BOJ Reports & Research Papers

Payment and Settlement Systems Report Annex Series

Payment and Settlement Systems Report – Annex

Privacy Enhancing Technologies: Payments and Financial Services in a Digital Society

Payment and Settlement Systems Department
Bank of Japan
January 2023

Payment and Settlement Systems Report Annex Series

The Bank of Japan regularly publishes the Payment and Settlement Systems Report with the aim of providing an overview and evaluating the development of its payment and settlement systems. The report also introduces the engagement of the Bank of Japan and relevant organizations to improve the safety and efficiency of the payment and settlement systems.

The Payment and Settlement Systems Report Annex Series provides in-depth analyses of specific themes concerning those systems. This report focuses on privacy enhancing technologies. It is part of the Bank's ongoing work on "privacy protection and handling of end-user information," which was identified in "The Bank of Japan's Approach to Central Bank Digital Currency," released in October 2020, as one of the topics related to its institutional arrangements.

This document is an English translation of the Japanese original published on September 29, 2022.

Privacy Enhancing Technologies:
Payments and Financial Services in a Digital Society

Executive Summary

In recent years, businesses in Japan and abroad have created many services through the collection of customer data, and the utilization of such data has become an important motivation for business development in the area of payment and settlement, as well as financial services. At the same time, the importance of anti-money laundering and combating the financing of terrorism (AML/CFT) for payments is becoming increasingly recognized, and international discussions are underway. The use of data is becoming more important in order to advance AML/CFT and establish an effective mechanism. Against this backdrop, technologies are being developed to contribute to the protection of users' privacy while using their data for business creation and soundness of transactions.

Specifically, there are the concepts of *anonymization*, in which data is altered so that individuals cannot be identified, and *differential privacy*, in which the possibility of identification from analysis results is suppressed by adding noise or by other means. Other methods include *secure computation*, in which analysis is performed while data is kept secret, and *trusted execution environment*, in which hardware technology is used to perform computational processing in a secure area where confidentiality is maintained. In addition, *federated learning*, in which users' information contained in each party's data is kept secret from other parties while collaborating with them to perform machine learning, is also being studied. As a privacy-related concept, *self-sovereign identity* is also attracting attention.

Discussions on privacy enhancing technologies and their application to payments and financial services may have implications for studies on Central Bank Digital Currency, which have been underway in many countries including Japan.

The Bank of Japan has been steadily proceeding with experiments and explorations into institutional arrangements to ensure the stability and efficiency of the entire payment and settlement system, and to be able to respond appropriately to various changes in the environment in the future. As part of its explorations into the institutional arrangements, the Bank will continue to research and study privacy protection related to digital currencies with a wide range of stakeholders.

Table of Contents

I. Introduction

In line with "The Bank of Japan's Approach to Central Bank Digital Currency" released in October 2020,[1] the Bank of Japan has been conducting experiments and exploring institutional arrangements. The publication lists "privacy protection and handling of end-user information" as one of the items to be explored in terms of future institutional arrangements. As part of this exploration, this paper focuses on technologies that contribute to the protection of user privacy, provides an overview of these technologies and examples of their application, and discusses their implications for the areas of payment and settlement and of Central Bank Digital Currency (CBDC).

In recent years, many services have been created through the collection of customer data by businesses in Japan and abroad, and the utilization of data has become an important motivation for business development in the various areas including payment and settlement. At the same time, there has been growing awareness of the importance of anti-money laundering and combating the financing of terrorism (AML/CFT) in payments, and international discussions are underway. In order to improve the accuracy of AML/CFT and establish an effective framework, utilizing data is becoming increasingly important.

With the growing demand for data utilization, there is also a growing public concern about privacy. In many countries, including Japan, privacy is recognized as a personal right (i.e., a right to an individual's personality). To ensure appropriate privacy protection in the use of data, several aspects must be taken into account, such as protection of interests of personality and ethics. In addition to complying with laws and regulations such as the Act on the Protection of Personal Information, based on these points, service providers are required to consider privacy protection when deciding what data to collect and store, the purposes for which the data is used, and who is allowed to access the data. In this context, "preventing information from being known or inferred by the parties who should not know the information" is considered to be one of the important factors in protecting privacy.

Against this backdrop, technologies that contribute to the protection of users' privacy while using data for business creation and soundness of transactions have been developed in recent years. This paper aims to frame these technologies as privacy enhancing technologies.

---

[1] Bank of Japan, "The Bank of Japan's Approach to Central Bank Digital Currency," https://www.boj.or.jp/en/announcements/release_2020/data/rel201009e1.pdf, 2020.

Section 2 describes the basics of these technologies, and Section 3 introduces examples of their application to payments and financial services. Section 4 discusses points that should be kept in mind when planning the implementations of these technologies in society, as they may also have implications for CBDC, which is currently under consideration in many countries. The term "privacy enhancing technologies" is ambiguous, and there are various ways of interpreting the specific technologies that fall under this category. The technologies and concepts introduced in this paper are only examples.

II. Overview of Privacy Enhancing Technology

In general, entities that acquire user information in the process of providing services are required to take various measures in terms of systems and technologies to protect the privacy of the users. Specifically, for example, a policy of acquiring, processing, and storing only the minimum necessary information (the concept of data minimization[2]) in the process of providing services is considered the norm. In addition, databases containing acquired data should be operated under a framework well organized to ensure information security (i.e., adequate information security measures). On top of that, if the user consents, or if the granularity of the information is such that individuals cannot be identified, there will be room for a wide range of entities to utilize the information stored in the database while protecting privacy. In the payment and settlement area, there may be necessary uses as required by laws and regulations for AML/CFT purposes. When these uses are made, on top of the premise of not using information that data subjects do not intend to be used, it is extremely important to prevent information from being known or inferred by the parties who should not know the information. For this reason, it is necessary to devise various measures to ensure that individuals cannot be identified from the dataset itself or from the results of analysis of the dataset. This paper introduces technologies that contribute to such privacy protection measures.

First, there are techniques to alter data so that individuals cannot be identified by the entity handling the dataset or analysis results (*anonymization*, Section A). Techniques that suppress identifiability from analysis results by adding noise are also considered useful (*differential privacy*, Section B). In addition, in recent years, a method of conducting analysis while

[2] The ISO/IEC 29100 Privacy Framework, an international standard published in 2011, lists data minimization as one of its 11 principles, and the EU General Data Protection Regulation and other national personal data protection laws also include data minimization as a principle.

keeping data confidential has been developed as an alternative to altering the contents of a dataset or adding noise (*Secret Computation*, Section C). There are also methods that use hardware technologies to ensure that the intended computation is performed in a confidential state (*TEE*, Section D). Furthermore, there has been research on methods of collaboratively learning while keeping user information contained in an organization's dataset that is kept secret from other organizations, such as by utilizing the above-mentioned technologies in data used for machine learning (*federated learning*, Section E).

The technologies described above are not necessarily used in isolation but rather in combination to achieve privacy protection. It should be noted that these technologies are still developing, and experts have varying classifications and terminologies. In addition, the use of the technologies described in this paper does not necessarily mean the immediate fulfillment of legal requirements.

A. Anonymization: Altering Data

To prevent individuals from being identified in a dataset, "identifiers" (information that directly identifies specific individuals, such as names and important ID numbers) are first removed or replaced. Furthermore, it is important to remove or replace "quasi-identifiers" (e.g., postcode, age, sex), each of which cannot identify a specific individual by itself but may be combined to identify an individual. The following methods can be used to suppress the risk of individual identification through data alteration for *anonymization* by removing or replacing identifiers and quasi-identifiers (Figure 1).

Figure 1: Common methods of suppressing risk of individual identification through data alteration

Attribute (Column) removal	Removing an attribute (e.g., name).
Pseudonymization	Replacing an attribute or combination of attributes (e.g., name, date of birth, etc.) with a code or number.
Generalization	Replacing the value of an attribute with a higher level value or concept; for example, replacing "age per year" with "age in increments of 10 years," "cabbage" with "vegetable".
Top/Bottom coding	Summarizing attribute values that are particularly large or small for a numerical attribute; for example, all the people who are 100 years old or older are indicated as "100 years old or older."
Micro aggregation	After grouping the original data, replacing each attribute value of a record in the same group with a representative value of the group.
Synthetic data	Artificial data generated to reproduce the original data statistically.
Record (Row) removal	Removing records of outliers, such as particularly large values; for example, records with a value of 120 years old or more are removed.

Note: Based on the report of the Technical Study WG of the Study Group on Personal Data (December 2013), etc.

