R&D on Secure Computation Technology for Privacy Protection

Koji Chida, Dai Ikarashi, Teruko Miyata, Hiroyoshi Takiguchi, and Naoto Kiribuchi

Abstract

Demand is growing for a means of securely processing highly confidential data on the cloud. To meet this demand, the NTT Secure Platform Laboratories has developed practical, fast, multi-party secure computation technology that provides both confidentiality and usability. This article presents some background issues concerning this technology and explains efforts to develop the basic mechanism into a commercial product.

Keywords: secure computation, cloud security, personal data

1. Introduction

Innovation in information and communication technology is making it possible to collect and analyze large amounts of diverse data, and there are increasingly high expectations for value creation by using big data. In particular, various efforts have been made recently to enable secondary use of personal data for the development of society and industry. The term *personal data* is not limited to the data defined by the Personal Information Protection Law, but is used in the broader sense of all information that concerns individuals [1]. Businesses that handle personal data are responsible for giving sufficient consideration to the privacy of the individuals who provide that information and for taking appropriate security management measures. Article 20 of the Personal Information Protection Law, Measures for Security Management, states that "Businesses that handle personal data must prevent the leaking, loss, or damage of that personal data and devise other appropriate measures for managing the security of personal data." In general, obtaining the *consent of the person* in respect to the purposes for using personal data is different from taking measures required by the Measures for Security Management. Leakage of personal data after the business obtains consent is an issue of responsibility, and such leakages reduce the public's trust.

The problem, then, is what security management measures are necessary and appropriate for the secondary use of personal data. Consider, for example, the situation in which individuals and businesses provide personal data to cloud providers, the cloud providers process the personal data for provision to secondary users, and the secondary users use the results of that processing (**Fig. 1**). In this case, the processing results consist of analysis results or anonymized personal data. The cloud provider is required to take security management measures for the obtained personal data and also for the processing results provided to the secondary user. In particular, the suppliers of the personal data cannot directly manage or control the personal data supplied to the cloud provider, and therefore, they probably worry about the possible leakage or unintended use of that data. This concern about security is cited as the most important factor in decisions to refrain from using cloud services. The concept known as *secure computation* has been attracting attention as a technological solution for security management that solves that problem by preventing the leakage or misuse of personal data and by maintaining privacy. In the remainder of this article, we present an overview of technology related to secure computation and describe the work being done at the NTT Secure Platform Laboratories.

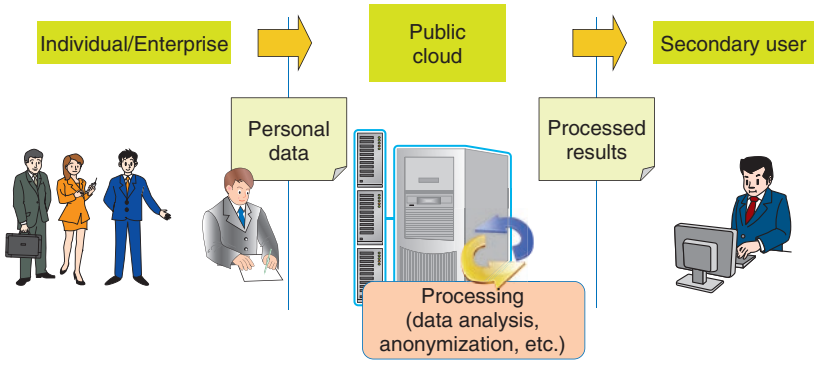


Fig. 1. Example process of secondary use of personal data.

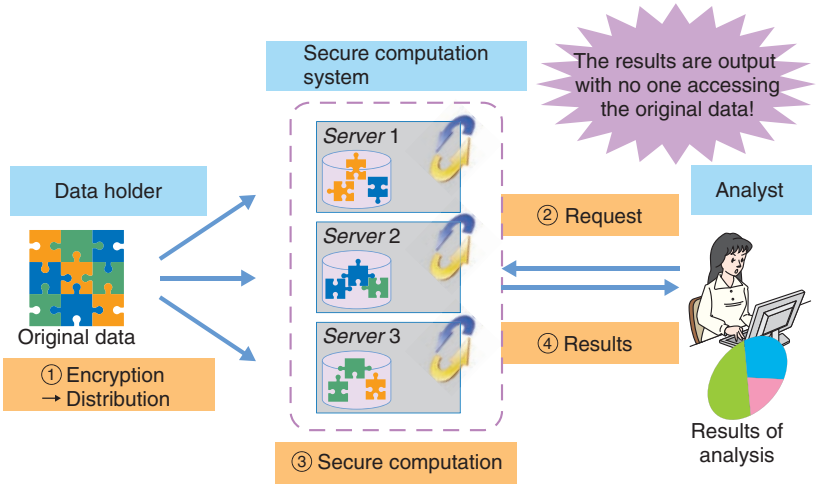


Fig. 2. Basic system model of secure computation.

