August 26, 2021

Toshiba Corporation

Tohoku University Tohoku Medical Megabank Organization

Tohoku University Hospital

National Institute of Information and Communications Technology

Toshiba, ToMMo, Tohoku University Hospital, and NICT Demonstrate the Use of Quantum Cryptography Communication and Secret Sharing Technologies for Distributed Storage of Genome Analysis Data

-Contributing to safe data management in the fields of genomic research and medicine-

Toshiba Corporation, Tohoku University Tohoku Medical Megabank Organization (ToMMo), Tohoku University Hospital, and the National Institute of Information and Communications Technology (NICT) have developed a distributed storage technology that combines quantum cryptography communication and secret sharing technologies, successfully demonstrating the world's first experimental large-scale[1] genome analysis data backup to multiple sites[2]. These technologies will allow data backup that prevents data leaks and tampering over long periods and are expected to contribute to safe data management in the fields of genomic research and medicine.

A part of this work was performed for Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), "Photonics and Quantum Technology for Society 5.0" (Funding agency: QST). Toshiba, ToMMo, Tohoku University Hospital, and NICT will present the details of these technologies and their experimental demonstration at the 11th International Conference on Quantum Cryptography (QCrypt 2021).

Secure protection of personal and other confidential information is becoming increasingly important, and cybersecurity measures are being strengthened in the medical field as well. Since genome analysis data is legally treated as personal information that identifies individuals, storing and transmitting such data requires taking strict measures for secure protection against data leaks. Also, utilization in genomic medicine, which is expected to become increasingly widespread, will also require distributed backups of genome analysis data to prevent data loss or damage due to system failures or natural disasters. Until now, consideration of the risk of data leaks has necessitated backup of large-capacity and highly confidential genome analysis data to storage media such as disks or tapes, in some cases at remote locations, rather than backing up

over a network. However, the safe transport and storage of such media are expensive and time-consuming, so establishing backup methods for large-capacity and highly confidential data, reducing backup costs, and ensuring convenience and availability of backups have been outstanding issues.

The group developed a distributed storage technology combining 1) a quantum cryptography communications technology that realizes secure cryptographic communication against any eavesdropping and unauthorized decryption based on the principles of quantum mechanics with 2) a secret sharing technology that realizes secure data storage by converting original data into multiple data fragments (shares) and demonstrated its ability to back up large-capacity genome analysis data. The developed distributed storage technology ensures information-theoretic security[3] for both communication and storage. Quantum cryptography communication technologies realize secure communication against all forms of eavesdropping and unauthorized decryption, while secret sharing technology ensures confidentiality of the original data, even if the stored data is partially damaged or leaked due to a system failure or natural disaster. In addition, the original data can be restored from the remaining data stored undamaged.

When storing genome analysis data, 1) an algorithm called the “XOR based secret sharing method” that enables high-speed secret sharing divides the original genome analysis data into multiple shares (data that has been made meaningless), and 2) one-time-pad encryption communication[4] using encryption keys generated and stored by quantum key distribution transfers each share and stores them in a different location. When restoring genome analysis data, same one-time-pad encrypted communications are applied to bring shares at various sites to a single site, where the “XOR based secret sharing method” algorithm restores the original genome analysis data from multiple shares.

The high-speed secret sharing and one-time-pad encrypted communication are realized with the technologies, 1) a “direct access technology” that specifies storage address of each data of shares at the disk sector level for high speed reads and writes of shares at each site, and 2) a “parallel software technology” that performs high speed one-time-pad encryption using large numbers of encryption keys that are generated and stored by quantum key distribution, which guarantees information-theoretic security.

Using the developed distributed storage technology, the group conducted a demonstration experiment in which they distributed and stored genome analysis data at three sites in Sendai city, Japan: Toshiba Life Science Analysis Center (LSA), ToMMo, and Tohoku University

Hospital. First, three shares (A, B, and C) were calculated at ToMMo using the NICT-developed secret sharing technology for genome analysis data. After that, share A was stored at ToMMo, share B was transmitted by encrypted transmission using Toshiba's quantum cryptography technology and stored at Tohoku University Hospital, and share C too was transmitted using quantum cryptography technology and stored at LSA. When restoring the original genome analysis data, two of the three shares were brought to the ToMMo site for restoration using the secret sharing technology (Fig. 1).

Measuring the time and throughput required for distributed storage and restoration of one genome analysis data sample (about 80 GB), distributed storing process required about 30 minutes (356 Mbps) and restoration process required about 21 minutes (502 Mbps)[5]. Converting to minimum backup units at ToMMo (one hundred samples), this would take about 50 hours. This is comparable to practical speeds in the conventional use case of physically transporting genome analysis data using tape or other media from a remote storage location.

