3-2 Information Theoretically Secure Distributed Storage with QKD and Password-Authenticated Secret Sharing

Mikio FUJIWARA, Atsushi WASEDA, Ryo NOJIMA, Shiho MORIAI, Wakaha OGATA, and Masahide SASAKI

Distributed storage plays an essential role in realizing robust and secure data storage in a network over long periods of time. Distributed storage systems consist of a data owner machine, multiple storage servers and channels to link them. In those systems, secret sharing (SS) scheme is widely adopted, in which secret data are split into multiple pieces and stored in each server. To reconstruct them, the data owner should gather plural pieces. Shamir's (k, n)-threshold scheme, in which the data are split into n pieces (shares) for storage and at least k pieces of them must be gathered for reconstruction, furnishes information theoretic security, that is, even if attackers could collect shares of less than the threshold k, they cannot get any information about the data, even with unlimited computing power. Behind this scenario, however, assumed is that data transmission and authentication must be perfectly secure, which is not trivial in practice. Here we propose a totally information theoretically secure distributed storage system based on a user-friendly single-password-authenticated SS scheme and secure transmission using quantum key distribution (QKD), and demonstrate it in the Tokyo metropolitan area (≤90km). Our system will also be useful for highly secure data relay with a QKD network, greatly relaxing the security assumptions on the key relay nodes as well as enhancing the ability of risk management.

1 Introduction

It is urgent to develop cryptographic technologies for secure confidential communications of certain types of information, such as genome data in the pharmaceutical and medical fields because their confidentiality needs to be kept for generations over hundreds of years. However, there are concerns that the current cryptographic systems, which rely on computationally complex of factorization of prime numbers or distributed logarithms, may not be able to securely protect confidentiality with the advent of quantum computers [1]. In addition, the current cryptographic techniques may not be able to secure safety in 30 years due to steadily increasing computational capabilities. On the other hand, lattice-based systems [2][3] are proposed as the most promising quantum-resistant cryptographic systems. However, their performance evaluations are scheduled to be conducted from 2020 to 2022 at the National Institute of Standards and Technology (NIST) [4]. As such, it will take several years before solutions to the safety issues we are currently facing become available. Furthermore, changing cryptographic systems will likely necessitate a change in the length of public keys. We therefore might have to abandon the current communication protocol. In other words, we might have to drastically modify the current communication devices if protocols for each layer of the OSI model are changed. Alternatively, risks of information leakage may be eliminated by attaching a new system equipped with a dedicated line to the current communications system. This method involves a combination of quantum key distribution, which enables sharing of theoretically safe random numbers between the two parties, and Vernam's one-time pad [5][6]. The method will completely eliminate the risks of information being eavesdropped. Experiments to distribute quantum keys using laid fiber began in 2000, and a high-speed quantum key distribution device driven by a GHz-order clock has been developed [6]-[8]. And, network operations of quantum key distribution have been conducted in several countries [9]-[11]. While quantum key distribution secures safety when data is transmitted, it does not help preserve data. In contrast, Shamir's secret sharing scheme in modern cryptography had been known to be a safe data preservation means in theory [12]. However, information theoretically secure transmission of the data called "share" that is indispensable for recovering secret data, was only "assumed." In other

words, the integration of quantum key distribution and secret sharing can be a very reasonable approach as they make up for each other's shortcomings.

NICT and Tokyo Institute of Technology jointly proposed a totally information theoretically secure distributed storage system based on a user-friendly single-password-authenticated secret sharing scheme and secure transmission using quantum key distribution, and demonstrated a distributed storage system with information theoretically secure data transmission, storage, and authentication in 2016 [13]. In this paper, we describe the protocol and the system.

2 Information theory-based, safe single-password secret sharing protocol

2.1 Shamir's (k, n) threshold scheme

In this section, we discuss Shamir's (k, n) threshold scheme, on which our scheme is based [12]. The (k, n) threshold scheme works as follows: first, n owner of secret data *S* (integer) creates n individual values called "shares" out of S; second, the data owner secretly transfers each of the values to each of 1 to n shared servers; then, the data owner erases the secret data S; the secret data S is recovered through a predefined computation on the k shares collected from k servers in collaboration—k is defined as the threshold. The computation is performed using the formula below.