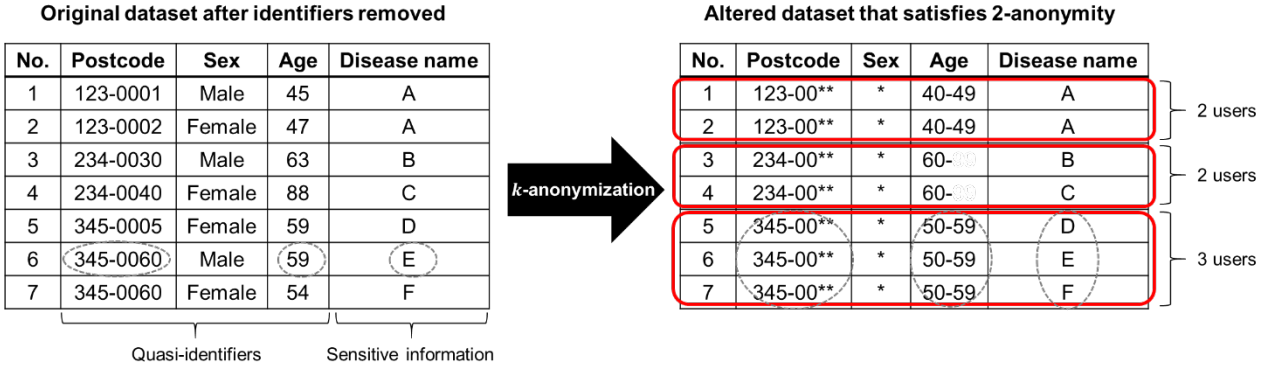
The following indicators have been devised for evaluating the degree of effectiveness achieved when combining the above methods to control the risk of individual identification. Experts propose altering the data to satisfy the properties of these indicators. It is important to use these indicators because there is a possibility that (1) excessive alternation of values in the dataset will result in reduced utility (i.e., its analysis results will become less useful) or (2) seemingly appropriate data alternation will result in the inference of an individual's identity.

1. k-anonymity

Suppose a dataset from which user identifiers (e.g., names) have been removed. If an analyst has some knowledge of the quasi-identifiers (e.g., postcode, age, sex) of a user in the dataset, the analyst may narrow down and identify the individual using a combination of quasi-identifiers and link the individual to information that the user in the dataset does not want others to know ("sensitive information" in this paper). For example, if "disease name" is

considered sensitive information in the dataset as displayed in the left table in Figure 2, an analyst who knows both "postcode: 345-0060" and "age: 59 years old" of user X would be able to link user X to the data "disease name: E."

Figure 2: Data satisfying *k*-anonymity



Note: Created based on various sources.

To address this problem, an indicator called *k-anonymity* has been devised to show the degree to which anonymity has been achieved. That is for every individual's record, there are records of at least *k* people whose quasi-identifiers are identical. A dataset is described as having 5-anonymity when there are at least five people in each group who have target attributes. A larger *k* value means that the individuals are less likely to be identified.

The data processing technique for achieving *k*-anonymity is called *k-anonymization*. Anonymization techniques such as generalization, in which values are expressed in ranges, are used to ensure that there are at least *k* records with the same quasi-identifier combination and thus satisfy the predetermined *k*-anonymity. For example, in the right table in Figure 2, anonymization makes it impossible to identify that the aforementioned user's disease name is E from this dataset (note that one can tell that it is either D, E, or F).

2. *l*- diversity

Even if *k*-anonymity is satisfied, if there are only a few variations in the values of sensitive information associated with an individual, it may be possible to guess the sensitive information about that individual. For example, as shown in No. 1 and No. 2 of the right table in Figure 2, if there is only one variation of disease name in the group with the same quasi-identifiers that includes a particular user, the disease name is known to be A, even if *k*-anonymity is satisfied.

An indicator called *l-diversity* has been proposed to solve this problem. While *k*-anonymity focuses on quasi-identifiers, this indicator evaluates anonymity by focusing on sensitive information in addition to quasi-identifiers. When considering groups with the same quasi-identifiers resulting from *k*-anonymization, *l*-diversity means that for any one of such groups, there are *l* or more variations of values of sensitive information contained in the group. This means that even if one knows the combination of a user's quasi-identifiers and which group the user belongs to, there are *l* types of possible sensitive information contained in that group, making it difficult to estimate and thus complementing the aforementioned *k*-anonymity.

3. *t*-closeness

The aforementioned *l*-diversity focuses on the number of variations of values of sensitive information in the same group (*l* variations are sufficient), but even if *l*-diversity is satisfied, a bias in the distribution of sensitive information may enable someone to detect trends in the data. For example, suppose the sensitive information is continuous data such as annual income. Even if there are *l* variations of values of annual income within the same group, if most of them fall within a certain range, meaningful information about the individual can be obtained without having to identify the specific annual income.

To prevent such identification, an indicator called *t-closeness* has been proposed, in which grouping criteria is devised. This refers to the property that for any group, the distance between the distribution of the values of sensitive information in the group and that in the overall dataset is *t* or less. Records are processed and grouped so that the distribution of the sensitive information in each group is close to that in the overall dataset, making it difficult to infer a user's sensitive information.

B. Differential Privacy[3]: Adding Noise to Analysis Results

If an indicator of privacy protection strength is constructed assuming specific background knowledge and attacking capabilities, it will be vulnerable to attacks that the indicator does not assume. *Differential privacy* is an indicator designed to suppress privacy disclosure below a certain level against an attacker with arbitrary background knowledge and against

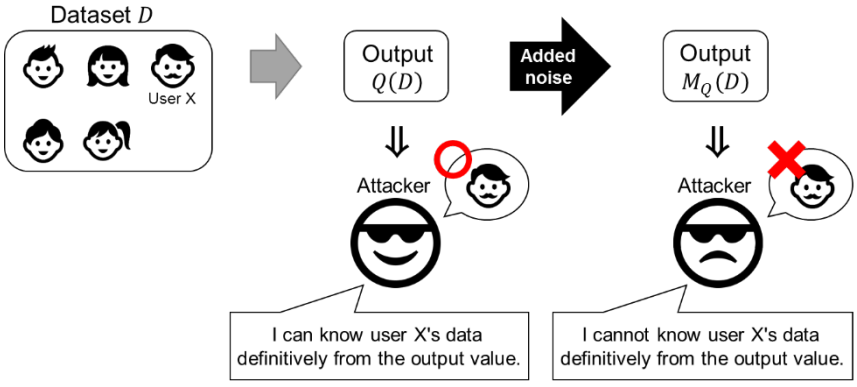
[3] The descriptions in this section are based on the following documents.
Terada, Masayuki, "Sabun-puraibashi towa nanika," *Systems, control and information*, vol. 63, no. 2, pp. 58-63, 2019.

any algorithmic attack. The output of a dataset analysis can be randomized by adding noise to satisfy this indicator, making it difficult to identify information in the dataset.

When a query is made to extract data from a dataset D that contains information about an individual, the output $Q(D)$, which is the result of the query, may expose information that could identify the individual contained in D if used as is. Terada (2019) presented the following example: A query is performed to extract "gender of examinee" and "pass/fail" results from a dataset that records the examination results in a class of 20 students. If the query results are "Female students: 10 passed / 5 failed, Male students: 5 failed," it can be seen that all the passing students were female because adding all of the numbers indicated in the query results will yield the number of students in the class. Thus, even if a male student wanted to keep the results of his own examination secret, it would be revealed that he had failed.

Therefore, a mechanism M_Q for adding noise to $Q(D)$ and use $M_Q(D)$ to denote the output value is considered. $M_Q(D)$ is similar to $Q(D)$, but the output is randomized due to the addition of noise, making it difficult to identify individuals from the output values (Figure 3).[4] The greater the noise added, the harder it becomes to identify individuals, but the more changes made to the results of an analysis, the less useful the results become.

Figure 3: Addition of noise to output values



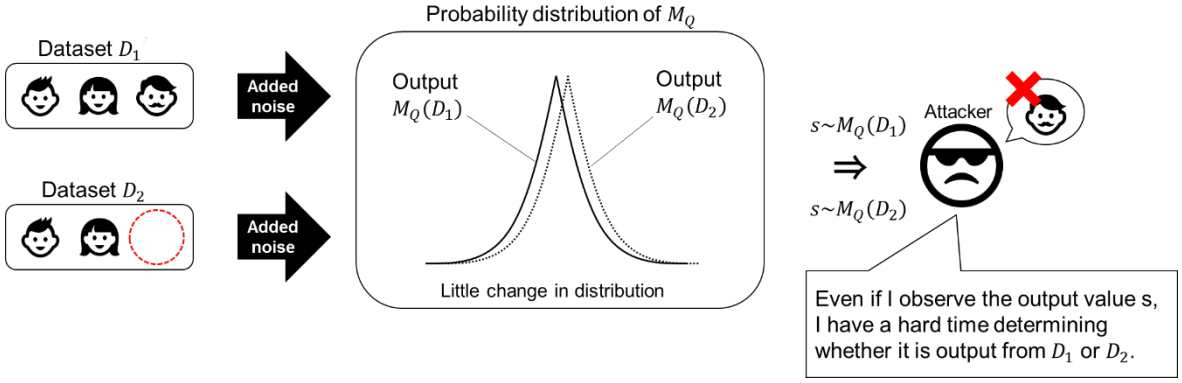
Note: Created based on various sources.