2. Secure computation

Secure computation technology allows statistical processing and other kinds of data processing to be performed on data that remains in a secure form. One security measure used when information is entrusted to the cloud is data encryption. When the data processing is also done in the cloud, the processing is not usually performed with the data in the encrypted state, so the data must be decrypted in the cloud, which creates a risk of data leakage.

Secure computation is a technology that reduces that risk. A secure computation system is illustrated schematically in **Fig. 2**. First, the possessor of the data to be used in statistical processing or another form of processing sends the anonymized data to a secure computation system. Next, the analyzing party that wants to do the processing sends a request for analysis to the secure computation system. The secure computation system then performs the processing on the data that is still in the anonymized form and sends only the results to the party that requested the analysis. The special feature of secure computation is that the data is never decrypted, and the requesting party never receives anything other than the analysis results.

The secure computation technology that is currently under development is implemented by the interworking of multiple computers (secure computation servers) that communicate with each other. The input data is first anonymized with a secret-sharing algorithm and then distributed among the secure computation servers. The secure computation servers perform the processing while exchanging the

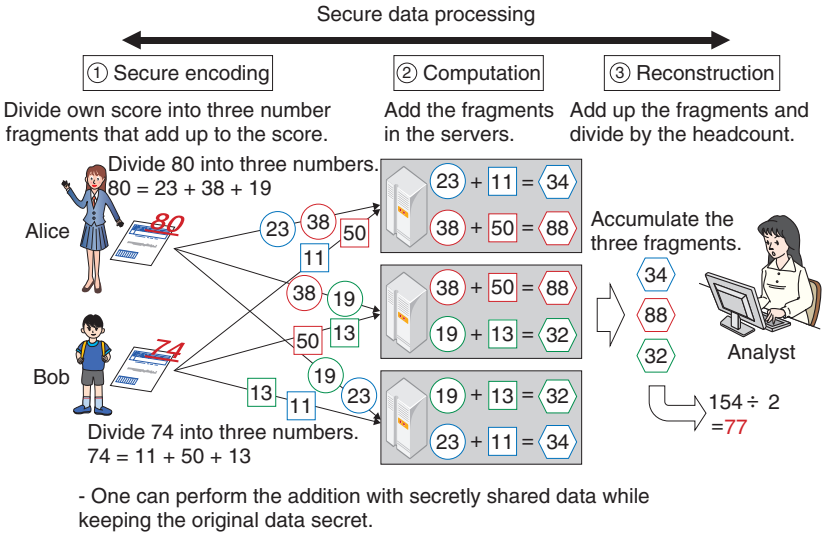


Fig. 3. Simple example of secure computation processing.

anonymized data among themselves. The simplest example of secure computation processing is illustrated in **Fig. 3**.

In Fig. 3, person A (Alice) and person B (Bob) have test scores. We want to obtain the average score for these two people without anyone learning what the individual test scores are. To obtain that result, the two test scores are first processed by implementing secret distribution. Specifically, each score is split into three numbers whose sums are the original scores. Those three distributed numbers are sent to secure computation servers, which then add up the received data. Finally, the resulting numbers are summed, and the sum is divided by the number of persons (two in this case) to obtain the ultimate result, which is the average test score for the two people (77 in this example). The numerical values used within the secure computation servers are unrelated to the original test scores, so the privacy of person A and person B is preserved.

The method depicted in Fig. 3 can only be used to obtain an average value. In 1982, however, A. C. Yao proposed a method that enables secure computation for any type of processing [2], although it was impractical for processing very large amounts of data. Research to improve efficiency has continued since that time, however. Research on secure computation began over ten years ago at the NTT Secure Platform Laboratories, and we achieved the world's fastest processing in 2005 [3]. Good compatibility between secure computation and statistical processing has also been discovered in recent years, and experiments that verify practical performance have been conducted [4].

3. Overview of the Trust-SC secure computation platform

Trust-SC is the first on-line statistical analysis system that was developed applying secure computation technology for commercial use. NTT Secure Platform Laboratories keeps striving to be a world leader in research on secure computation technology, and this on-line service makes it possible to analyze data and obtain results of statistical analysis without reconstructing any master data.

An overview of the Trust-SC platform is presented in **Fig. 4**. The main features are explained in the following subsections.

