Information theoretically secure distributed storage system with quantum key distribution network and password authenticated secret sharing scheme

M. Fujiwara[1*], A. Waseda[1], R. Nojima[1], S. Moriai[1], W. Ogata[2] and M. Sasaki[1]

[1]National Institute of Information and Communications Technology (NICT), 4-2-1 Nukui-kita, Koganei, Tokyo 184-8795, Japan

[2]Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552 Japan

*fujwara@nict.go.jp

A quantum key distribution (QKD)[1] allows two users to share random numbers with the unconditional security based on the fundamental laws of physics. Combining a QKD with one time pad encryption (OTP), communication with unconditional security can be realized. A QKD system, however, does not guarantee the security of stored data. Shamir's (*k, n*)-threshold secret sharing (SS)[2] 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, provides information theoretical security. Therefore, combination of a QKD system and SS scheme is a combination for secure data transmission and storage. However, assumed is 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 secret sharing scheme and secure transmission using quantum key distribution, and demonstrate it in the Tokyo metropolitan area (≤90km)[3].

Our newly proposed password-authenticated secret sharing scheme is based on Shamir's SS scheme. In the SS scheme, secret data *D* is divided into *n* pieces of shares $f_{\mathrm{D}}(a_1), f_{\mathrm{D}}(a_2), \dots, f_{\mathrm{D}}(a_n)$, where f_{D} is a random polynomial of degree at most $k-1$ with a free coefficient representing the secret data *D* itself, and $a_1, a_2, \dots, a_n$ are public values. Then, knowledge of any *k* or more pieces of $f_{\mathrm{D}}(a_i)$ makes *D* easily computable with Lagrange interpolation. For example, to determine the free coefficient *D* in the quadratic function $f_{\mathrm{D}}(x)$, at least three shares are required because there are three unknown parameters in the quadratic function $f_{\mathrm{D}}(x)$. On the other hand, knowledge of any $k-1$ or fewer pieces of $f_{\mathrm{D}}(a_i)$ leaves *D* completely undetermined (in the sense that all its possible values are equally likely). That is, the attacker cannot recover the original data from less than the threshold *k* of shares even by using unlimited computational resources. Such a scheme is called (*k, n*)-threshold scheme. Shamir's scheme enables secret calculations (addition and multiplication). In fact $f_{\mathrm{D}^{(1)}}(a_i) + f_{\mathrm{D}^{(2)}}(a_i)$ becomes a share of addition of two secret data $D^{(1)}$ and $D^{(2)}$. Likewise, $f_{\mathrm{D}^{(1)}}(a_i) \times f_{\mathrm{D}^{(2)}}(a_i)$ is used as the share of $D^{(1)} \times D^{(2)}$. In multiplying process, however, the degree of the polynomial $f_{\mathrm{D}^{(1)}}(x) \times f_{\mathrm{D}^{(2)}}(x)$ is $2k-2$. So, $2k-1$ of shares are necessary to reconstruct $D^{(1)} \times D^{(2)}$. In our password-authenticated SS scheme, these characteristics are utilizes. Shares of the secret data and the password are calculated, sent, and stored in the storage servers. We assume there are four storage servers (n = 4) denoted as 1, 2, 3, and 4, and an attacker can corrupt at most one storage server. In this case, the secret data *D* are shared by the (3,4)-threshold scheme. On the other hand, the password shares must be generated by using the (2,4)-threshold scheme due to the multiplying process. Our password-authenticated secret sharing scheme comprises three phases: (1) **Registration phase** in which data shares $f_{\mathrm{D}_i}(1)$, $f_{\mathrm{D}_i}(2)$, $f_{\mathrm{D}_i}(3)$, $f_{\mathrm{D}_i}(4)$ and password shares $f_{\mathrm{P}}(1)$, $f_{\mathrm{P}}(2)$, $f_{\mathrm{P}}(3)$, $f_{\mathrm{P}}(4)$ are computed and stored in the storage servers; (2) **Pre-computation and communication phase** in which each storage server generates a random number by a physical random number generator, and computes their shares $f_{\mathrm{R}j}(1), f_{\mathrm{R}j}(2), f_{\mathrm{R}j}(3), f_{\mathrm{R}j}(4)$, as well as prepares shares of the data “0” $f_{0j}(1), f_{0j}(2), f_{0j}(3), f_{0j}(4)$ to randomize the shares in data reconstruction without changing the value of shared data. Then all the storage servers send the shares to each other; (3) **Data reconstruction phase** in which the data owner requests the storage servers to send back the data shares using the password P'. The data owner sends the password shares $f_{\mathrm{P}'}(1)$, $f_{\mathrm{P}'}(2)$, $f_{\mathrm{P}'}(3)$ to storage servers , and receives $F_i(x) = \left(f_{\mathrm{P}}(x) - f_{\mathrm{P}'}(x)\right)f_{\mathrm{R}}(x) + f_0(x) + f_{\mathrm{D}_i}(x)$. If the password is correct, then the original data is reconstructed.