Share: random k-1th order polynomial of which the constant term is the secret data

$$f(x) = a_{k-1}x^{k-1} + ... + a_1x + a_0 \quad (1)$$

where $a_{k-1},....a_1,a_0$ are random integers and a_0 is the secret data *S*.

Holders of the ith share receive $(i, f(i))$, where "i" is the share-holder identifier. In a reconstruction operation, $a_0 (= S)$ is calculated from the pairs of $(i, f(i))$ collected from k shared servers. The secret data S is recovered as follows. When the identifier of k shared servers in collaboration is defined as $\{i_1,...i_k\}$, the following equations represent shares existing at each share server:

$$\begin{aligned} f(i_1) &= a_{k-1}i_1^{k-1} + ... + a_1i_1 + a_0 \\ &\vdots \\ f(i_k) &= a_{k-1}i_k^{k-1} + ... + a_1i_k + a_0 \end{aligned} \quad (2)$$

If $(i_1, f(i_1)),\cdots, (i_k, f(i_k))$ are substituted with numerical values, k individual linear equations with k variables of $a_{k-1},....a_1,a_0$ are generated. Therefore, by solving the simultaneous equations, all the unknown variables can be obtained. Then, the secret data S can be reconstructed.

Lagrange Interpolation is applied for actual secret-data reconstruction.

Figure 1 shows an instance of the (3, 4) threshold scheme. Substitution of more than three $(i, f(i))$ pairs is sufficient for the reconstruction of secret data S.

2.2 Password secret sharing protocol

This protocol protects any secret data from information leakage. In addition, it allows addition and multiplication between shares. For instance, the share obtained from addition of data $D^{(1)}$ and $D^{(2)}$ is $f_{D^{(1)}}(a_i) + f_{D^{(2)}}(a_i)$, and similarly, the share obtained from $D^{(1)} \times D^{(2)}$ is $f_{D^{(1)}}(a_i) \times f_{D^{(2)}}(a_i)$. In the multiplying process, however, the degree of the polynomial $f_{D^{(1)}}(x) \times f_{D^{(2)}}(x)$ is $2k-2$. So, $2k-1$ of shares is necessary to reconstruct $D^{(1)} \times D^{(2)}$. We took advantage of these characteristics to implement the password-sharing protocol, which requires only a single password for safe authentication as defined in information theory. Our scheme can be roughly divided into three phases: the "registration phase," where the share of the secret data and the share of the password are transmitted; the "pre-computation phase," where shared computations are executed to secure data secrecy at the time of data reconstruction; and the "data-reconstruction phase." More details are described below using an example of the (3, 4) threshold.

(1) Registration phase

(1-1) Since each calculation in the finite field with prime order $q = 2^m - 1$ can deal with only blocks of length at most $m - 1$ bits, secret data *D*, which has gener-

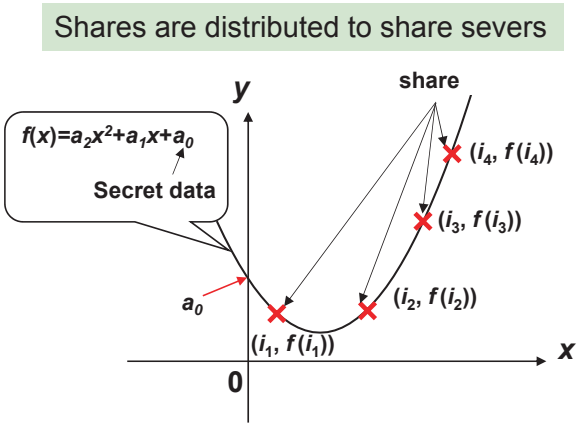

Fig. 1 Example of Shamir's (3, 4) threshold scheme

ally a much longer length, needs to be divided into pieces of $(m-1)$-bit block, say l pieces; $D = D_l|D_{l-1}|\cdots|D_1$. The data owner sets a $(m-1)$-bit password P, which should have sufficient entropy against an online dictionary attack, then computes a message authentication code, $MAC = D_l P^l + D_{l-1}P^{l-1} + \cdots + D_1 P$, which is denoted as D_{l+1}, and finally adds it to the data for later purpose of message authentication.