The *ε-differential privacy* is an indicator for quantitatively evaluating the strength of privacy protection. Given two datasets $D_1 \subset D$ and $D_2 \subset D$ that differ only in terms of

[4] There are various methods for adding noise. A commonly used method is the Laplace mechanism, through which noise is added to the output value based on a distribution called the Laplace distribution.

whether they contain data for a certain individual, the output values $M_Q(D_1)$ and $M_Q(D_2)$ after noise addition are random variables. M_Q satisfies ε-differential privacy[5] if the ratio of the probabilities that $M_Q(D_1)$ and $M_Q(D_2)$ output s ($\Pr[M_Q(D_1) = s]$ / $\Pr[M_Q(D_2) = s]$) is less than e^{ε} for any output value s. In this case, as ε is smaller, the difference between the probability distributions of $M_Q(D_1)$ and $M_Q(D_2)$ becomes smaller (harder to distinguish), and it becomes more difficult to determine upon observation whether a certain output value s is from D_1 or D_2. In other words, it becomes difficult to read the personal information that is the difference (Figure 4). If this property is satisfied for any combination of D_1 and D_2, then it is difficult to identify any individual in D from the output M_Q after adding noise.[6]

Figure 4: Overview of ε-difference privacy



Note: Created based on various sources.

C. Secret Computation: Processing Data in a Confidential State

There is a growing need for entities to integrate their datasets with external entities for analysis, or to request external experts to analyze datasets. Accordingly, there has been an increasing number of cases where datasets for analysis are stored on external servers.

---

[5] For details on ε, see the following:
Uno, Yosuke, Akira Sonoda, and Masaki Bessho, "The Economics of Privacy: A Primer Especially for Policymakers," Bank of Japan Working Paper Series No. 21-E-11, 2021.
Kan, Kazutoshi, "Nozomashii praibashi-hogo no arikata wo megutte: sabun-praibashi no yūkō-sei to genkai," *Kinyū Kenkyū*, vol. 41, no. 4, pp. 69-106, 2022.

[6] A specific example of the use of differential privacy-related technologies is the attempt by the U.S. Census Bureau. In the U.S. Census, it was noted that mechanisms existed in the past to avoid disclosing respondent information from statistics. However, in recent years, the increased computing power of computers and relatively inexpensive access to data sources that can be matched have increased the risk of re-identification from the statistics. The Bureau conducted a series of empirical studies related to differential privacy and implemented differential privacy-related technologies when they published statistical data based on the 2020 Census.

However, even if datasets are anonymized or otherwise processed, when they are passed onto other entities, it is necessary to manage the risk of privacy infringement, including with whom they are shared, and users are likely to feel uncomfortable with such a situation. It is important to improve the level of privacy protection and create a system in which the users feel confident.

In response to these issues, research has been conducted in recent years on methods that enable data to be processed in a confidential manner. If the data can be analyzed without knowing the original data, the level of privacy protection can be improved by integrating and analyzing the dataset without allowing others to access it.

1. Technology for handling data in a confidential manner

With encryption, data is converted into a string of characters (ciphertext) whose meaning cannot be read without modification. The ciphertext is returned to its original form (plaintext) through the process of decryption. This process of encryption and decryption is performed using cryptographic keys. Since only the holder of the cryptographic key can perform these operations, it is possible to encrypt data using the encryption key when one does not want others to know the contents of the data, or to allow decryption by giving the encryption key to the party to whom one wants to inform.[7]

While encrypted data cannot be used unless it is decrypted to plaintext, data in plaintext is vulnerable to privacy violations. Since the 1980s, research has been conducted on technologies that make it possible to perform various processes in encrypted form. While encryption protects privacy during data transfer and storage, the technology of processing in encrypted form can be said to protect privacy during the process of data analysis and operations as well.[8]

The method of processing data and obtaining analysis results while maintaining a certain level of data confidentiality among multiple entities is called secure computation (also known as multi-party computation or secure multi-party computation), and has been gaining

[7] Key transfer protocols with cryptography such as public key cryptography are used for such key deliveries. See footnote 13 for more information on public key cryptography.

[8] For more information on the technology of processing in encrypted form, see:
Seito, Takenobu, and Junji Shikata, "kōkinō-angō wo katsuyō shita jōhō-rōei taisaku 'angōka-jōtai shori gijutsu' no saishin dōkō", *Kinyū Kenkyū*, vol. 33, no. 4, pp. 97-132, 2014.

interest in recent years due to its high versatility.

2. Overview of secure computation and its implementation

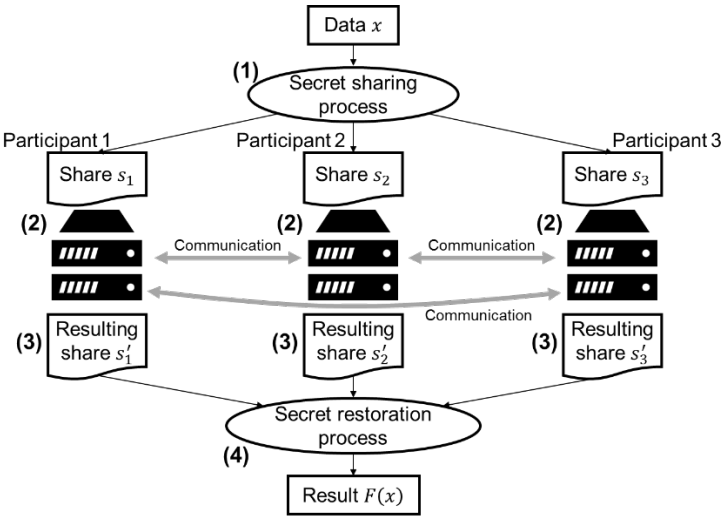
There are two main techniques used to implement secure multi-party computation: secret sharing schemes and homomorphic encryption.

a. Secret sharing schemes

Secret sharing schemes are technologies which manage secrets by distributing it into multiple entities which work together. In this scheme, data is divided into multiple fragments called "shares". The original data can only be restored when a certain number of shares are collected. Due to this property, information about the original data cannot be obtained unless a certain number of participants collude to gather the shares.[9] After allocating these shares to multiple servers and having them perform computation in the form of shares, the computation results, also in the form of shares, at each server are aggregated to derive the overall result. Throughout the entire process, each server is only exposed to its allocated shares and thus has no knowledge of either the original data or the overall results. An example of a specific process is in Figure 5.

[9] For example, in a secret sharing scheme called the (k, n) threshold scheme, when n shares are created and assigned to n servers, even if k-1 of the servers are compromised or colluded, the information in the original data cannot be recovered.

Figure 5: Secure multi-party computation using secret sharing method



(1) The entity providing the data (client) divides the input data x into multiple shares (secret sharing process), sends them to the participants (servers), and requests a computation.

(2) The participant (server) receives the shares s_n and performs the computation (communicating with other participants as necessary).[10]

(3) The participant (server) obtains the resulting shares s'_n.

(4) The client collects the resulting shares (secret restoration process) and restores the final result $F(x)$ from them.

Note: Created based on various sources.

In this case, the entity providing the data divides the data into shares, allocates the shares to the data analysis service providers, and receives the shares of the analysis results. The entity providing the data can obtain the analysis results without disclosing the data to the analysis providers.[11]

b. Homomorphic encryption

The alternative method for secure computation is *homomorphic encryption*, which allows

[10] Computations are performed with the shares as input, and operations such as addition and multiplication are performed without restoring the data. Some computations require procedures in which a part of the shares are mutually shared between specific participants, and for this purpose, communications take place between the participants.
Ohara, Kazuma, "Himitsu bunsan-hō wo mochiita himitsu-keisan," *Systems, control and information*, vol. 63, no. 2, pp. 71-76, 2019.

[11] Figure 5 represents a model in which data held by a single entity is used as input and analyzed, but there are other models, such as one in which data held by different organizations in a distributed manner is used as input. The organizations collaborate to analyze the data while keeping it secret from each other.

