# The Future of Data Distribution and Its Security Technology

*Toshiyuki Miyazawa, Toshinori Fukunaga, Gen Takahashi, Ryo Kikuchi, Seiji Takahashi, and Satoshi Hasegawa*

**Abstract**

With the acceleration of digital transformation, the value of data is ever increasing, while concerns about security risks and privacy are growing stronger. At NTT Secure Platform Laboratories, we intend to address these issues with cryptographic technology and create a world in which data owned by individuals and organizations are provided and used at the minimum level necessary according to the purpose to enable problem solving. In this article, specific initiatives undertaken at NTT laboratories to support secure data distribution in such a world are introduced.

*Keywords: multi-party key sharing, secure-computation AI, anonymous processing*

## 1. The Future of data distribution

With the acceleration of digital transformation (DX) in various fields, it is expected that the digitization of people, goods, processes, etc. of companies and organizations will progress and that advanced analytical processing will make it possible to address issues such as value creation and improving operational efficiency. While the value of data has increased in line with these trends, security risks and privacy concerns have also intensified.

With the globalization of corporate activities and the spread of the cloud and Internet of Things, various and diverse entities (people, terminals, organizations, etc.) have become interconnected. As various types of data are exchanged and shared by these entities, the risk of theft or leakage of the data affecting companies and individuals has also increased. Concerns about legal restrictions and privacy regarding using personal and corporate information for artificial intelligence (AI) and machine learning—which play important roles in DX—are also growing.

To eliminate such risks and concerns, NTT Secure Platform Laboratories wants to use cryptography to create a world in which data concerning products or individuals can be safely exchanged according to the purpose and enable problem solving (**Fig. 1**). The following technologies that support such secure data distribution are introduced in this article: (i) data encryption and related technologies for protecting all communications end-to-end; (ii) secure-computation AI for enabling advanced integrated analysis while protecting corporate secrets and privacy; and (iii) anonymization technology for processing personal data in a manner that does not identify individuals and encourages the use of that information.

## 2. Data encryption and related technologies

Important data used in data distribution can only be exchanged and shared between multiple entities (individuals, organizations, terminals, etc.), and it is important that information not be leaked to other entities. In consideration of the risk of information leakage from service providers recently, it has become increasingly necessary to keep data confidential even from the companies that provide data-distribution services and their system administrators.

To satisfy this requirement, it is desirable that (i) a secret key to the encryption is shared between trusted

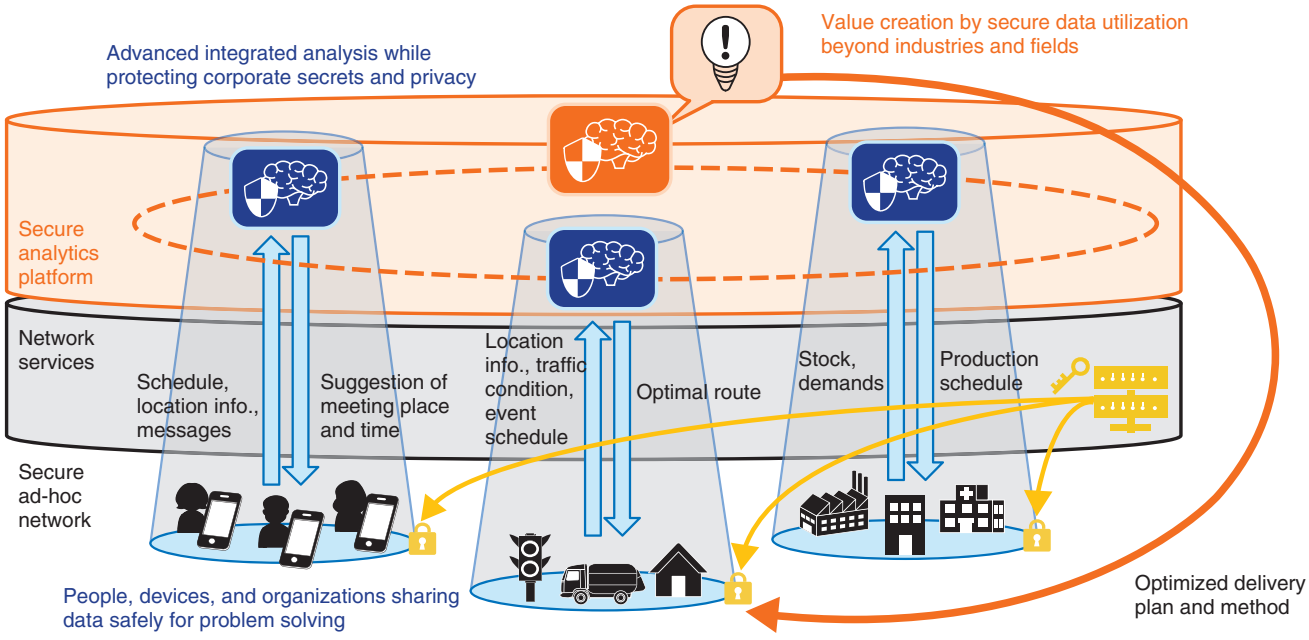



Fig. 1. Secure data distribution.