(1-2) For each data block, data shares $f_{D_i}(1)$, $f_{D_i}(2)$, $f_{D_i}(3)$, $f_{D_i}(4)$ are created for storage server 1, 2, 3, and 4, respectively, by using polynomial f_{D_i} of degree at most 2, where $i = 1, \ldots, l+1$. Password shares $f_P(1)$, $f_P(2)$, $f_P(3)$, $f_P(4)$ are created by using polynomial f_P of degree at most 1.

(1-3) They are then sent to the corresponding storage servers.

(1-4) Each server stores the set of shares.

(2) Pre-computation and communication phase

(2-1) Each server generates a random number, denoted as R_j for the j-th storage server, and makes its shares $f_{Rj}(1), f_{Rj}(2), f_{Rj}(3), f_{Rj}(4)$ by using polynomial f_{Rj} of degree at most 1. Furthermore, each server generates shares of the "0" $f_{0j}(1), f_{0j}(2), f_{0j}(3), f_{0j}(4)$ by using polynomial f_{0j} of degree at most 2, such that $f_{0j}(0) = 0$ should hold so as to keep confidentiality of the share in the data reconstruction phase without changing the value of the data share.

(2-2) The storage servers send these shares to each other.

(2-3) Each server receives three shares of three random numbers and three shares of the "0," and stores them together with the ones produced by itself.

For ITS, the above procedure has to be iterated l+1 times before each data reconstruction of l blocks of secret data. That is, j-th storage server has to keep $l+1$ sets of $(f_{R1}(j), f_{R2}(j), f_{R3}(j), f_{R4}(j), f_{01}(j), f_{02}(j), f_{03}(j), f_{04}(j))$.

(3) Data reconstruction phase Let P' be the password in the data owner's memory.

(3-1) The data owner chooses three storage servers among the four. We may assume that they are storage server 1, 2, and 3 without loss of generality, denote them as a set L={1, 2, 3}.

(3-2) The data owner generates shares of P', $f_{P'}(1)$, $f_{P'}(2)$, $f_{P'}(3)$ by using polynomial $f_{P'}$ of degree at most 1.

(3-3) Each set $(L,\ f_{P'}(j))$ is sent to each corresponding storage server (request).

(3-4) If $|L| \neq 3$, the request is rejected regarding it as an improper request. Otherwise, for each data block, each server, say j-th one, computes $R = f_{R1}(j) + f_{R2}(j) + f_{R3}(j)$, $Z = f_{01}(j) + f_{02}(j) + f_{03}(j)$ and

$$F_{ji} = \left(f_P(j) - f_{P'}(j)\right)R + Z + f_{D_i}(j) \tag{3}$$

The F_{ji} $(i = 1, \ldots, l+1)$ are then sent to the data owner (response). Here, note that R and Z should be discarded at each request-response for ITS.

(3-5) For each data block, the data owner finds polynomial $F_i(x)$ of degree 2 that satisfies $F_i(j) = F_{ji}$ for all j. $F_i(0)$ is the reconstructed block.