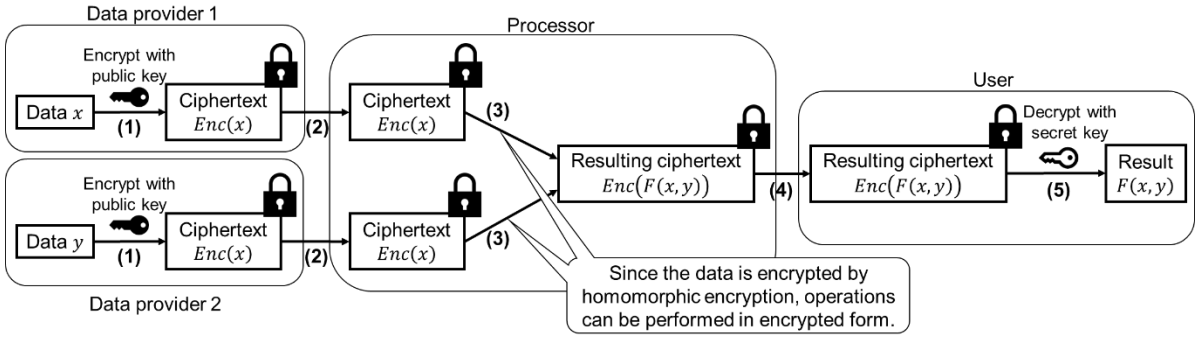
operations such as addition and multiplication to be performed while the data is encrypted. With this type of encryption, a certain operation is performed on the ciphertext of a value, and the plaintext obtained when the resulting ciphertext is decrypted matches the result of the operation on the original value.[12]

As a model case of secret computation using homomorphic encryption, consider a situation in which a data provider passes data to a processor, the processor performs a computation, and the result is provided to a user (Figure 6). In this example, by using public key cryptography[13] with the above properties of homomorphic encryption to process the original data in encrypted form, the data provider can provide only the processing results to the user without sharing the original data with other entities (other data providers, processors, and users). Since this is a mechanism that uses encryption keys, it is important that the keys are managed in a manner that allows the entities to agree on the handling of keys.

[12] For any plaintext m_1, m_2, for example, with respect to addition, such that $Enc(m_1) \oplus Enc(m_2) = Enc(m_1 + m_2)$, the operation $\oplus$ exists, and the additive method can be computed for the original data without encryption. Here, $Enc(\cdot)$ is the ciphertext of input (plaintext). An encryption scheme that allows both additive and multiplicative operations is called a fully homomorphic encryption.

[13] Public key cryptography is an encryption scheme that uses two different keys: a public key and a private key. For example, the recipient of the information creates two keys and only gives the public key to the sender. The sender encrypts the information with the public key and sends it, and only the recipient with the private key can decrypt the information.

Figure 6: Secret computation using homomorphic encryption



(1) Data providers 1 and 2 encrypt data x and y, respectively, with their public keys and generate ciphertext $Enc(x)$ and $Enc(y)$.

(2) Each data provider sends the ciphertext $Enc(x)$ and $Enc(y)$ to the processor.

(3) The processor performs the operation on the ciphertext and obtains the resulting ciphertext $Enc\big(F(x,y)\big)$.

(4) The user obtains the resulting ciphertext $Enc\big(F(x,y)\big)$ from the processor.

(5) The user decrypts the resulting ciphertext with the secret key to obtain the result $F(x,y)$.

Note: Created based on various documents.

D. TEE: Computation under Hardware Data Protection

In a *trusted execution environment (TEE)*, a software is executed securely in an isolated and secure area. The functions and protection schemes provided by a TEE depend on the underlying processor. In the case of Intel's Software Guard Extensions (SGX), the software is loaded and executed in a cryptographically protected memory area called an enclave. In memory and on the system bus (data transmission paths), programs and data are encrypted and can only be decrypted and executed with the cryptographic key unique to the processor.

This feature of a TEE allows a user's intended program to run correctly while being isolated from the outside world, even in environments where OS vulnerabilities or malware may exist, or on hardware that is hosted externally, such as a cloud. This property of a TEE can be utilized to share data and logic among multiple parties and to collaborate with them in a privacy-preserving manner.

When a TEE is incorporated in a system, its trustworthiness need to be verifiable from the outside. For this purpose, a technique called *remote attestation* is used to verify over the network that the target environment is not a device forged as a TEE and that the code to be run on it matches what the user intends.

It should be noted that aspects of TEEs are highly dependent on hardware vendors, such as hardware key management and identification mechanisms.

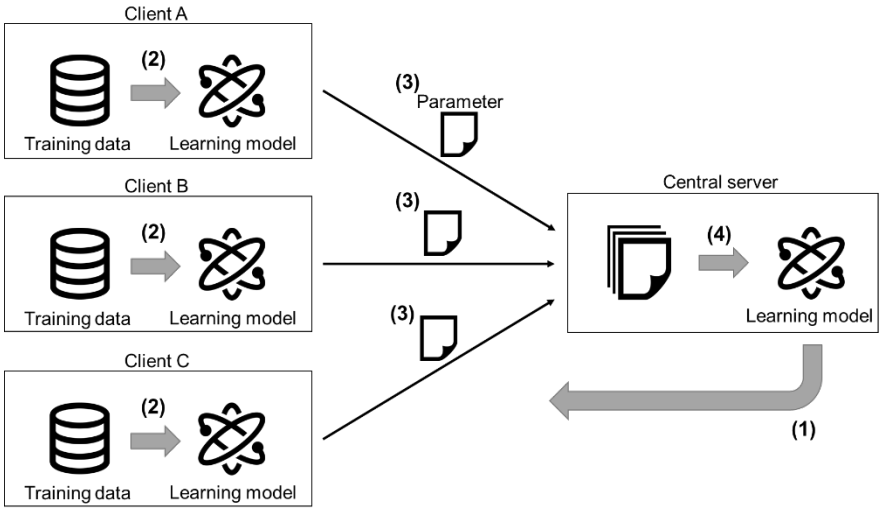
E. Federated Learning: Privacy-Aware Machine Learning

In recent years, research has been progressing on *federated learning*, in which data is used for machine learning while taking privacy into consideration. In addition, to enhance privacy in federated learning, methods leveraging the aforementioned differential privacy and secure computation are also being studied.

1. Federated learning[14]

Federated learning is a method of machine learning in which data is distributed without aggregation. An example of a federated learning process in a client-server model is illustrated in Figure 7.

Figure 7: Process of federated learning



(1) Each client (e.g., smartphone) obtains the latest learning model from the central server.
(2) Each client performs machine learning on the basis of its own data and improves the parameters of the learning model.
(3) Each client sends the improved parameters to the central server.
(4) The central server updates the learning model on the basis of the aggregated parameters.
(5) Return to (1)

Note: Based on Kairouz et al. (2021).

[14] The description in this section is based on the following study.
Kairouz, Peter et al., "Advances and Open Problems in Federated Learning," Foundations and Trends® in Machine Learning, vol. 14, no. 1–2, pp. 1-210, 2021.

In federated learning, the original data remains distributed across multiple clients, and only the parameters of the learning model extracted by each client are integrated at the central server. In terms of privacy protection, the advantage over general machine learning is that the original data held by the client remains with the client.

2. Privacy-preserving federated learning[15]

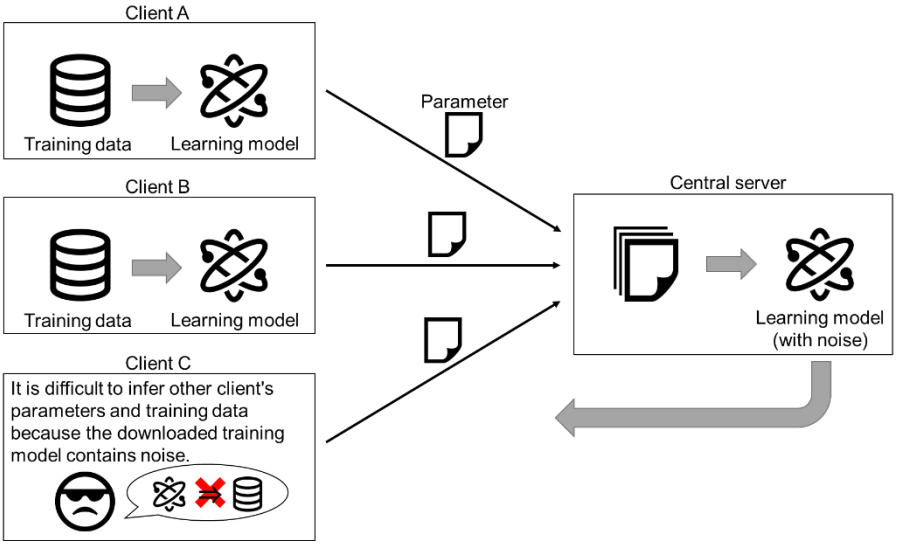
In the above-mentioned federated learning, the concern remains that the client or central server may infer the original training data from the information obtained through a series of processes. To counter these threats, privacy-preserving federated learning has been proposed, which uses the concept of differential privacy and secure computation to provide countermeasures.

a. Countermeasures against attacks originating from clients (federated learning with differential privacy)

A client participating in federated learning or an attacker who obtains information by attacking a client may be able to infer the training data of other clients from the learning model provided by the central server. One possible way to reduce this risk is for the central server to provide learning models that satisfy differential privacy. That is, noise is added to the learning model provided to each client by the central server to satisfy differential privacy. This makes it difficult to infer the training data of others from the learning model for a malicious client or for an attacker who obtains information from a client (Figure 8).

