Trial Service of Secure Computation System San-shi™

Hiroyuki Kitajo, Takuya Yamaguchi, Sanami Nishiyama, Gen Takahashi, Asami Miyajima, Keiichi Hirota, Shoko Nishida, and Junko Hashimoto

Abstract

To enable the safe and secure use of corporate secrets, personal data, and other types of data that must be kept confidential, NTT has developed Secure Computation System San-shi™ that can perform tabulation and statistical processing securely and with practical performance without decrypting the data. As an initiative to stimulate the use of data, NTT is providing San-shi as a free trial service for a limited period so that many users can experience this secure computation technology. The advantage is that it enables integrated analysis without mutual disclosure of data among organizations while keeping the data encrypted. This article describes this initiative and introduces secure computation technology.

Keywords: secure computation technology, San-shi, cybersecurity

1. Background

Digital transformation is currently underway in a variety of fields, and this transformation is driving change toward a service economy, open systems, social networking, and smart systems. At the same time, the accumulation of cross-sector data and the skillful use of that data are expected to foster innovation and promote development and economic growth in a wide range of fields. However, the risk of incidents and the high social responsibility associated with data management and the need for data security measures to protect corporate strategy are factors that have hindered the expanded use of data.

To help eliminate these obstacles to data usage, NTT has taken a global lead in the research and development of secure computation technology that enables data processing while keeping the data encrypted. The advantage of secure computation technology is that data operations are invisible to everyone, except for the results of computation (**Fig. 1**), thereby enabling a new form of integrated analysis using data that up to now has been difficult for organizations to mutually disclose. Application examples of this technology have already been tested in several fields including multi-facility clinical research data analysis [1] and genome data analysis [2]. NTT has expanded the operations and functions of this technology, as well as enhanced the performance and made other improvements in developing Secure Computation System San-shi™ (referred to below as San-shi) [3].

2. Overview of San-shi trial service

NTT's San-shi has been developed as a system having the advantage of secure computation technology that enables integrated analysis of data without mutual disclosure of data among organizations providing and/or using such data, while keeping the data encrypted. NTT has begun a free trial service of San-shi to enable many users in a variety of fields to experience this value. The trial period began on August 20, 2018, and runs through March 2019. At present, trial users can experiment with San-shi in various fields including healthcare, manufacturing, and system integration.

Using San-shi implemented on the cloud, users can

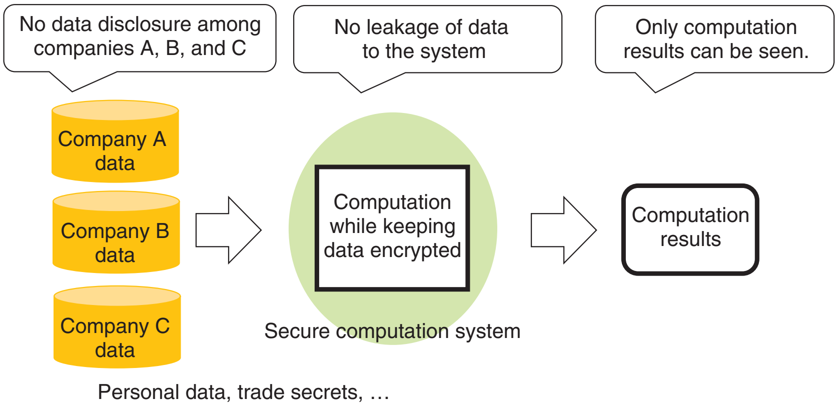

No data disclosure among companies A, B, and C
No leakage of data to the system
Only computation results can be seen.
Company A data
Company B data
Company C data
Computation while keeping data encrypted
Computation results
Secure computation system
Personal data, trade secrets, …

Fig. 1. Advantages of secure computation technology.

Table 1. Typical analysis scenarios provided in the trial service for customer use.