entities and the data to be exchanged and shared are encrypted with the secret key and (ii) the shared data can be searched. However, enabling these processes faces two major technical challenges. The first challenge is efficient key sharing among multiple entities. It is required to share keys among many entities; however, it is inefficient and impractical to repeatedly execute a key-sharing protocol between two parties using multiple systems.

In light of the above-described circumstance, NTT is studying technology that enables efficient sharing of keys among multiple entities via a key-mediation server installed by a service provider. The key-mediation server guarantees in principle that the shared key cannot be reconstructed. We developed a method for enabling efficient key sharing for a fixed period regardless of the number of entities. This method enables exchange and sharing of data concealed from information-distribution service providers by using only an arbitrary number of entities involved in communication at that time.

The second challenge is to search for encrypted shared data by using the computer resources of the data-distribution service provider. Since data-distribution services are often provided in cloud form, it is desirable to search for shared data by using the computer resources of the data-distribution service provider; however, if the encrypted data are decrypted for such a search, the data cannot be concealed from the service provider. With this issue in mind, NTT is studying a method of enabling this concealed search by encrypting the search index separately from the data and searching with the encrypted search index. As mentioned above, the processing is complicated because shared data are frequently re-encrypted every time an entity is added or deleted. We are thus devising a method of doing this efficiently.

For technology related to the two above-mentioned challenges, we have also developed technology for efficiently re-encrypting data with a new key without encrypting the shared key (encrypted when the shared key was updated) each time an entity is added or deleted. Combining the above-mentioned method and technology, we have already commercialized communication services such as telephony and online chat that do not leak data to service providers [1].

## 3. Secure-computation AI

It is generally necessary to restore (decrypt) the original data during processing for data utilization, even if they were encrypted during communication or storage. Data owners know that this processing may leak information, so many users and organizations are reluctant to use data related to trade secrets and personal privacy. This is considered a major obstacle,

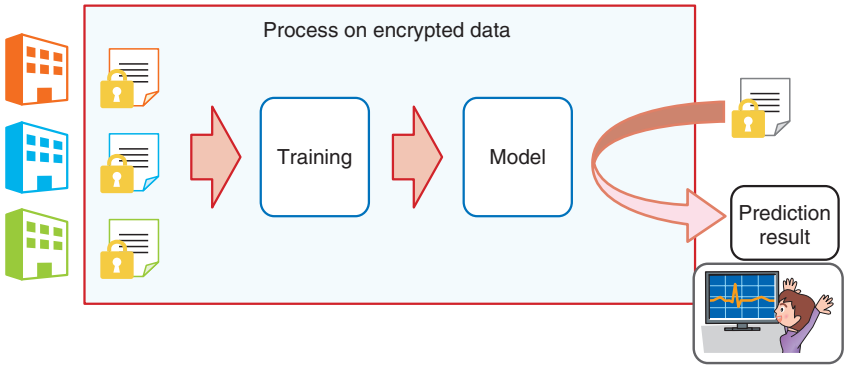



Fig. 2. Secure-computation AI.

especially when data are to be provided, actively used, and transferred from their owner to another party or even distributed within the same organization.

To help eliminate such obstacles, NTT is leading the world in research and development (R&D) of secure-computation technology that can process data while encrypting them. With this technology, data encrypted using the confidential-data-sharing technology standardized by the ISO (International Organization for Standardization) can be analyzed without first restoring them to the original form. This is expected to contribute to the creation of a world in which information (such as confidential information of companies and information concerning personal privacy) can be safely and securely provided and used. This technology has reached the practical stage.

NTT is currently researching and developing secure-computation technology that enables more advanced analysis. We have recently developed a world-first technology that can process a standard algorithm used for deep learning, which has begun to be used in the AI field, without restoring encrypted data to the original data [2]. In other words, all the steps necessary for data utilization in deep learning, namely, data provision, data storage, learning processing, and prediction processing, can be carried out on data in the encrypted state (**Fig. 2**). We believe that this technology will allow data owners to provide data with peace of mind when using the data with AI, leading to an increase in the amount and types of data while enabling advanced analysis with improved accuracy. For example, the technology is expected to make the following possible: (i) by learning personal-location information and schedules in conjunction with weather and corporate-event information, etc., it will be possible to anticipate the most appropriate purchases and staffing resources for restaurants; and (ii) by learning medical data (such as X-rays, magnetic resonance imaging, computed tomography scans, and micrographs) while keeping them concealed, it will be possible to quickly and accurately determine whether malignant tumors are present in test results.