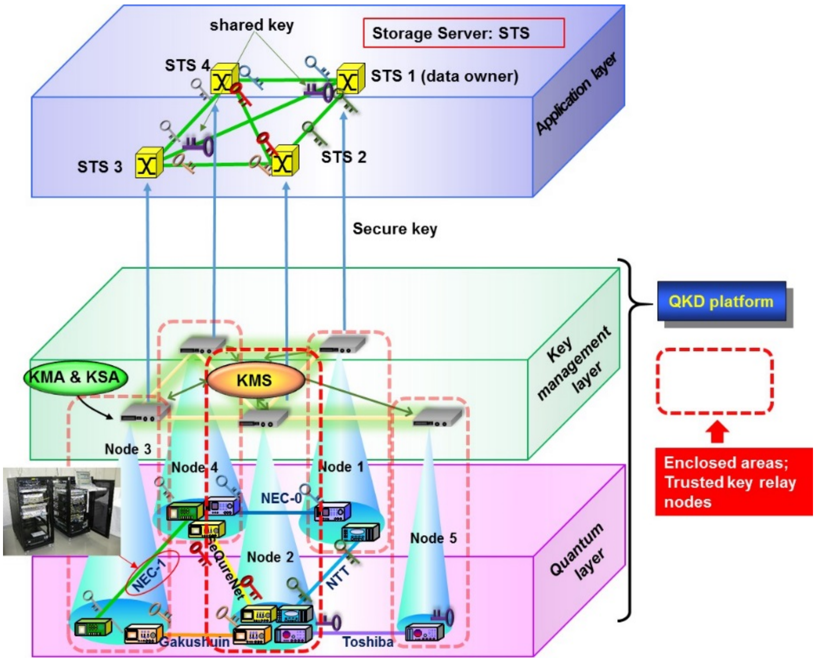

Figure 1. Schematic view of the layer structure of our distributed storage system.

The whole system consists of two blocks, the application layer and the QKD platform. The password-authenticated secret sharing scheme is implemented in the application layer on which the data owner and the storage servers (STSs) are setup. The keys are generated in each QKD link in the quantum layer, pushed up to key management agents (KMAs) in the key management layer. The KMAs are in the trusted key relay nodes, store the keys, and if necessary, relay the keys. To support various applications, key supply agents (KSAs) are introduced at each KMA. A key management server (KMS) is located and carry out the centralized key management. Having requests from the data owner and the storage servers, the KSAs supply them the random number key streams.

Such a distributed storage system with (3,4) threshold for secret data is physically constructed with four nodes which are denoted as Node 1 to 4 of the QKD platform[3] in Fig. 1, and the data owner is located in the same node as "storage server 1" just for experimental convenience. Here we should assume that in this node, "storage server 1" may be corrupted by attackers but the data owner can be protected against attackers. All the communications between the data owner machine and the storage servers and among the storage servers are carried out based on the QKD-enhanced Internet protocol (IP) [4], in which all IP packets are OTP-encrypted and further authenticated with message authentication code (MAC) based on Wegman-Carter protocol[5] by using the keys supplied from the QKD platform. The QKD links named, NEC-0, NEC-1, Toshiba, NTT, Gakushuin, and SeQurenet in Fig. 1 denote QKD link providers.

We prepare secret data D whose size is 6955, 13695, and 46000 bytes. All calculations are made in a finite Galois field with prime order q. Mersenne primes $q = 2^m - 1$ with indices m=521, 1279, 2203, 3217, 4253, 10041, 11213, 19937, 23209, 44497, and 86243 are used to calculate shares. Secret data D, which has generally a much longer length, needs to be divided into pieces of $(m-1)$-bit block, say l pieces. Performance of our system depends on the size of q. This is because (1) the computational time of the shares increases roughly in the square of bit length of q and (2) using a smaller prime q increases the number of blocks l, and hence a longer processing time is required for dividing/managing the blocks and sending IP packets. There would be a good balance point, conditioned on the maximum payload size of 1500 bytes. In our experiments, the best performance can be found in the range of q with $11213 \leq m \leq 23209$. The total length of keys for OTP-encryption required to store and retrieve is about 30 times as long as the original secret data, and all process can be finished within 30 s for 46000 bytes data.

Our newly proposed distributed storage system will also be useful to realize secure data relay via classical nodes. And, depending on a QKD network topology, we can relax requirements to the current trusted key relay node in the QKD network, not resorting to a greatly costly quantum repeater paradigm.

References

(1) Gisin, N., Ribordy, G., Tittel, W. & Zbinden, H. Quantum cryptography. *Rev. Mod. Phys.*, **74**, 145–195 (2002).

(2) Shamir, A. How to share a secret. *Communications of the ACM*, **22**, 612-613 (1979).

(3) Sasaki, M. *et al.* Quantum Photonic Network: Concept, Basic Tools, and Future Issues. *J. Selected Topics in Quant. Elec.*, **21**, 6400313 (2015).

(4) Fujiwara, M., Domeki, T..Moriai, S. & Sasaki, M. Highly secure network switches with quantum key distribution systems. *Int. J. Network security* **17**, 34-39 (2015).

(5) Wegman, M. & Carter, L. New hash functions and their use in authentication and set equality. *J. Comp. and Sys. science*, **22,** 265-279 (1981).