[15] The description in this section is based on the following reference.
Li, Tian et al., "Federated Learning: Challenges, Methods, and Future Directions," IEEE Signal Processing Magazine, vol. 37, issue 3, pp. 50-60, 2020.

Figure 8: Application of differential privacy to federated learning

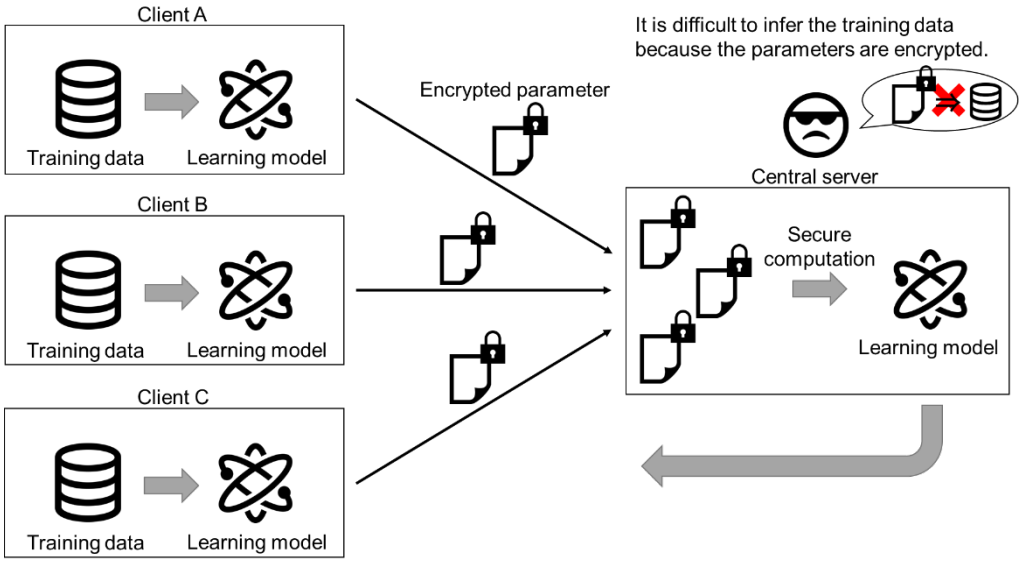


Note: Based on Li et al. (2020).

b. Countermeasures against attacks originating from central server (federated learning with secure computation)

A central server participating in federated learning, or an attacker who obtains information by attacking the central server, may be able to infer the training data of the clients from the parameters sent by each client. One way to reduce this risk is by using a method that introduces secure computation into federated learning. That is, the client encrypts its parameters and sends them to the central server, which uses them in encrypted form in the computation of the learning model. This makes it difficult at the central server to know the plaintext of the encrypted parameters and to infer the training data (Figure 9).

Figure 9: Use of secure computation for federated learning



Note: Based on Li et al. (2020).

F. Other Technologies and Concepts Related to Privacy Enhancement

In addition to the technologies described thus far, there are many other techniques and concepts that contribute to privacy enhancement. For example, as the digital domain becomes increasingly important, there is a growing interest in how digital identities are managed. Self-sovereign identity has been proposed as a new identity management model, and since it is considered to be closely related to privacy enhancement, it is addressed in Box 1 at the end of this paper.

III. Applications Related to Payments and Financial Services

This section introduces attempts by financial institutions and public organizations to apply the privacy protection technologies described in Section 2 to payments and financial services.

A. Provision of Synthetic Data

Synthetic data is artificial data generated from original data. By statistically reproducing the characteristics and trends of the original data using algorithms, the use of the original data itself can be avoided. The Financial Conduct Authority (FCA) in the UK offers companies synthetic data that can be used to experiment with product development during

TechSprints,[16] a workshop organized by the FCA for financial institutions and other companies. The July 2019 session was focused on money laundering and financial crime prevention[17] and provided synthetic data created with the help of outside organizations.

B. Privacy-Aware Machine Learning and Its Application to AML/CFT[18]

In Japan, there have been cases in recent years where financial institutions introduced automatic fraud detection systems using AI technology. While the effectiveness of AI-based fraud detection should increase through cooperative learning carried out by multiple financial institutions, there are many challenges when taking financial transaction data including customer information outside of each financial institution. To address this issue, the National Institute of Information and Communications Technology (NICT), Kobe University, Eltes, MUFG Bank, and other financial institutions are working on an experiment to develop a system that can automatically detect fraudulent remittances and other irregularities while protecting privacy. This system uses the NICT's technology, which enables machine learning without disclosing data that includes customer information to outside parties.

Specifically, the aforementioned institutions experimented on the machine learning performed on data held by each party, in which the parameters (gradient information) locally trained in each party are encrypted and sent to the central server, and the global parameters (weights) of the learning model are updated without decrypting the sent parameters. This updating process uses homomorphic cryptography, which allows addition in encrypted form. Each party can download the parameters updated at the central server for more accurate analysis. Here, the parameters are statistical information that aggregates multiple data, making it difficult to identify individuals, and, because they are encrypted, the risk of the central server inferring each party's training data from the parameters is reduced (Figure 10).[19]

---

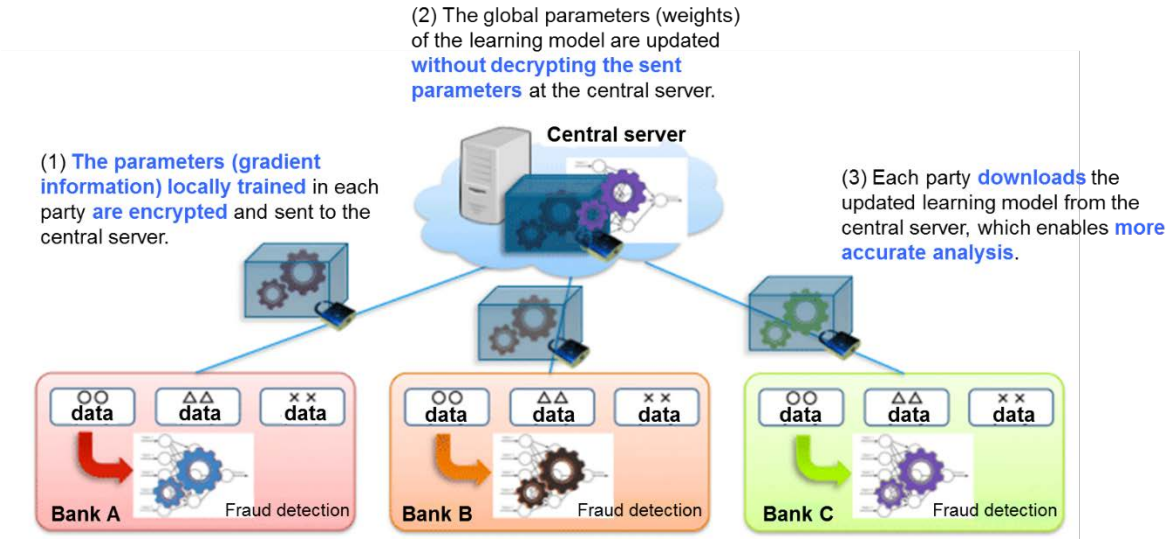
[16] FCA, "TechSprints." https://www.fca.org.uk/firms/innovation/regtech/techsprints

[17] FCA, "2019 Global AML and Financial Crime TechSprint,"
https://www.fca.org.uk/events/techsprints/2019-global-aml-and-financial-crime-techsprint

[18] The statements in this section are from press releases and various media reports from the NICT.
https://www.nict.go.jp/press/2020/05/19-1.html

[19] For details, see:
Phong, Le Trieu et al., "Privacy-Preserving Deep Learning via Additively Homomorphic Encryption," IEEE Transactions on Information Forensics and Security, vol. 13, issue 5, pp. 1333-1345, 2018.

