

Genome Sequence Data Storage System using distributed storage system on QKD network

Kazuaki Doi¹, Ririka Takahashi¹, Akira Murakami¹, Mamiko Kujiraoka¹, Alexander R. Dixon¹, Yoshimichi Tanizawa¹, Hideaki Sato¹, Muneaki Shimada³, Yasunobu Okamura^{2, 4}, Fuji Nagmi^{2, 4}, Mikio Fujiwara⁵

¹Corporate Research & Development Center, Toshiba Corporation

²Tohoku Medical Megabank Organization, Tohoku University

³Tohoku University Hospital, Tohoku University

⁴Advanced Research Center for Innovations in Next-Generation Medicine, Tohoku University

⁵National Institute of Information and Communications Technology

E-mail: kazuaki.doi@toshiba.co.jp

Abstract

We developed a genome sequence data storage system using a distributed storage system on a quantum key distribution (QKD) network [1] and have successfully demonstrated secure storage and data reconstruction for genome sequence data. The proposed system thus has potential for use as a distributed storage system in genome analysis.

Methods

We developed a highly secure storage system based on a QKD network and secret sharing technology. In this system, genomic data are divided in to 3 pieces of shares by using random numbers, and shares are stored in share holders. To reconstruct data, two of shares are necessary at least. Communication channels between are encrypted by Vernam's one-time pad (OTP) using secure key provided from a QKD network.

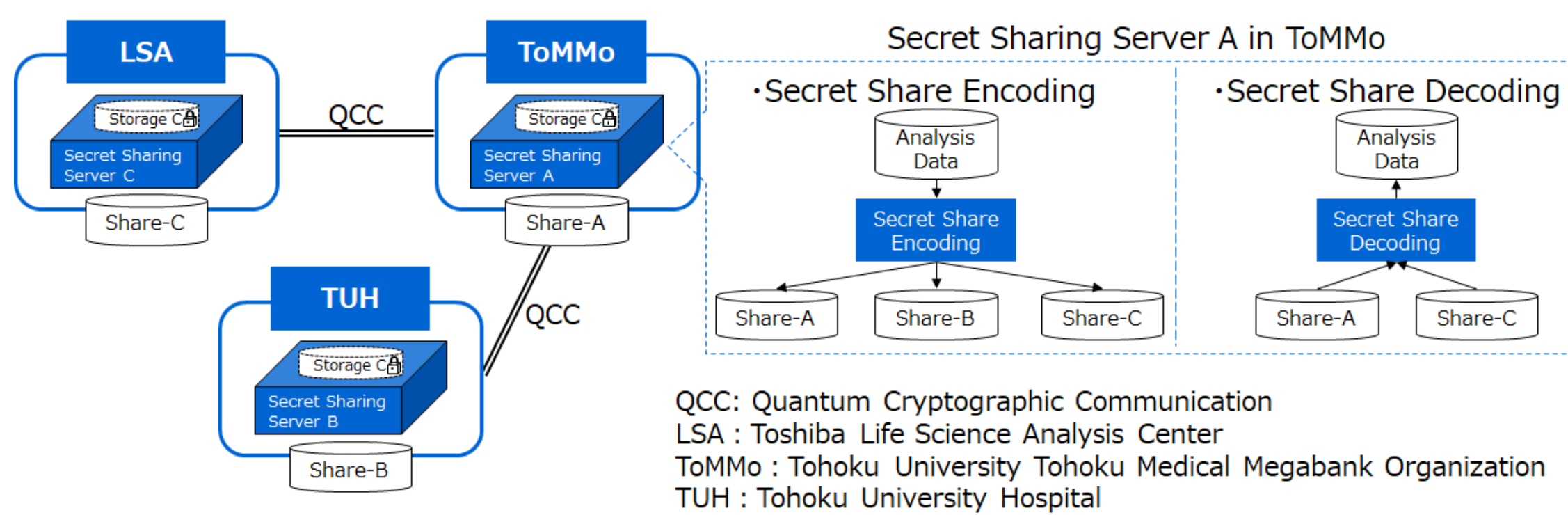


Figure 1. Overview of the genome sequence data storage system in Sendai.

Table 1. System specifications for the genome sequence data storage system.

Item	Details
Cryptographic method	Vernam's One-time pad (OTP)
Use of encryption keys	Pre-shared and pre-stored on a secret-sharing server
Secret-sharing method	(2,3) threshold XOR based secret-sharing method