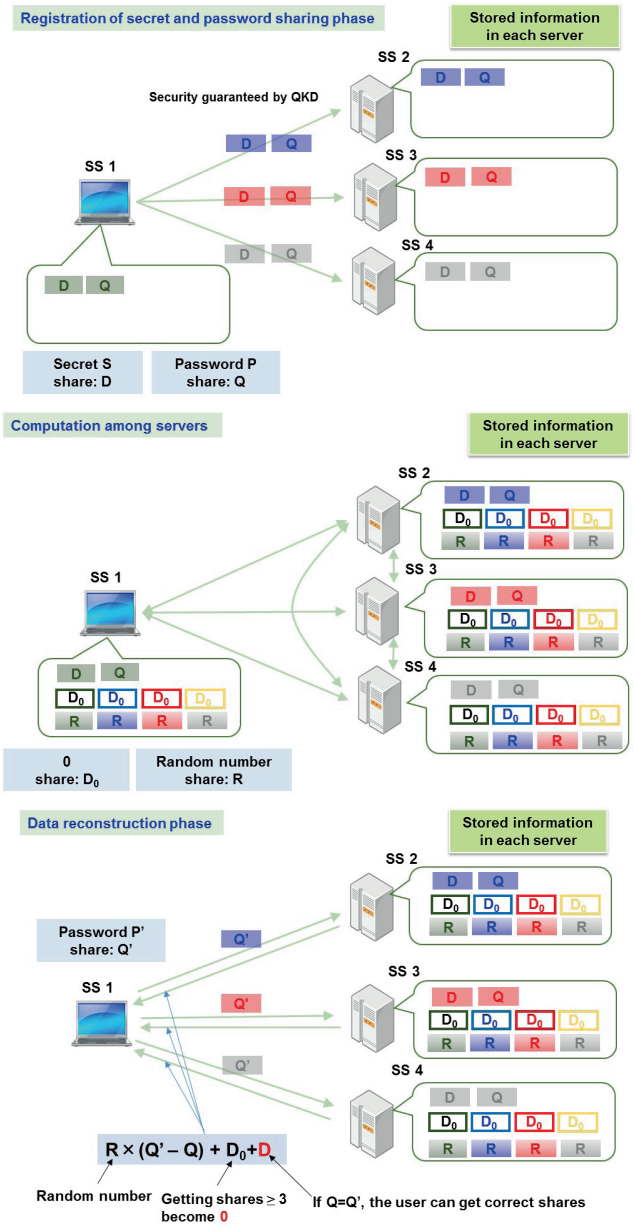

Fig. 2 Schematic diagram of password secret sharing protocol

(3-6) The data owner calculates *MAC* from $F_1(0), \ldots, F_l(0)$ as in the first phase. If $F_{l+1}(0)$ is equal to the calculated *MAC*, the data owner successfully reconstructs the secret data D.

In the procedure described above, note that there are no risks of data leakage when $P' \neq P$, because secret data is masked by f_{Rj} and f_{0j}. Therefore, the procedure ensures safe data transmission, data preservation, password authentication and data reconstruction in line with information theory. Figure 2 shows an outline of the protocol.

3 Implementation of and experiments on QKD network

To put the scheme we developed into practice, a communications network with networked QKD links is required. Since 2010, NICT and JGN have been cooperatively operating Tokyo QKD Network, whose operation center is based at NICT Headquarters (Koganei) [14]. Even with QKD systems made in Japan having world-highest performances, our QKD network is susceptible to transmission path losses because the system uses a series of single photons as transmission media. However, the key generation rate of these systems is 1 M bps with transmission of 50 km optical fiber[7] [8]. To extend the effective service area of QKD, we have been conducting operations on a connection of QKD links—a number of QKD links are interconnected at a connector. Such a connector is called a node, and it preserves the key information as a normal bit-stream. Such a node, because it is strictly protected to secure its safety so that no risks of intrusion and information theft assumedly exist, is defined as a "trusted node."

Service-area extension is realized by relaying the other link's key stored in a node after being exclusive-OR processed from a link to another link. The operation of such a QKD network requires a network architecture enabling strict key management. Since 2010, NICT has been developing a QKD network architecture and communication applications using a QKD key [11] [15] [16]. The network architecture we proposed has a structure of 3 layers following the OSI model; the layer called the quantum layer contains QKD links; in the key-management layer, keys generated in individual QKD systems are format-converted

