li, Ninghui, Tiancheng Li, and Suresh Venkatasubramanian, "*t*-Closeness: Privacy Beyond *k*-Anonymity and *l*-Diversity," Proceedings of 2007 IEEE 23rd International Conference on Data Engineering, IEEE, 2007, pp. 106–115.

———, Wahbeh Qardaji, and Dong Su, "On Sampling, Anonymization, and Differential Privacy or, *k*-Anonymization Meets Differential Privacy," Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security, 2012, pp. 32–33.

Lian, Huanhuan, Tianyu Pan, Huige Wang, and Yunlei Zhao, "Identity-Based Identity-Concealed Authenticated Key Exchange," Proceedings of 26th European Symposium on Research in Computer Security, Lecture Notes in Computer Science, 12973, Springer, 2021, pp. 651–675.

Machanavajjhala, Ashwin, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam, "*l*-Diversity: Privacy beyond *k*-Anonymity," *ACM Transactions on Knowledge Discovery from Data*, 1(1), 2007.

McMahan, Brendan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 54, ML Research Press, 2017, pp. 1273–1282.

McSherry, Frank, and Kunal Talwar, "Mechanism Design via Differential Privacy," Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science, 2007, pp. 94–103.

Meyerson, Adam, and Ryan Williams, "On the Complexity of Optimal *k*-Anonymity," Proceedings of the 23rd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, 2004, pp. 223–228.

Nakagawa, Hiroshi, *Puraibashi Hogo Nyumon: Hoseido To Suriteki Kiso* (Introduction to Privacy Protection: Legal System and Mathematical Foundations), Keiso Shobo, 2016 (in Japanese).

Nissim, Kobbi, Sofya Raskhodnikova, and Adam Smith, "Smooth Sensitivity and Sampling in Private Data Analysis," Proceedings of the 39th Annual ACM Symposium on Theory of Computing, 2007, pp. 75–84.

O'Neil, Cathy, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown Press, 2016.

Organisation for Economic Co-operation and Development (OECD), "The OECD Privacy Framework," OECD, 2013 (available at https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf).

Ramacher, Sebastian, Daniel Slamanig, and Andreas Weninger, "Privacy-Preserving Authenticated Key Exchange: Stronger Privacy and Generic Constructions," Proceedings of European Symposium on Research in Computer Security 2021, Lecture Notes in Computer Science, 12973, Springer, 2021, pp. 676–696.

Sakuma, Jun, *Deta Kaiseki Ni Okeru Puraibashi Hogo* (Privacy Protection in Data Analysis), Kodansha, 2016 (in Japanese).

Shokri, Reza, Marco Stronati, Congzheng Song, and Vitaly Shmatikov, "Membership Inference Attacks against Machine Learning Models," Proceedings of 2017 IEEE Symposium on Security and Privacy, 2017, pp. 3–18.

Sogabe, Masahiro, Shuya Hayashi, and Masahiro Kurita, *Johoho Gaisetsu Dai 2 Han* (Introduction to Information Law second edition), Kobundo, 2019 (in Japanese).

Stollnitz, Eric J., Anthony D. DeRose, and David H. Salesin, *Wavelets for Computer Graphics: Theory and Applications*, Morgan Kaufmann Publishers, 1996.

Sweeney, Latanya, "*k*-Anonymity: A Model for Protecting Privacy," *International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems*, 10(5), World Scientific Publishing, 2002, pp. 557–570.

Terada, Masayuki, Ryohei Suzuki, Takayasu Yamaguchi, and Sadayuki Hongo, "Daikibo Shukei Deta Heno Sabunpuraibasi No Tekiyo (On Publishing Large Tabular Data with Differential Privacy)," *Johoshorigakkai Ronbunshi* (Transactions of Information Processing Society of Japan), 56(9), Information Processing Society of Japan, 2015, pp. 1801–1816 (in Japanese).

United States Census Bureau, "A History of Census Privacy Protections," United States Census Bureau, 2019 (available at https://www2.census.gov/library/visualizations/2019/communications/history-privacy-protection.pdf).

Warner, Stanley L., "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American Statistical Association*, 60(309), 1965, pp. 63–69.

Willenborg, Leon, and Ton de Waal, *Elements of Statistical Disclosure Control*, Lecture Notes in Statistics, 155, Springer, 2001.

Wilson, Royce J., Celia Yuxin Zhang, William Lam, Damien Desfontaines, Daniel Simmons-Marengo, and Bryant Gipson, "Differentially Private SQL with Bounded User Contribution," Proceedings on Privacy Enhancing Technologies Symposium, 2020(2), De Gruyter, 2020, pp. 230–250.

Xiao, Xiaokui, Guozhang Wang, and Johannes Gehrke, "Differential Privacy via Wavelet Transforms," Proceedings of 2010 IEEE 26th International Conference on Data Engineering, 2010, pp. 225–236.

Yang, Mengmeng, Lingjuan Lyu, Jun Zhao, Tianqing Zhu, and Kwok-Yan Lam, "Local Differential Privacy and Its Applications: A Comprehensive Survey," arXiv:2008.03686, 2020.