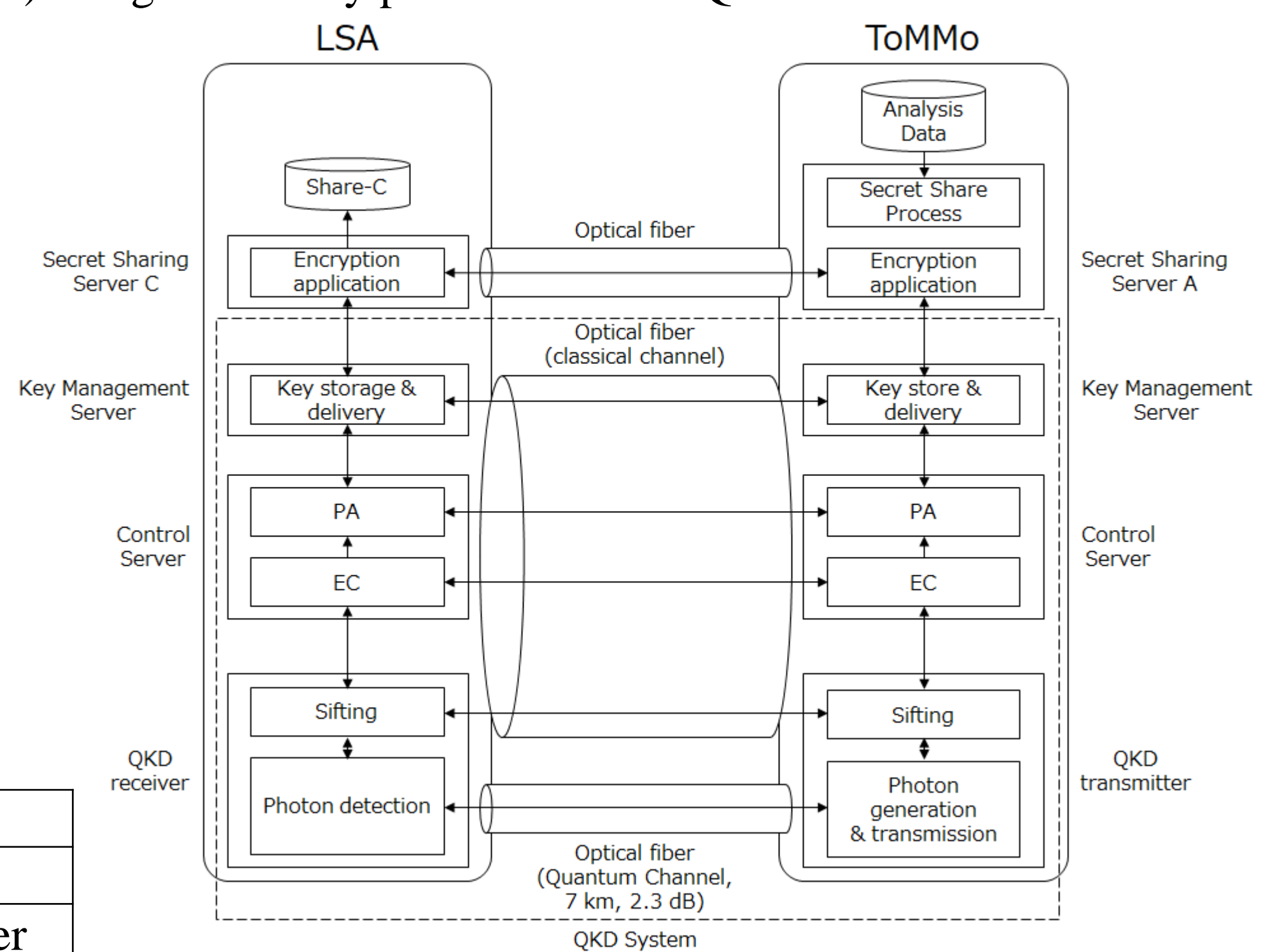


Figure 2. Schematic diagram of the genome sequence data storage system. [2]

Share holders are set in LSA, ToMMo, TUH. QKD enhanced OTP encryption links are applied between ToMMo and LSA, and between ToMMo and TUH respectively. At first three shares of genomic data (named share A, share B, and share C) are calculated in ToMMo by a (2,3) threshold XOR based secret-sharing method. After calculation, Share A is stored in ToMMo, and share B, and share C are transferred to TUH and LSA using OTP encryption, respectively. Cryptographic communications are performed between secret-sharing servers, and quantum cryptographic keys for transmitting shares are acquired from the key management server before transmitting the share [3].

Results

In this test, genome sequence data were stored in FASTQ files. The data size for one sample of the whole genome sequence was 81.35 GB. Processing time for decentralized storage was 30 min 32 sec and decentralized storage throughput was 355.24 Mbps. Processing time for data reconstruction was 21 min 35 sec, and data reconstruction throughput was 502.24 Mbps. Processing time for remote backup of 100 samples should be approximately 2.12 day in theory.

Table 2. Test results for data storage and data reconstruction.

Parameters	Results
Amount of transmitted genome analysis data	81.35 GB (whole genome analysis data for HapMap sample, one sample, FASTQ files)
Processing times	(A) Decentralized storage: 30 m 32 sec (355.24 Mbps) (B) Data reconstruction: 21 m 35 sec (502.54 Mbps) (C) Remote backup (100 samples, in theory): 2.12 day (355.24 Mbps)

Conclusion

We developed a genome sequence data storage system using quantum secure storage technology and succeeded in field tests of distributed storage and data reconstruction. In the future we will investigate developments for improving process speed and convenience.

REFERENCES

- [1] M.Fujiwara et al., Scientific Reports volume 6, Article number: 28988 (2016)
- [2] Z. Yuan et al., J. Lightwave Technol., doi: 10.1109/JLT.2018.2843136 (2018).
- [3] ETSI GS QKD 014

Acknowledgement

This work was supported by the Council for Science, Technology and Innovation (CSTI) and the Cross-ministerial Strategic Innovation Promotion Program (SIP) "Photonics and Quantum Technology for Society 5.0" (Funding agency: QST).