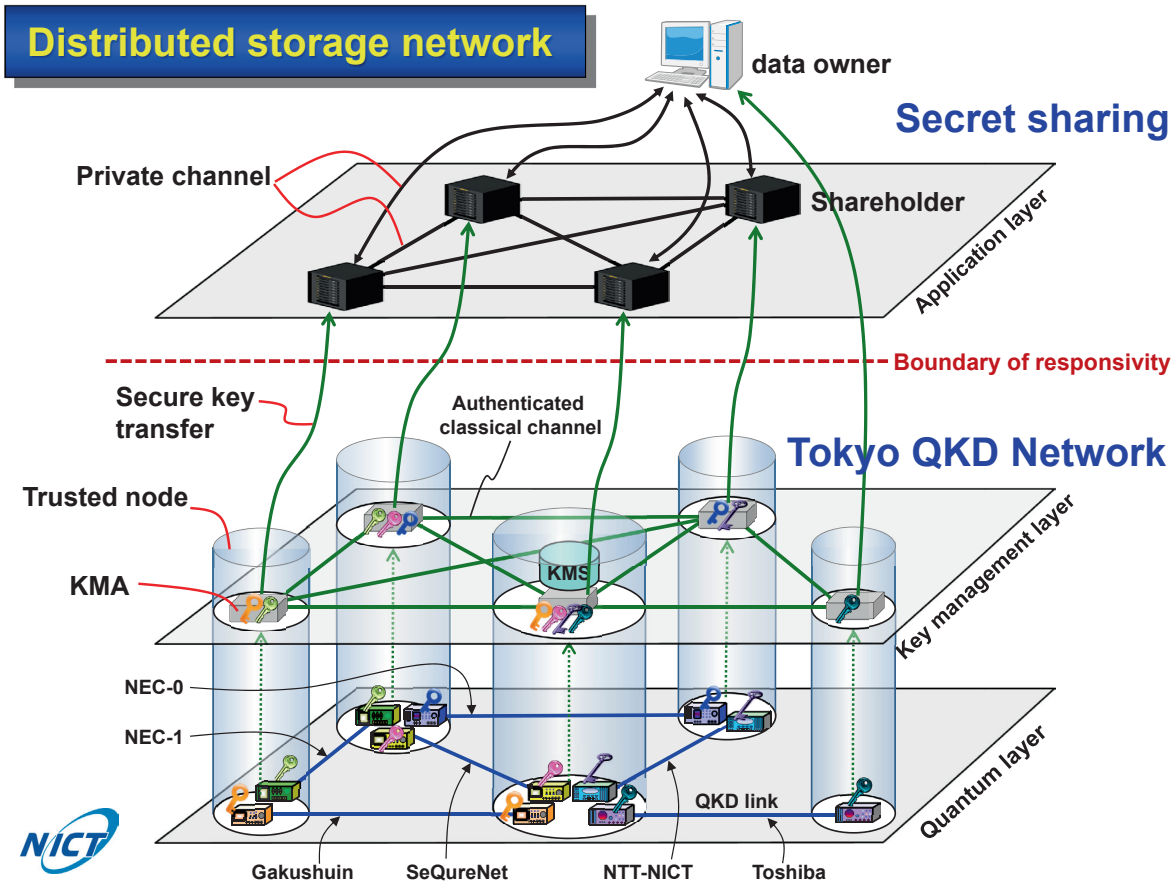

Fig. 3 Schematic diagram of QKD platform and shared storage network constructed on Tokyo QKD Network

Table 1 Protocol and communication distance / loss in Tokyo QKD Network

	Protocol	Transmission	
		Length (km)	Loss (dB)
NEC-0	BB84 with decoy	50 (Spooled fiber NICT premise)	10
NEC-1	BB84 with decoy	22 (field installed 95% areal line)	13
Toshiba	BB84 with decoy	45 (field installed 50% areal line)	14.5
NTT-NICT	DPS-QKD	90 (field installed 50% areal line)	28.6
Gakushuin	CV-QKD	2 (NICT premise)	2
SeQureNet	CV-QKD	2 (NICT premise)	2

to the predefined format that enables relays between QKD links and the provision of safe keys to various kinds of applications; on top of the key-management layer, the application layer is placed, where various kinds of communication applications using safe keys are developed in line with information theory. We call the combination of the quantum layer and the key-management layer the QKD platform. Below, we describe in detail the configuration of the key-management layer. Each "trusted node" is equipped with a key management agent (KMA) which is in charge of collection and management of keys from QKD links. Each KMA, interconnected with other KMAs via an authenticated communication path, is in charge of key-relaying. At the time of key-relaying, other additional information such as a key-ID attached when the key is created is transmitted.

The QKD platform has a key management server (KMS), which is placed in a trusted node. KMS, in charge of the collection from KMAs of QKD-link information such as the error rate, key creation rate, or the stored key amount, determines the key-relay route using the stored key amount and key creation rate, issues a route change order on the occasion of a drop in the stored key amount or an incident, and furthermore, monitors the life-cycle of the keys stored in KMAs, and issues an order to KMA for removing a key that lives longer than a predefined time from birth.

A key supply agent (KSA) is a device placed in each node, which works as an interface device from the QKD platform to the application layer, supplies a key in the format the application specifies, makes a record of the key ID, the application type and the date of supply, and delivers such information to the KMS.