Figure 10: Fraud detection model using machine learning



Source: National Institute of Information and Communications Technology

C. Using TEE for Secret Computation for Marketing Purposes[20]

Royal Bank of Canada (RBC) has built Virtual Clean Room (VCR), an information sharing platform that takes data confidentiality into consideration, in an environment provided by Microsoft (Azure confidential computing), and conducted an experiment in which analysis is performed on the data distributed across the platform. VCR protects data in use by performing computations in hardware-based trusted execution environments.

For example, when a customer makes a purchase at a retail store using an RBC credit card, RBC obtains data showing which retailer the customer purchased from but does not have data regarding items the customer purchased at that retailer. The retailer acquires data showing items the customer purchased at its own store but does not want to share it externally to maintain the confidentiality of customer data. By using VCR, RBC can conduct an analysis that combines the data with its own data without touching the contents of each retailer's individual customer purchase data, and only the results of that analysis can be acquired and utilized.

[20] The description in this section is from the Microsoft website.
Microsoft, "RBC creates relevant personalized offers while protecting data privacy with Azure confidential computing," https://customers.microsoft.com/ja-jp/story/1356341973555285762-royalbankofcanada-banking-capital-markets-azure

IV. Conclusion and Implications for Digital Currencies

This paper has provided an overview of privacy enhancing technologies and their applications in relation to payments and financial services. There are certain points to keep in mind towards planning future social implementation of these technologies.

First, many of these technologies are only in the experimental stage, and they are still in the process of being researched.[21] For example, secure computation requires more time compared to computation in plaintext for communication and processing. Self-sovereign identity is at the stage where various organizations are working on related standards and specifications, and there are only a few cases put into practice. However, it is believed that these technologies will not only lead to the creation of business opportunities in various domains but may also become an important foundation for supporting privacy in a digital society. A wide range of entities, including public institutions, research institutes, financial institutions, and service providers, are conducting research, development, and experiments, and further progress is expected.

Even if privacy enhancing technology progresses to a level where it can be implemented in society, technology alone cannot solve all problems. It is important to consider that effective privacy enhancing mechanisms can only be achieved by applying technologies in conjunction with various mechanisms, such as strong privacy and information security policies, governance, and operational frameworks.

In addition, discussions on privacy enhancing technologies in a digital society can also provide implications for studies on Central Bank Digital Currency (CBDC), which have been underway in many countries, including Japan.

In general, payment service providers obtain data from users at the customer onboarding, such as account opening, and at the time of individual transactions, such as remittance. In CBDC, the division of roles between the central bank and private operators, i.e., who should obtain and manage data, to what extent, and under what conditions, should be determined while considering the requirements regarding the handling of user information.

---

[21] The San Francisco Fed's staff survey paper on Privacy Enhancing Technology (PET) also states that "the development and use of PETs is in early stages."
Asrow, Kaitlin, and Spiro Samonas, "Privacy Enhancing Technologies: Categories, Use Cases, and Considerations," FinTech Edge Special Report, Federal Reserve Bank of San Francisco, 2021.

There have been various international discussions related to CBDC and privacy. "Public Policy Principles on Retail Central Bank Digital Currencies (CBDC)" released in October 2021 by the G7, states that in order to gain trust and credibility, any CBDC should have "rigorous standards of privacy, accountability for the protection of user's data, and transparency on how information will be secured." In addition, it notes the need for "commitment to mitigate their use in facilitating crime" and that the design of CBDCs should strive to incorporate "advancements in technology and/or innovative solutions that may improve the authentication and verification of transactions" into measures for combating illicit finance (Figure 11).

Figure 11: Public Policy Principles for Retail Central Bank Digital Currencies (CBDCs) (excerpts)

Principle 3. Data privacy	Rigorous standards of privacy, accountability for the protection of users' data, and transparency on how information will be secured and used is essential for any CBDC to command trust and confidence. The rule of law in each jurisdiction establishes and underpins such considerations. - Public and private sector entities in any CBDC ecosystem should only access, hold, process, or share users' personal data where this data is necessary to achieve clear, open and legal purposes, for example, to mitigate money laundering or terrorist financing (ML/TF) risks. - Users of any CBDC should have a high degree of transparency regarding the use of their personal data, centered around the principles of data minimisation and control for the user (wherever possible). Access to individual users' data beyond the minimum required should be supported by a strong consent framework where entities (public and private) should clearly and transparently lay out those additional requirements that are necessary to providing a viable and functional service.
Principle 6. Illicit finance	Any CBDC needs to carefully integrate the need for faster, more accessible, safer and cheaper payments with a commitment to mitigate their use in facilitating crime. - CBDCs and accompanying regulatory frameworks should commit to countering their use in facilitating crime and be designed to comply with anti-money laundering (AML), counter-terrorist financing (CTF), and counter-proliferation of weapons of mass destruction obligations. They should also mitigate the risk of evasion of financial sanctions and comply with Financial Action Task Force (FATF) Standards. - When designing CBDCs, public authorities should look to build in ways of countering fraud and other forms of illicit finance by looking to incorporate advancements in technology and/or innovative solutions that may improve the authentication and verification of transactions. - Public authorities' powers and use of data in any CBDC ecosystem to counter their use for illicit finance should be set out transparently in national legislative frameworks and those powers should not be used for other purposes.

Within the international discussions among central banks, particularly in the G7 countries, there are many references to the importance of balancing privacy with other public interests,

such as AML/CFT, while also explicitly stating that privacy is an important issue regarding CBDC (Figure 12). In addition, in communicating externally, central banks, on many occasions, have explicitly expressed that they have no interest in utilizing information for purposes other than ALM/CFT.[22]

[22] In light of this, some central banks, such as MAS, have stated that they "could thus offer a CBDC payment system that confers a higher degree of privacy to consumers by default compared to existing models of electronic payment, while ensuring protections against illicit money flows."
Monetary Authority of Singapore, "A Retail Central Bank Digital Currency: Economic Considerations in the Singapore Context," pp.26, 2021.

Figure 12: Communications of major central banks on privacy aspect of CBDC

CBDC and privacy fundamentals	Euro area	Privacy emerges as the most important feature of a digital euro. Protecting user's personal data and ensuring a high level of confidentiality will therefore be a priority in our work, so that the digital euro can help maintain trust in payments in the digital age (ECB Panetta (2021), Member of the Executive Board).[23]
	U.S.	Protecting consumer privacy is critical (if CBDC is introduced) (FRB public consultation paper (2022)).[24]
	U.K.	The Bank (BOE) recognises that privacy is a critical consideration in any CBDC system, and that appropriate privacy must be ensured if CBDC is to command users' trust and confidence (BOE (2021)).[25]
Privacy and its relationship to other public interests such as AML	Euro area	Privacy is an important prerogative because it influences people's personal lives and fundamental rights. It must nonetheless be carefully assessed against other important considerations in the general interest. Digital euro payments could guarantee different degrees of privacy, involving different trade-offs with other policy and regulatory objectives such as the need to combat illicit activities (ECB Panetta (2021), Member of the Executive Board).
	U.S.	Any CBDC would need to strike an appropriate balance, however, between safeguarding the privacy rights of consumers and affording the transparency necessary to deter criminal activity (FRB public consultation paper (2022)).
Central bank interest in collecting payment data	Euro area	As a public and independent institution, the ECB has no interest in monetizing or even collecting users' payment data (ECB Director Panetta (2021)).
	U.K.	BOE has no commercial incentive to gather user's data. ...if we were to launch a CBDC, the central bank should only collect the very minimum set of information needed to operate the system and meet our legal and compliance obligations (BOE Mutton (2021), Director).[26]
	Singapore	MAS is not driven to collect and utilize transaction data or personal information for profit motives (MAS (2021)).

[23] Panetta, Fabio, "A digital euro to meet the expectations of Europeans – Introductory remarks by Fabio Panetta, Member of the Executive Board of the ECB, at the ECON Committee of the European Parliament, Frankfurt am Main, 14 April 2021," https://www.ecb.europa.eu/press/key/date/2021/html/ecb.sp210414_1~e76b855b5c.en.html, 2021.

[24] Board of Governors of the Federal Reserve System, "Money and Payments: The U.S. Dollar in the Age of Digital Transformation," 2022.

[25] Bank of England, "Responses to the Bank of England's March 2020 Discussion Paper on CBDC," https://www.bankofengland.co.uk/paper/2021/responses-to-the-bank-of-englands-march-2020-discussion-paper-on-cbdc, 2021.

[26] Mutton, Tom, "Central Bank Digital Currency: An update on the Bank of England's work," https://www.bankofengland.co.uk/speech/2021/june/tom-mutton-pre-recorded-keynote-speech-the-future-of-fintech-digital-conference, 2021.