In the future, we plan to demonstrate the effectiveness of deep learning using secure computation by conducting proof experiments in cooperation with partners who have AI expertise.

## 4. Anonymization technology

The use of personal data has recently become the focus of attention, and the market for such use is about to be activated in earnest. Under such circumstances, NTT is researching data-processing technology that promotes safe use of personal data.

According to the revised Act on the Protection of Personal Information of 2017, “anonymously processed information” (namely, personal information processed in a manner that a specific individual cannot be identified and the personal information cannot be restored) can be provided to third-parties and used for tasks other than the specified purpose without the consent of the person in question. For example, by allowing manufacturers to use anonymously processed information made from purchase-history data possessed by retailers, the manufacturers could develop new products in accordance with consumer attributes and purchasing habits.

If only anonymity is to be enhanced, the characteristics of the original data will be greatly impaired, and the data will be less useful. Accordingly, to satisfy

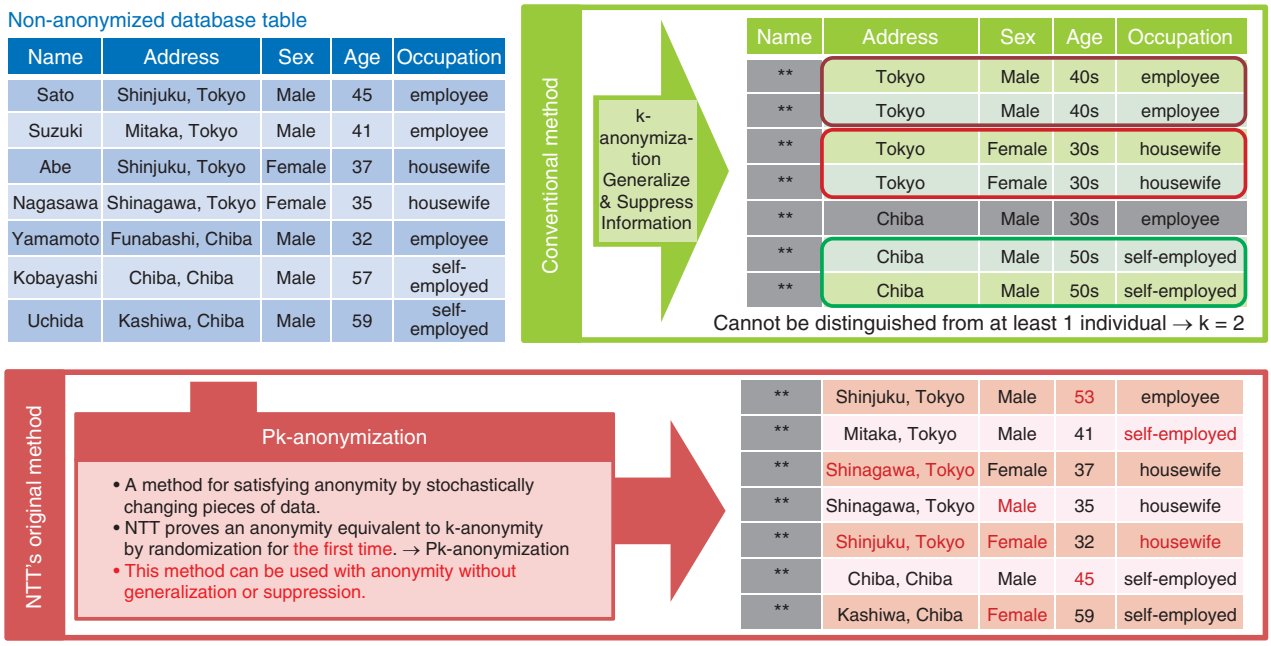



Fig. 3. Pk-anonymization.

both the data owner's need to reduce the risk of leaking personal identification and the data user's need to obtain data that retain the characteristics of the original data, optimal anonymous processing is needed. NTT has thus developed anonymously processed information-creation software for creating anonymously processed information that balances anonymity and usefulness. This software has been available in Japan from NTT TechnoCross since 2018, and its market development is progressing mainly in the medical and financial fields.

This software features various anonymization and evaluation techniques to comply with Processing Standards Nos. 1 to 5 for anonymously processed information specified by the Personal Information Protection Committee. One of these techniques is NTT's original anonymization technique called "Pk-anonymization," which ensures high usability without changing the data granularity by using a perturbative method. As the conventional technology, K-anonymization secures k-anonymity* by abstracting data by for example, changing "33 years old" to "30s" and "Chiyoda ward, Tokyo" to "Tokyo"; however, it is affected by information loss. In contrast, when Pk-anonymization (which disturbs data) is introduced, no information is lost, and more accurate and broader analysis is possible (**Fig. 3**).