Feature	Scenario
Strengthen ties with other companies in the same industry.	Integrate and analyze the sales data of multiple companies in a region to enhance product assortment and avoid the loss of sales opportunities, thereby invigorating an entire regional trading zone.
Link data between different industries.	Combine and analyze purchase data of online sales companies and vital-signs data (BMI, pedometer data, etc.) held by companies providing health-support apps, and apply results to the marketing of health-related products and offering of product recommendations (expand advertising revenues).
Experience San-shi's operations and functions.	Discover what operations are possible against various types of data groups such as household attributes, household expenses, and food expenses.

BMI: body mass index

experience actual functions for tabulating and statistically processing data while keeping the data encrypted. Three types of scenarios and trial data that highlight the use of San-shi have been prepared to make it particularly easy for users to try out the system (**Table 1**).

The first scenario—strengthen ties with other companies in the same industry—increases the quantity of data (row data). The idea here is to invigorate an industry or solve problems in that industry without having to disclose information to a competitor. With a regional trading zone as an example, San-shi enables the user to securely register the sales data of multiple companies and store the data in a state that joins all of the data together. In this way, the user (data analyst) can check total sales figures for an entire trading zone, what merchandise had strong sales in different sales periods, and other details.

The second scenario—link data between different industries—increases the number of data items (column data). The ability to combine data belonging to different industries is expected to reveal new trends and generate new business value. To give an example, San-shi could be used to combine and securely register the purchase data of online sales companies and the vital-signs data of companies providing health-support apps, which would enable the user (data analyst) to determine the average number of steps taken by customers purchasing health food products by age group.

Finally, the third scenario—experience San-shi's operations and functions—gives users a free hand in trying out the operations and functions supported by San-shi. With San-shi, the user (data analyst) can securely register publically available statistical information (general-purpose micro data) and perform operations on that data associated with consumer expenditures, food expenses, insurance and medical expenses, and other details.

Additionally, for users who wish to experience San-shi beyond these typical scenarios, NTT supports the trial use of individual analysis scenarios based on data that users themselves possess.

3. Secure computation technology

Secure computation is the capability to perform computations on data while keeping the data encrypted. With most cyphers, data must first be decrypted to perform calculations with it, but this runs the risk of data leaks to data analysts, system operators, or elsewhere. Secure computation technology, however, enables computations to be performed while keeping the data encrypted, which prevents data analysts or system operators from seeing any data, including the results of computations in progress. This means that even confidential corporate information and trade secrets can be safely used for data computation.

The framework of secure computation technology was first established in the 1980s based on the theory of secure multi-party computation in the fields of computer science and cryptographic theory. However, the time required for computation was excessive (as the process was slow), presenting an obstacle to practical use. More recently, though, much research has been done to find ways to increase the computation speed and achieve practical implementations of secure multi-party computation. At NTT, we have developed a high-speed secure computation system based on secret sharing.

3.1 Encryption by secret sharing

NTT adopts secret sharing as the mechanism for encryption in its secure computation technology. Secret sharing is a scheme that enhances confidentiality by dispersing data into fragments called *shares*. In this scheme, information cannot be leaked from individual shares, and data can be recovered even if some shares are lost. This secret sharing scheme makes use of ISO/IEC 19592-2, a standard of the International Organization for Standardization (ISO). NTT members edited this standard and contributed to its formulation.

3.2 Multi-party computation based on secret sharing

NTT adopts multi-party computation based on secret sharing as the mechanism for computing while keeping the data encrypted. A multi-party computation system consists of multiple servers and a set procedure for exchanging data between those servers and performing operations on the data. Each server registers shares dispersed under the secret sharing scheme—data are always handled in this state of dispersed shares.

3.3 Safety of secure computation technology

There is no way that original data or computation results can be restored from individual shares on a server. However, given that shares are dispersed to and registered on multiple servers, it would be possible to restore the data if shares were to be obtained in an unauthorized manner from a certain number of servers. For this reason, appropriate management of each server is a precondition for safety.