Some foreign central banks have expressed hope for the technologies discussed in this paper as a means of achieving both privacy protection and AML/CFT. For example, the BOE stated that "privacy enhancing technologies, such as zero-knowledge proofs, and digital identity frameworks, could offer opportunities to both enhance transparency and increase security and privacy" (BOE Mutton (2021)[26]).[27] Nevertheless, they are also aware of the computational burden of the new technologies and the impact it may have on CBDC performance (BOE (2020)[28]). These are the same perceptions as those discussed above in this paper.

The above discussions related to privacy will be very important in the process of considering CBDCs in our country. It would be meaningful to follow the development trends and attempts at implementing the technologies that may contribute to privacy enhancement, as mentioned in this paper.

Whether or not a CBDC should be issued in Japan would be a judgment by the people in the future, taking into account domestic and global trends. Nevertheless, the Bank of Japan is steadily proceeding with its technical experiments and explorations into institutional arrangements to ensure the stability and efficiency of the entire payment and settlement system and to be able to respond appropriately to future changes in the environment. As part of its explorations into the institutional arrangements, the Bank will continue to conduct research and study on privacy protection related to digital currencies with a wide range of stakeholders.

[27] For zero-knowledge proofs, see Box 2.

[28] Bank of England, "Central Bank Digital Currency: opportunities, challenges and design," 2020.

(Box 1) Self-Sovereign Identity: Identity Controlled by Oneself

In payment and settlement services, it is necessary to provide the right service to the right person, and the handling of the user's identity is important for this purpose. Nowadays, as services are often provided via smartphones and networks, it is particularly important to consider how to handle the user's identity expressed in digital form (digital identity). In the following, we introduce *self-sovereign identity* as a new management model for digital identity and demonstrate the importance of this concept in terms of user privacy enhancement.

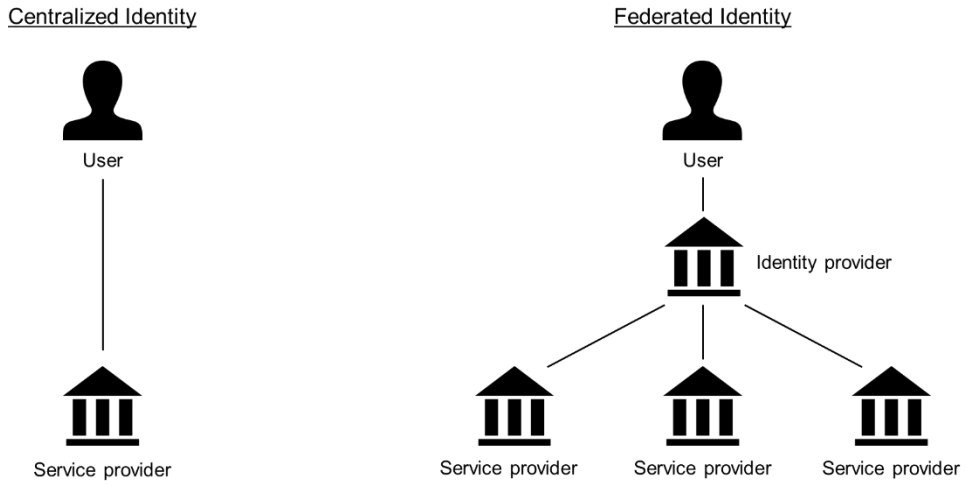
(1) Existing identity management models and self-sovereign identity

The ISO/IEC, an organization that promotes international standardization, defines an identity as a "set of attributes related to an entity."[29] "Attributes" here refer to items such as personal information (e.g., the user's name), credentials, and rights. Therefore, an identity can be thought of as a set of attributes such as personal information and credential information related to an individual. A digital identity is a digital representation of such information that can be processed on a computer.

The first model for managing digital identity is one in which a service provider directly manages the user's identity (Centralized Identity; Figure B1-1, left). In this model, the user's identity is only used by the service provider. With the advancements in digitalization and increasing number of services available, there is a growing social need to use and provide identities across multiple services. The method of linking a user's identity to multiple service providers by the identity provider selected by the user has become prevalent in this context (Federated Identity; Figure B1-1, right). In common practice, when using a service, the provided identity is stored by the service provider, who then links it to other service providers as an identity provider. In this model, the user's identity is now used for a wide range of services rather than a single service, and, at the same time, the identity providers become the hubs, resulting in a situation where users' identities are dependent on a small number of identity providers.

[29] ISO/IEC 24760-1, https://www.iso.org/obp/ui/#iso:std:iso-iec:24760-1

Figure B1-1: Typical identity management model



Note: Based on Reed (2018).[30]

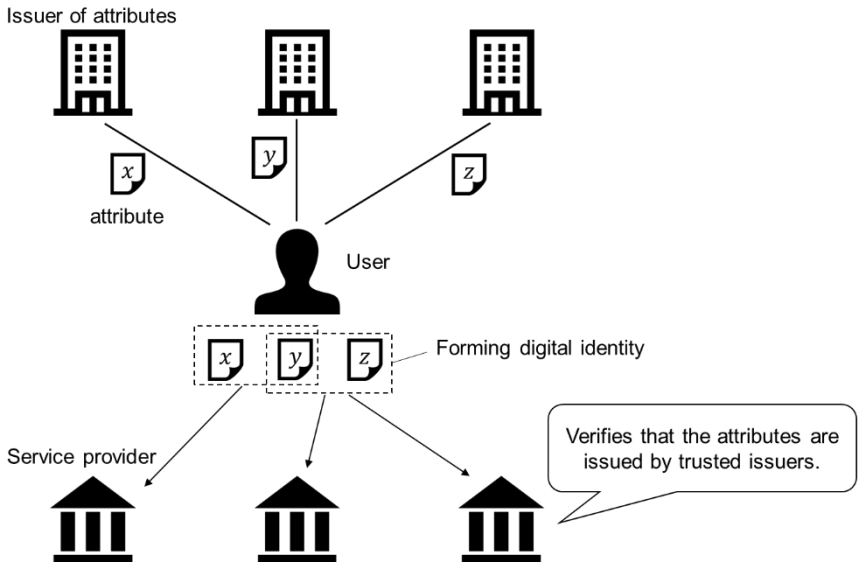
Thus, the currently prevailing models for managing digital identity are based on the assumption that identities are dependent on a particular organization. Such models entail the following risks: (1) that the provision or linkage of identities will be suspended due to the circumstances or decision of a particular organization, (2) that identities are linked against the users' will due to negligence or willful intent, and (3) data tampering.

Self-sovereign identity (SSI) is attracting attention as a digital identity management concept that can mitigate such risks. The concept of SSI is that an individual controls their identities without the intervention of an administrative organization.[31] This can be structured as shown in Figure B1-2.

[30] Reed, Drummond, "The Story of SSI Open Standards," https://ssimeetup.org/story-open-ssi-standards-drummond-reed-evernym-webinar-1/, 2018.

[31] Sovrin Foundation, https://sovrin.org/faq/what-is-self-sovereign-identity/

Figure B1-2: Example of self-sovereign identity structure



- Entities trusted by the service providers are the issuers of the user's attributes (e.g., credentials, background, and rights).
- The user is issued his/her attributes from these entities. Attributes are issued in a state associated with a decentralized identifier generated by the user. The combination of this decentralized identifier and the attributes associated with it form an identity.
- The user provides the identity formed above to the service provider.
- The service provider verifies the distributed identifier and attributes that form the identity by (1) confirming that the identifier belongs to the user using the user's public key or by other means, and (2) verifying that the identifier is associated with the attributes issued by trusted entities. If there are no problems, the service is provided to the user on the basis of the attributes forming the identity provided.

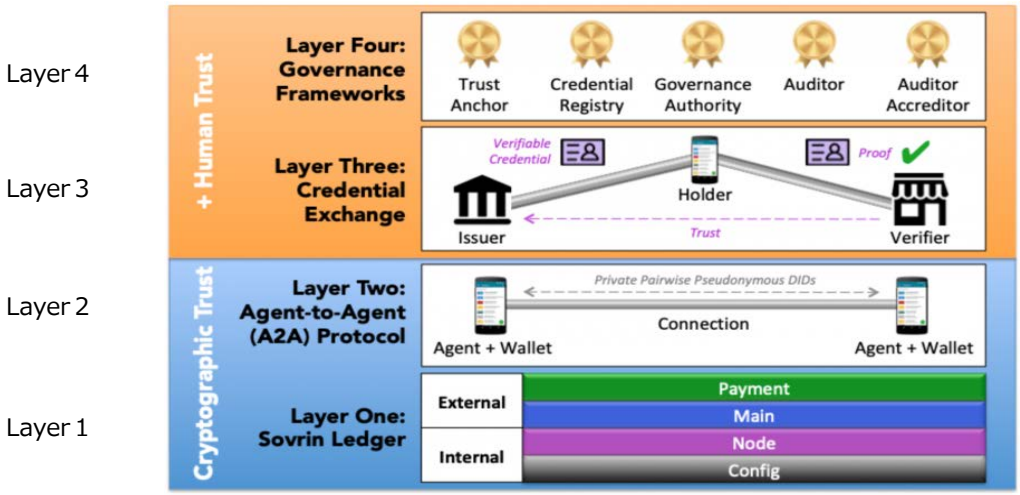
Note: Created based on various documents.