One of the most significant ideas on the QKD platform is that the boundary of responsibility is set between the QKD platform and the application layer. Therefore, only very limited information crosses the border—no means is

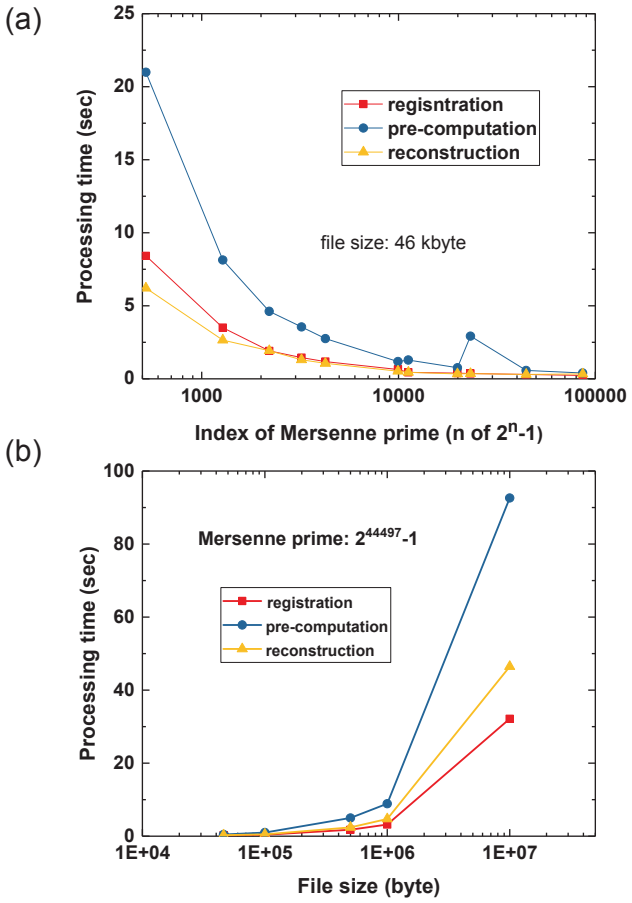

Fig. 4 Processing time in three phases (registration, pre-computation and reconstruction. (a) Processing time dependency on mersenne prime index size at data size of 46 kilobytes. (b) Processing time dependency on file size at the mersenne prime of 2^{44497}-1.

available in the application to access any data but the key supplied by the QKD platform, and at the same time, the QKD platform has no means to access any data belonging to the application layer such as what kinds of contents are being handled in an application. So, such strict authority separation in terms of access to devices belonging to the other side ensures the essential protection of network security.

The QKD links which form the Tokyo QKD Network are managed and operated by NEC, Toshiba, Gakushuin University, NTT-NICT and SeQureNet [7][8][17]–[19]. The link operated by NTT-NICT uses JGN's dark fiber between Koganei and Otemachi [14]. Leaving the details of those systems to other chapters, we summarize their QKD link protocols, transmission distances and transmission loss in Table 1.

We constructed a system capable of safe password authentication, data transmission, data preservation and data reconstruction in line with information theory, on the distributed storage. Figure 4 shows secret sharing processing time in the three phases (registration, pre-computation and reconstruction). Figure 4 (a) shows the dependency of processing time on the Mersenne prime index size when the data size is 46 kilobytes, and Figure 4 (b) shows the dependency of processing time on file size when a Mersenne prime of 2^{44497}-1 is used.

Note that the results shown in Fig.4 were obtained using conventional-type PCs; so, a drastic processing performance improvement would be expected if high-performance servers were used. However, the most significant factor limiting the processing speed of the current system is the key-synchronization process (i.e., key-sorting process which is executed on a server) that is essential to OTP encryption using the QKD platform supplied. So, we expect that software improvement will enable high-speed processing. On the other hand, we have proved by experience that our current system is capable of completing the process from registration to reconstruction of 10 M of data—often used for mail transmission—in approximately two minutes. This is the world's first successful demonstration of a system capable of safe password authentication, data transmission, data preservation and data reconstruction in line with information theory.