3.1 Fast processing and variety of statistical functions with combination of *R* functions

The statistical functions provided by the Trust-SC secure computation platform are listed in **Table 1**. The first group of functions includes basic statistical functions (maximum, minimum, median, mean, and distribution), the second group consists of applied statistical functions such as table operations and a t-test function, and the functions in the third group are related to database processing operations (search and shuffle). Most all-purpose statistical functions can be covered by combining Trust-SC and "R" statistical

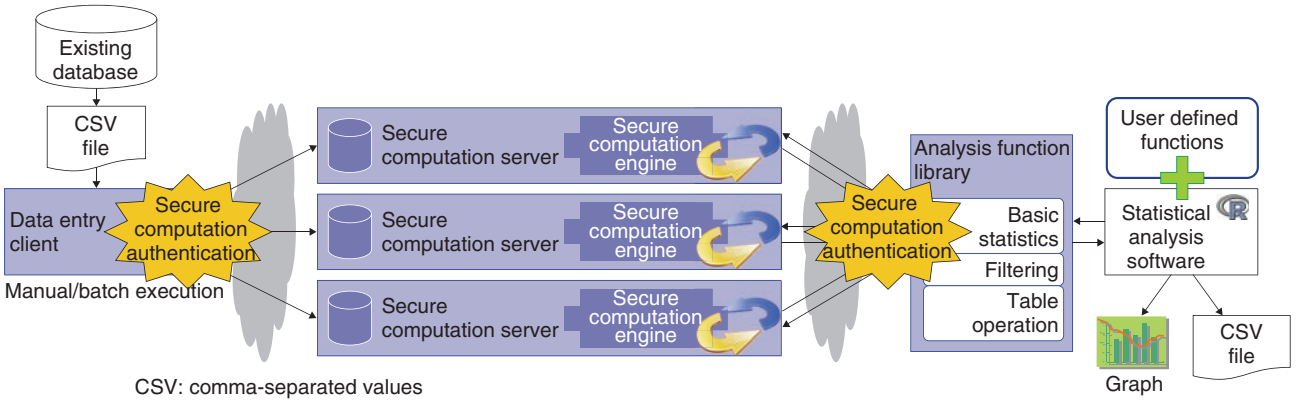


Fig. 4. Trust-SC secure computation platform.

Table 1. Statistical functions provided by Trust-SC.

Type	Function
Basic statistical functions	Maximum Minimum Median Mean Distribution
Applied statistical functions	Table operations t-test Kaplan-Meier estimator
Additional functions	Item count Item filter Shuffle

functions. Furthermore, the performance of the Trust-SC functions has the possibility of achieving the world's fastest speed class while still keeping data in an anonymized form.

This performance was largely possible thanks to NTT's research on a secure computation algorithm. The Trust-SC platform does not require any special personal computer (PC) power; it can be installed and perform those functions at high speed on ordinary PC servers (e.g., a four-core CPU (central processor unit) and 32 GB of RAM (random access memory)).

3.2 Statistical analysis with R

The Trust SC is operable with the R language, a major statistical computing and graphics OSS (open source software) tool that is used extensively in various fields, including medicine, finance, and the environment (**Fig. 5**). The statistical analysis functions supplied by the Trust-SC platform are all implemented as R function libraries. By using R, the basic statistical functions listed in Table 1 can be combined to define other required R functions. Also, the many R library functions can be used to output results in various formats such as graphs and files.

3.3 Secure computation authentication

This system prevents leaking of the original data even if a single secure computation server is breached. A security policy of this kind, however, requires that the user perform authentication with a different password for each secure computation server. The secure computation authentication process uses a proprietary method that achieves both usability and security. The user uses a single password for authentication, but the passwords stored in the servers are managed in secret-sharing encryption form, and password verification is performed with secure computation. Even in the event that one secure computation server is invaded, only a single item of secret-sharing data can be leaked, and it is not possible for someone to obtain the data needed to pose as the user.

4. Future development

Research on secure computation was previously only theoretical, but practical research has suddenly accelerated with the demand for security management in the secondary use of personal data, and practical goals have been established. The NTT Secure Platform Laboratories has developed the world's first commercial-level secure computation system and confirmed the ability of the system to carry out data analysis processing on a data scale of 100,000 items in a practical amount of time.