In the example shown in Figure B1-2, users generate their own *decentralized identifiers (DIDs)*, link them to attributes (credentials and personal information), and then combine the attributes to form identities. No organization is needed to centrally manage the correspondence between decentralized identifiers and users, and users can construct and use their own identities without depending on a particular organization.

The Sovrin Foundation, a non-profit organization that aims to implement self-sovereign identity, has organized schemes, such as that shown in Figure B1-2, into four layers (Figure B1-3). In this stack, ledgers that manage decentralized identifiers designed to be self-generated and owned are positioned as the foundational layer (Layer 1). On top of this layer, connections are established between users (Layer 2), and attributes are exchanged in a verifiable manner (Layer 3). The data model for this is called verifiable credentials (VCs) and has been standardized by the World Wide Web Consortium (W3C)

along with decentralized identifiers.[32] Layer 4 is the non-technical layer, which is the governance layer that forms business and legal agreements between issuers so that information such as credentials and personal information can be widely and uniformly handled in society. For Layer 1, the use of distributed ledger technology (DLT) is attracting attention as a possible means to create a structure that does not depend on a particular entity.

Figure B1-3: Hierarchical diagram SSI infrastructure from the Sovrin Foundation



Source: Sovrin Foundation[33]

(2) Self-sovereign identity and privacy enhancement

In the concept of self-sovereign identity, users provide their own information to service providers using a standardized data model. This provides the following features in terms of privacy enhancement.

① Controlling the content of information to be provided

In self-sovereign identity, users provide their information to service providers in accordance with the purpose of use. The information collected by the service provider can be kept to the minimum necessary (data minimization), and the issuers will never know to which service provider the information is provided.

[32] For details, see the World Wide Web Consortium's distributed identifier standards related documents. https://www.w3.org/TR/did-core/

[33] Sovrin Foundation, https://sovrin.org/2020-how-ssi-went-mainstream/

Furthermore, in combination with technologies related to zero-knowledge proofs, a method will likely be devised in which unnecessary items of information are kept secret at the discretion of the user and only the information necessary to fulfill the eligibility conditions for service use is provided without having to re-issue the set of attributes that has already been issued (for zero-knowledge proofs, see Box 2).

② Controlling the destination of the information

Because users provide the attributes themselves, it can be assumed that they are doing so willingly. At least for the primary provision of attributes to service providers, this is not done without the user's involvement.

③ Deterring the linking of data between the parties to which the data is provided

Users can choose a specific identifier that is used in her/his provision of attributes to each service provider, which mitigates the risk of privacy infringement due to user identity unification by collusion of multiple service providers, leveraging a single common identifier.

(3) Cases related to self-sovereign identity

The Province of British Columbia, Canada, has two initiatives closely related to digital identity as part of its digital government initiative: OrgBook BC and Verifiable Credentials for People.[34] OrgBook BC is a live, public registry of business licenses and permits issued by various government agencies to businesses and others as verifiable attributes, searchable through a web service that, at the time of writing this paper, has over 4 million verifiable attributes registered.[35] The service is built on a distributed ledger that handles DIDs/VCs, making identity management independent of any particular organization. Verifiable Credentials for People is a concept to apply identity management using such a distributed ledger to individuals rather than legal entities, and a framework that allows citizens to gain control over their own attributes is being considered.

A relevant case in the financial sector is Verified.Me, also from Canada, which is based on a framework by the Digital Identification and Authentication Council of Canada

[34] Province of British Columbia, "BC Digital Trust," https://digital.gov.bc.ca/digital-trust/

[35] Province of British Columbia, "OrgBook BC," https://www.orgbook.gov.bc.ca/search/

(DIACC), a non-profit organization. Verified.Me was developed by SecureKey Technologies, along with a consortium of seven major Canadian financial institutions, and the service has been provided since May 2019. The consortium of financial institutions operates a distributed ledger that constitutes the identity infrastructure, and users can provide their attributes to service providers by sending instructions to the issuers of those attributes, such as financial institutions, credit bureaus, and telecommunications carriers. Verified.Me is a new service that combines traditional identity (Federated Identity) and self-sovereign identity and is designed to take advantage of both.[36]

[36] Secure Technologies, "A Primer and Action Guide to Decentralized Identity," https://securekey.com/wp-content/uploads/2020/07/VerifiedMe_OWIWhitepaper_APrimertoDecentralizedIdentity.pdf, 2020.

(Box 2) Concept of Zero-Knowledge Proof and Related Techniques

Zero-knowledge proof is a general term for a method in which the prover proves (convinces the verifier) that the prover's claim is true without revealing any other knowledge.[37] It typically takes the form of a problem whose solution is considered extremely difficult to find, such as a discrete logarithm problem, and convincing the verifier that the prover knows the solution without giving any information about the solution.

Zero-knowledge proofs are incorporated into cryptographic techniques such as digital signatures[38] so that they are able to keep certain information secret. Specifically, zero-knowledge proofs are used to achieve strong anonymity in digital signature techniques such as group, ring, and blind signatures (Figure B2-1).

[37] Zero-knowledge proof methods satisfy properties of completeness, soundness, and zero-knowledge.
- Completeness: Given that the prover's assertion is true, the verifier can verify that the assertion is true.
- Soundness: Given that the prover's assertion is false, the verifier can detect with high probability that the assertion is false.
- Zero-knowledge: The verifier has no knowledge other than that the prover's assertion is true.

[38] Digital signatures are a technique that ensures that a document conveyed to others was created by a known document creator and that the content has not been modified. A verifier can verify that the document has not been modified by comparing the document with the content of the signature. The digital signature is made with the secret information of the document's creator, so the document and its creator are linked by the signature.

Figure B2-1: Examples of embodiment of zero-knowledge proofs

Group Signatures Ring Signatures	Group signatures are techniques whereby any member of a preconfigured group creates a signature representing the group without revealing oneself as the signer. The group has an administrator who can identify the signer and break the anonymity of the members when necessary. Ring signatures are a simplified version of group signatures, in which the signer borrows public keys from others to form a group of possible signers, completing the entire signing procedure by her/himself. Unlike group signatures, ring signatures do not require a group administrator. The signer can prove that the document has been endorsed by someone in the group associated with the public keys without revealing that the actual signer is her/himself.
Blind signatures	Blind signatures are techniques that delegate signing a document to an authority or other entity (signer) while keeping the contents of the document secret from the signer. This allows the document creator to obtain a third-party signature that authenticates the document without disclosing the contents to the signer, and preventing the creator from being identified by any third party. Zero-knowledge proofs can be used to strengthen anonymity by only proving that a document is signed without leaving any trace of the interaction between the document creator and the signer, and to tighten anti-counterfeiting property that eliminates opportunities for third parties to sign without the involvement of a valid signer.

The idea of zero-knowledge proofs has gained interest in recent years, in part due to the following privacy protection-related factors. One is the growing interest in data minimization. There are hopes that zero-knowledge proofs will make it possible to eliminate the need to present information that was conventionally required for user authentication and other similar objectives. This would enable more granular selection of information to be presented and lead to enhanced user privacy.

The second is in relation to distributed ledger technology (DLT). Efforts are underway to implement distributed ledgers in many fields. In general, information on a distributed ledger is shared among the participants due to the nature of DLT, which may pose a hurdle to ensuring confidentiality. The introduction of zero-knowledge proof mechanisms is attracting interest because it allows others to verify without having to reveal all of the information to be entered into the ledger. For example, ZCash uses a zero-knowledge proof mechanism called zk-SNARKs, which allows others to verify that there are no inconsistencies in the transaction while keeping transaction information such

as the amount and destination of the money transfer confidential.[39] Ethereum also enables programs (smart contracts) to be executed with secrecy guaranteed by using zk-SNARKs. Such zero-knowledge proof mechanisms give users the option to virtually manage their own data, even in a public distributed ledger, where data is inherently public.

At present, due to the large amount of computation required, technologies based on zero-knowledge proofs have not yet been widely implemented in services, but there is interested in the concept and research is expected to progress.

[39] Ben-Sasson, Eli et al., "Zerocash: Decentralized Anonymous Payments from Bitcoin," IEEE Symposium on Security and Privacy, 2014.