3.4 Principle of secure computation technology

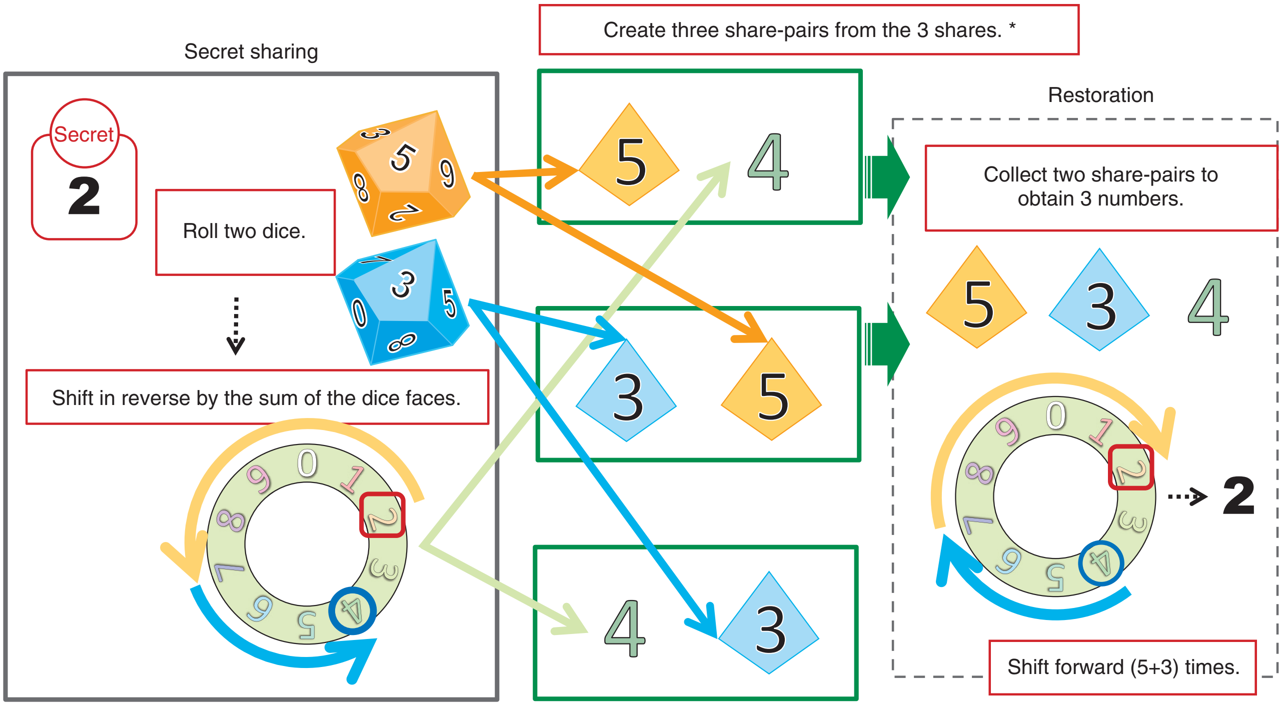
In secure computation technology, data are dispersed into multiple shares. Here, we introduce an example of dispersing “2” into three shares (**Fig. 2**). Generating shares in secret sharing is achieved by generating random numbers and performing computations based on those numbers. In this example, the share-generation process begins by generating two random numbers, each of which can take a value from 0 to 9. In the case where 5 and 3 are generated as random numbers, two of the three generated shares are taken to be 5 and 3. The process now computes the third share from these two shares by subtracting the sum of 5 and 3 (= 8) from original data 2 to obtain –6, which corresponds to 4 on the roulette wheel shown in the figure. The third share is therefore determined to be 4.

To restore the original data, the process collects the three shares 5, 3, and 4 and adds them up to get the value 12, which corresponds to 2 on the roulette wheel. The value of the original data is therefore determined to be 2.

Here, the process computes the shares generated in this way directly from each server. For example, a sum total, if desired, can be computed by simply summing the shares in the state in which they exist on each server. Finally, the result of the summation can be obtained by restoring the result of summing the values calculated on each server using the method described above.