4 Summary

Thus, by combining QKD and a newly developed password-authenticated secret sharing scheme, we demonstrated, for the first time to our best knowledge, a distributed storage system with information theoretically secure data transmission, storage, and authentication in a metropolitan area network. This system uses a QKD network capable of generating information theoretically secure keys. We are currently planning to construct a system enabling safe data secrecy preservation for a very long time, by periodically updating shares stored in servers and adding long-term data leakage prevention capabilities to our system. Even if various new types of cryptographic schemes and networks systems become available, risks of eavesdropping and code-breaking will always exist if they are incapable of ensuring safety in line with information theory. The system we developed is not technically decodable and is therefore expected to be very robust against the types of attacks anticipated in the near future. NICT is capable of quickly responding to potential decryption threats by providing safe system solutions to Japanese people. NICT is also responsible for continuously enhancing relevant technologies and preparing for future threats.

Our system uses secret computation, which is essential in secret sharing. It is promising to apply the secret computation technology in cloud services, such as statistical data computations while protecting users' privacy information attached to the data. NICT will continue to pursue functional enhancement in communications technology, thereby fulfilling our duty to offer safe information communications systems.

Acknowledgments

A part of this research and development was conducted with the support of the Innovative Research and Development Promotion Program (ImPACT) designed by the Council for Science, Technology and Innovation. We thank the staff of the organizations participating in the ImPACT project "Realization of Advanced Knowledge infrastructure serving for connecting Quantum Artificial Brains via Networks," for their support and fruitful discussion.

References

1 P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," Proceeding of the 35th Annual Symposimu on Fundations of Computer Science, pp.12–134 (IEEE Computer Society Press, Los Alamitos, 1994).

2 J. Hoffstein, J. Pipher, and J. H. Silverman, "NTRU: A Ring based Public Key Cryptosystem," ANTS-III Proceedings of the Third International Symposium on Algorithmic Number Theory, pp.267–288 (ANTS-III, London, 1998).

3 O. Goldreich, S. Goldwasser, and S. Halevi, "Public-Key Cryptosystems from Lattice Reduction Problems," Proceeding of CRYPTO 1997 pp.112–131 (Springer, Heidelberg, 1997).

4 http://csrc.nist.gov/groups/ST/post-quantum-crypto/documents/call-for-proposals-draft-aug-2016.pdf

5 C. H. Bennett and G. Brassard, "Quantum cryptography: public-key distribution and coin tossing," in Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing (Institute of Electrical and Electronics Engineers, New York, 1984), pp.175–179