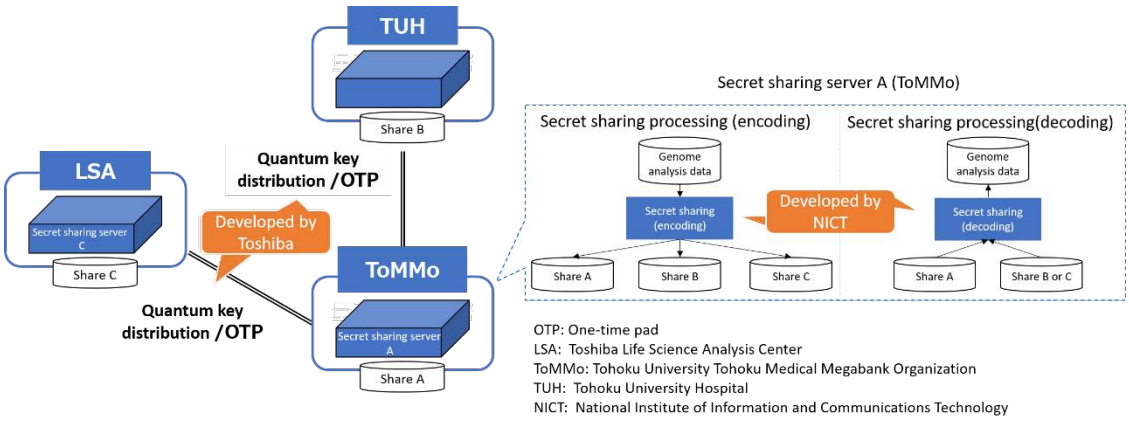

Fig. 1: Overview of the demonstration experiment for distributed storage of genome analysis data.

Toshiba will continue research and development of various quantum cryptography communication technologies, including system demonstrations that involve secret sharing technologies and promote the practical application of quantum cryptography communication technologies in various applications such as medicine, finance, governmental agencies, and communications infrastructure. ToMMo and Tohoku University Hospital will continue to promote the use of safe and secure ICT technology toward the realization of future forms of medical care based on genomic information. NICT will continue to research and develop technologies such as quantum cryptography and photon control with the goal of contributing to industry and to both basic and advanced research and development in the field of quantum communication.

Project structure

This demonstration was performed with the following system and division of tasks.

Toshiba: development of the quantum cryptographic communication system and high-speed one-time-pad cryptographic communication; overall system construction and operation

ToMMo and Tohoku University Hospital: providing verification sites and use cases; confirming applicability in the field of genomic medicine

NICT: development of the high-speed secret sharing system and data backup system

[1]Approximately 80 gigabytes.

[2]As of August 26, 2021, based on a Toshiba survey. This was the world's first successful demonstration of storing genome analysis data using data distribution storage technology that combines quantum cryptography communication technology and secret sharing technology.

[3]Unlike the computational security used in general cryptography, which relies on assumptions regarding the difficulty of solving a specific type of mathematical problem, information-theoretic security is a property that is guaranteed by the probabilistic independence of the confidential information from any information obtained by an attacker or eavesdropper. In theory, no eavesdropping attack will leak information.

[4]A method for encrypted communication in which a random number with the same length as the original data (unencrypted, "plaintext") to be transmitted is used as the encryption key, which is disposed of after each use. Different encryption keys are used for different plaintexts. The encrypted text ("ciphertext") is generated and transmitted as the exclusive OR of the plaintext data and the encryption key, and the plaintext is decrypted again on the receiving side as the result of exclusive OR of the ciphertext data and the encryption (decryption) key. This method satisfies information-theoretic security and is the strongest and most secure encryption method. It has been proved that even an attacker or eavesdropper with high computing power will be unable to decrypt from ciphertext to plaintext, regardless of the time available.

[5]The size of the shared data is about 80 GB, which is the same as the original genome analysis data.

<Media Contact>

Toshiba Corporation
Corporate Communications Division, Media Relations Office
Email: media.relations@toshiba.co.jp

Tohoku University (ToMMo and Tohoku University Hospital)
Group of Public Relations, Tohoku Medical Megabank Organization
Email: pr[at]megabank.tohoku.ac.jp
(Please replace [at] to @.)

Press Office
National Institute of Information and Communications Technology Public Relations Department, Press Office
E-mail: publicity@nict.go.jp