Our future tasks include developing the system so

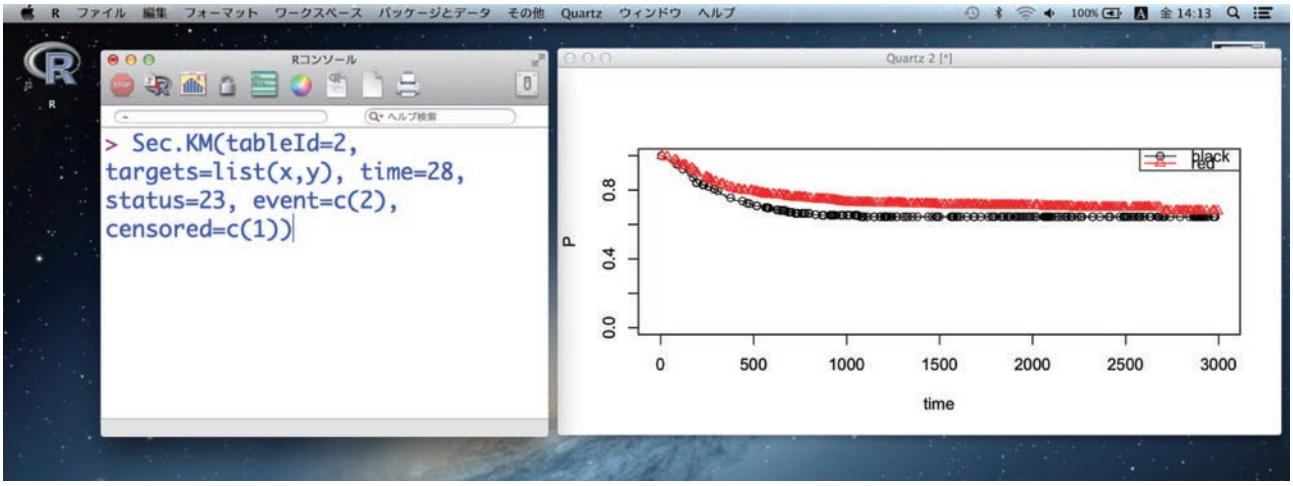


Fig. 5. Results of executing a Kaplan-Meier estimator with R.

that it is capable of handling data on a larger scale for use in big data applications, and adopting a social scientific approach to research that supports relevant laws in order to give users a greater sense of security.

References

[1] IT Strategic Headquarters, Personal Data Study Group (in Japanese). http://www.kantei.go.jp/jp/singi/it2/pd/index.html

[2] A. C. Yao, "Protocols for Secure Computations (Extended Abstract)," Proc. of 23rd Annual Symposium on Foundations of Computer Science (FOCS 1982), pp. 160–164.

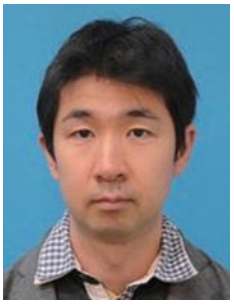
[3] NTT press release, "World's Fastest Secure Circuit Evaluation Algorithm Enabling Arbitrary Operation on Encrypted Data," published on October 25, 2005. http://www.ntt.co.jp/news/news05e/0510/051025.html

[4] NTT press release, "World's First Verification of Secure computation Technology Applied to Medical Statistical Processing," published on February 14, 2012 (in Japanese). http://www.ntt.co.jp/news2012/1202/120214a.html

Koji Chida

Senior Research Engineer, Information Security Project, NTT Secure Platform Laboratories.

He received the B.S. and M.S. from Waseda University, Tokyo, in 1998 and 2000, respectively. Since 2000, he has been engaged in research on cryptography and privacy enhancing technologies at NTT. He received the Dr. Eng. degree from Waseda University in 2006. He is a member of the Information Processing Society of Japan (IPSJ). He was awarded the IPSJ Best Paper Award in 2012.



Dai Ikarashi

Researcher, Information Security Project, NTT Secure Platform Laboratories.

He received the B.S. and M.S. from The University of Tokyo in 2005 and 2008, respectively. Since 2008, he has been engaged in research on cryptography and information security at NTT. He is a member of the IPSJ.

Teruko Miyata

Senior Research Engineer, Security Management & Operations Project, NTT Secure Platform Laboratories.

She received the B.S. and M.S. in mathematical science from Ochanomizu University, Tokyo, in 1991 and 1993, respectively. She joined NTT laboratories in 1993. Her current fields of interests are protection technology of privacy and identity management.

Hiroyoshi Takiguchi

Research Engineer, Security Management & Operations Project, NTT Secure Platform Laboratories.

He received the B.S. and M.S. in human-environment studies from Kyushu University, Fukuoka, in 2000 and 2002, respectively. He joined NTT laboratories in 2003. His current fields of interests are protection technology of privacy and identity management.

Naoto Kiribuchi

Security Management & Operation Project, NTT Secure Platform Laboratories.

He received the B.E. and M.E. in informatics from the University of Electro-Communications, Tokyo, in 2010 and 2012, respectively. He joined NTT laboratories in 2012. His current fields of interests are cryptography and information security.