With the aim of creating a world in which data are more actively used, NTT is working on reducing the risk of leaking personal identification as well as of inferring personal attributes. For data utilization, it is problematic that when calculation results such as statistical information are used, for example, if the average test scores of two people are used, a person who knows the score of one of them can estimate the score of the other. Privacy concerning such calculation results is called "output privacy," which has been widely studied in statistics. NTT is focusing on output privacy of machine learning, which is a promising data-analysis technology, and continuing to research risk-analysis and protection technologies related to machine learning.

## 5. Future directions

To enable secure data distribution, as introduced at the beginning of this article, NTT is designing and developing various encryption technologies in addition to those introduced in this article. Examples of these efforts are (i) development of cryptosystems and evaluation of their security in anticipation of the creation of quantum computers (the R&D of which is

* K-anonymity: The characteristic that the corresponding individual cannot be identified from the processed data with a probability of 1/k or more.

accelerating) and (ii) cryptographic program obfuscation, which makes the processing content of a program cryptographically non-analyzable and enables distribution of secure programs [3]. On the basis of our expertise in cryptographic theory and technology, we will conduct R&D on data security that contributes to the resolution of problems of NTT Group's customers and social issues.

## References


[1] R. Yoshida, Y. Okano, H. Okuyama, and T. Kobayashi, "A Secure Business Chat System that Prevents Leakage and Eavesdropping from the Server by Advanced Encryption Technology," NTT Technical Review, Vol. 15, No. 5, 2017.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201705fa4.html

[2] Press release issued by NTT, "Secure Computation for a Typical Training Algorithm of a Deep Neural Network," Sept. 2, 2019.
https://www.ntt.co.jp/news2019/1909e/190902a.html

[3] K. Xagawa, "Research Trends in Quantum-resistant Cryptography," NTT Technical Review, Vol. 17, No. 3, pp. 22–26, 2019.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201903fa4.html



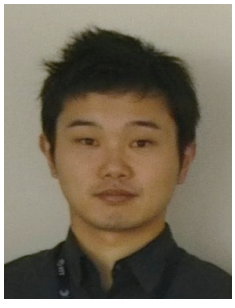
**Toshiyuki Miyazawa**
Manager, Research and Development Planning Department, NTT.
He received a B.E. and M.S. in mathematics from Waseda University, Tokyo, in 2000 and 2003. Since joining NTT in 2003, he has been engaged in R&D of cryptography and security management.

**Toshinori Fukunaga**
Senior Manager, Human Capital Management Group, Research and Development Planning Department, NTT Secure Platform Laboratories.
He received a B.E. and M.E. in electrical engineering from Chiba University in 1996 and 1998 and joined NTT in 1998. He has been engaged in cryptographic engineering and information security management system.

**Gen Takahashi**
Senior Research Engineer, Secure Data Sharing Project, NTT Secure Platform Laboratories.
He received a Master of media and governance from Keio University, Tokyo, in 2005 and joined NTT in 2006. His research interests include information security and cryptographic engineering. He received the Symposium on Cryptography and Information Security (SCIS) Paper Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2008.

**Ryo Kikuchi**
Research Engineer, NTT Secure Platform Laboratories.
He received a Dr. Eng. from Tokyo Institute of Technology in 2015. Since joining NTT in 2010, he has been engaged in R&D for secure data sharing through secure multiparty computation, anonymization, and statistical disclosure control. He was a visiting researcher at the National Statics Center from 2016 to 2018 and has been an expert in ISO/IEC JTC 1/SC 27/WG 2 since 2013. He received ACISP (Australasian Conference on Information Security and Privacy) 2019/ Computer Security Symposium (CSS) 2012/ CSS2013/CSS2017 best paper awards and SCIS 2017 innovation paper award.

**Seiji Takahashi**
Senior Research Engineer, NTT Secure Platform Laboratories.
He received a master's degree in science and engineering from Ehime University in 1999. Since he joined NTT in 2000, he has been engaged in system development applying information security technology.



**Satoshi Hasegawa**
Researcher, Data Security Project, NTT Secure Platform Laboratories.
He received a Master of computer science from University of Tsukuba, Ibaraki, in 2014.
Since joining NTT Secure Platform Laboratories in 2014, he has been engaged in R&D of enhanced privacy-preserving technologies, especially of anonymization and secure multi-party computation. He received the CSS Encouragement Award and DICOMO (Multimedia, Distributed, Cooperative, and Mobile) Symposium Best Paper Award from the Information Processing Society of Japan in 2016 and 2017.