4. San-shi features

NTT’s San-shi is a world-class secure computation


Create three share-pairs from the 3 shares. *
Secret sharing
Secret
2
Roll two dice.
Shift in reverse by the sum of the dice faces.
5
4
3
5
4
3
Restoration
Collect two share-pairs to obtain 3 numbers.
5
3
4
2
Shift forward (5+3) times.

* Three share-pairs: Created by combining two instances each of the 3 generated shares consisting of two dice faces and the shift result

Fig. 2. Principle of secure computation technology.

Table 2. Main operations of NTT's Secure Computation System San-shi™.

Data operation	Tabulation	Basic statistics		Tests
Table join	Frequency table (cross tabulation)	Total sum	Maximum	t-test
Filtering by conditions	Quantity table	Mean	Minimum	
		Variance	Median	Other
		Sum of products	Quantiles	Kaplan-Meier method

system that dramatically improves processing speed—a technical problem for many years in secure computation—while being capable of tabulating and statistically analyzing data on a scale of 100 attributes × 10,000,000 items within a realistic length of time. San-shi features an extensive set of tabulation functions and basic statistical operations, each of which can be executed at high speed.

4.1 Extensive operation variation

San-shi enables the user to execute the operations listed in **Table 2** on a graphical user interface (GUI) on a web browser or via an interface to "R" statistical analysis software without viewing the original data. The user can also create simple programs on "R" to perform regression analysis, principal component analysis, and other types of analyses according to the target application. The San-shi trial service provides the user with partial access to these interfaces.

In particular, San-shi's table-join function (a function that enables the data of multiple tables to be joined without leaking the join key) makes it possible to integrate the data of different companies and industries and obtain only the results of cross analysis without having to mutually disclose individually held data. This capability enables supply chains or customer

Table 3. Execution times of typical functions.

Function	Execution time (milliseconds)				
No. of data items	10^3	10^4	10^5	10^6	10^7
Addition	1	1	1	2	14
Multiplication	1	1	5	39	473
Sort	10	23	133	1274	12,255
Sum total	1	1	1	1	9
Sum of products	1	1	1	2	15
Quantity table creation	22	46	255	2252	22,676
Shuffle	1	1	8	60	731
Table join	19	65	518	4965	53,205
Data filter with prefix match	6	6	14	91	813
Data filter with numerical match	5	5	10	35	413

Measured with three personal computers (central processing unit: Intel Core i7 6900K, memory: 32 GB, solid state drive: 525 GB, operating system: CentOS 7.2) connected on a 10-Gbit/s network

data that overlap multiple companies to be analyzed, which can contribute to the creation of new value in the use of data that could not be achieved up to now within a single company or industry.

4.2 High-speed processing sufficient for practical use

In addition to adopting a secret sharing scheme [4], San-shi is able to provide both extensive operation variations as described above and faster processing through a proprietary speed-boosting algorithm and fast implementation method.

Secure computation technology based on secret sharing has two key advantages: the size of data basic to data processing is small, and the frequently used operations of addition and multiplication can both be executed at high speed. This means that San-shi can process a variety of operations at high speed compared with secure computation based on other types of encryption schemes such as homomorphic encryption.

In addition to the above, NTT has developed a basic algorithm for secure computation having extremely low computational and communication costs and has applied this algorithm using a fast implementation method. These measures have dramatically improved the processing speed, enabling NTT to achieve the world's highest speeds in executing operations of this type.

Execution times of typical functions are listed in **Table 3**. Sort processing of 10 million records can be performed in 12.2 seconds. This time can be compared with an execution time of about 1 second when sorting 10 million unencrypted records by a standard sorting algorithm. The difference in performance between secure computation technology and ordinary computer processing is therefore about one order of magnitude.