6 N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, "Quantum cryptography," Rev.

Mod. Phys. 74(1), pp.145–195 (2002)
7 K. Yoshino, T. Ochi, M. Fujiwara, M. Sasaki, and A. Tajima, "Maintenance-free operation of WDM quantum key distribution system through a field fiber over 30 days," Opt. Express 21(25), pp.31395–31401 (2013).
8 J. F. Dynes, I. Choi, A. W. Sharpe, A. R. Dixon, Z. L. Yuan, M. Fujiwara, M. Sasaki, and A. J. Shields, "Stability of high bit rate quantum key distribution on installed fiber," Opt. Express. 20(15), pp.16339–16347 (2012).
9 D. Stucki, M. Legre, F. Buntschu, B. Clausen, N. Felber, N. Gisin, L. Henzen, P. Junod, G. Litzistorf, P. Monbaron, L. Monat, J.-B. Page, D. Perroud, G. Ribordy, A. Rochas, S. Robyr, J. Tavares, R. Thew, P. Trinkler, S. Ventura, R. Voirol, N. Walenta, and H. Zbinden, "Long-term performance of the SwissQuantum quantum key distribution network in a field environment," New J. Phys. 13(12), 123001, 1–18 (2011).
10 M. Peev, C. Pacher, R. Alleaume, C. Barreiro, W. Boxleitner, J. Bouda, R.Tualle-Brouri, E. Diamanti, M. Dianati, T. Debuisschert, J. F. Dynes, S.Fasel, S. Fossier, M. Fuerst, J.-D. Gautier, O. Gay, N. Gisin, P. Grangier, A. Happe, Y. Hasani, M. Hentchel, H. Hübel, G. Humer, T. Länger, M.Legre, R. Lieger, J. Lodewyck, T. Lorünser, N. Lütkenhaus, A. Marhold, T. Matyus, O. Maurhart, L. Monat, S. Nauerth, J.-B. Page, E. Querasser, G. Ribordy, A. Poppe, L. Salvail, S. Robyr, M. Suda, A. W. Sharpe, A. J. Shields, D. Stucki, C. Tamas, T. Themel, R. T. Thew, Y. Thoma, A. Treiber, P. Trinkler, F. Vannel, N. Walenta, H. Weier, H. Weinfurter, I. Wimberger, Z. L. Yuan, H. Zbinden, and A. Zeilinger, "The SECOQC quantum key distribution network in Vienna," New J. Phys. 11(7), 075001/1-37 (2009).
11 M. Sasaki, M. Fujiwara, H. Ishizuka, W. Klaus, K. Wakui, M. Takeoka, A. Tanaka, K. Yoshino, Y. Nambu, S. Takahashi, A. Tajima, A. Tomita, T. Domeki, T. Hasegawa, Y. Sakai, H. Kobayashi, T. Asai, K. Shimizu, T. Tokura, T. Tsurumaru, M. Matsui, T. Honjo, K. Tamaki, H. Takesue, Y. Tokura, J. F. Dynes, A. R. Dixon, A. W. Sharpe, Z. L. Yuan, A. J. Shields, S. Uchikoga, M. Legre, S. Robyr, P. Trinkler, L. Monat, J.-B. Page, G. Ribordy, A. Poppe, A. Allacher, O. Maurhart, T. Langer, M. Peev, and A. Zeilinger, "Field test of quantum key distribution in the Tokyo QKD Network," Opt. Express, 19(11), pp.10387–10409 (2011).
12 A. Shamir, "How to share a secret," Communications of the ACM, 22, pp.612–613 (1979).
13 M. Fujiwara, A. Waseda, R. Nojima, S. Moriai, W. Ogata, and M. Sasaki, "Unbreakable distributed storage with quantum key distribution network and password-authenticated secret sharing," Sci. Reports, 6, 28988-1-8 (2016).
14 http://www.jgn.nict.go.jp/
15 M. Sasaki, M. Fujiwara, R-B. Jin, M. Takeoka, T. S. Han, H. Endo, K. Yoshino, T. Ochi, S. Asami, and A. Tajima, "Quantum Photonic Network: Concept, Basic Tools, and Future Issues," J. Selected Topics in Quant. Elec., 21, 6400313 (2015).
16 M. Fujiwara, T. Domeki, S. Moriai, and M. Sasaki, "Highly secure network switches with quantum key distribution systems," Int. J. Network security 17, pp.34–39 (2015).
17 T. Hirano, H. Yamanaka, M. Ashikaga, T. Konishi, and R. Namiki, "Quantum cryptography using pulsed homodyne detection," Phys. Rev. A68, 042331 (2003).
18 K. Shimizu, T. Honjo, M. Fujiwara, T. Ito, K. Tamaki, S. Miki, T. Yamashita, H. Terai, Z. Wang, and M. Sasaki, "Performance of long-distance quantum key distribution over 90-km optical links installed in a field environment of Tokyo metropolitan area," IEEE J. Lightwave tech. 32, pp.141–151 (2014).
19 http://www.sequrenet.com/datasheets/datasheet_cygnus.pdf

Mikio FUJIWARA, Ph.D

Research Manager, Qunatum ICT Advanced Development Center, Advanced ICT Reseach Institute
Quantum key distribution, Photon detection techology, Cryogenic electronics

Atsushi WASEDA, Dr. Eng.

Senior Researcher, Security Fundamentals Laboratory, Cybersecurity Research Institute
Information security

Ryo NOJIMA, Dr. Eng.

Research Manager, Security Fundamentals Laboratory, Cybersecurity Research Institute
Cryptology, cryptoprogocol, information security, privacy, security

Shiho MORIAI, Dr. Eng.

Director, Security Fundamentals Laboratory, Cybersecurity Research Institute
Cryptographic technology, security evaluation, privacy protection

Wakaha OGATA, Dr. Eng.

Pofesser, Tokyo Institute of Technology, School of Engineering
Cryptology, public-key cryptosystem, digital signature ,cryptographic protocol

Masahide SASAKI, Ph.D

Distinguished Researcher, Advanced ICT Research Institute
Quantum communication, Quantum cryptography