4.3 San-shi system

Secure computation technology enables multi-party computation over multiple servers operating in an integrated manner. The San-shi system consists of secure computation clients and three or four secure computation servers. The secure computation client that performs data registration divides data into shares under the secret sharing scheme and registers those shares on different servers. In addition, the secure computation client that performs data analysis issues requests to each server for computation (data analysis) and obtains only the results of computation. Here, data are registered in table format such as a relational database. The computation of mean or variance values, for example, can be requested by specifying the name of the table or column where the data are stored. Each server receiving such a computation request cooperates with the other servers to perform multi-party computation and returns computation results as shares to the secure computation client that performs the data analysis. This client then restores those shares to obtain the result.

5. Future development

Going forward, NTT aims to further promote the safe and secure use of confidential corporate and personal data through the San-shi trial service while endeavoring to develop and globally propagate data usage technology including secure computation technology.

References

[1] Press release issued by NTT on Feb. 14, 2012 (in Japanese). http://www.ntt.co.jp/news2012/1202/120214a.html

[2] Press release issued by NTT on July 12, 2016 (in Japanese). http://www.ntt.co.jp/news2016/1607/160712a.html

[3] Website of NTT on secure computation, http://www.ntt.co.jp/sc/project_e/data-security/secure_computation.html

[4] Press release issued by NTT on Oct. 23, 2017 (in Japanese). http://www.ntt.co.jp/news2017/1710/171023a.html

Hiroyuki Kitajo

Manager, R&D Produce Group, Research and Development Planning Department, NTT.

He received a Bachelor of Information Engineering from Tohoku University, Miyagi, in 2000. He joined NTT EAST in 2000. He has been in his current department since 2016, where he has been promoting information and communication technology (ICT) business and technologies for the medical and healthcare field.

Takuya Yamaguchi

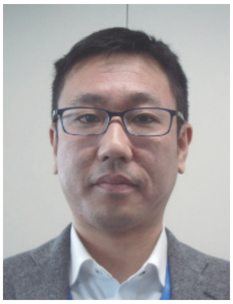
Manager, Produce Section (Security), Research and Development Planning Department, NTT.

He received a B.E. in physics from Sophia University, Tokyo, in 2000. He joined NTT EAST in 2000 and worked in corporate sales from 2000 to 2005. He was involved in developing security services at NTT EAST from 2006 to 2014. He has been in his current department since 2015, where he has been promoting security related business and technologies.

Sanami Nishiyama

Associate Manager, R&D Produce Group, Research and Development Planning Department, NTT.

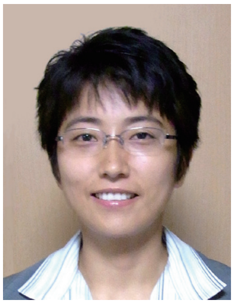
She received a Bachelor of Management Engineering from Nagoya Institute of Technology, Aichi, in 2000. She joined NTT WEST in 2000. She has been in her current department since 2018, where she has been promoting ICT business and technologies for the medical and healthcare field.



Gen Takahashi

Senior Research Engineer, NTT Secure Platform Laboratories.

He received a Master of Media and Governance from Keio University, Tokyo, in 2005. He joined NTT in 2006. His research interests include information security and cryptographic engineering. He received the SCIS Paper Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2008.



Asami Miyajima

Senior Research Engineer, NTT Secure Platform Laboratories.

She received a Master of Science and Technology from Keio University, Tokyo, in 2000. She joined NTT in 2000. Her research interests include information security.

Keiichi Hirota

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

He received a B.S. and M.S. from Mie University in 1995 and 1997, and a Ph.D. in informatics from the Graduate University of Advanced Study (SOKENDAI) in 2008. He joined NTT in 1997. His current research interests are security and privacy in information processing, information sharing, and data utilization. He is a member of the Information Processing Society of Japan.

Shoko Nishida

Research Engineer, NTT Secure Platform Laboratories.

She received an M.S. from Kyushu University, Fukuoka, in 2009. She joined NTT in 2009. Her research interests include information security.

Junko Hashimoto

Research Engineer, NTT Secure Platform Laboratories.

She received an M.S. from Kyushu University, Fukuoka, in 1999. She joined NTT in 1999. Her research interests include information security.