35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Applying Secret Sharing for HIS Backup Exchange

Tomohiro Kuroda, Eizen Kimura, *Member, IEEE*, Yasushi Matsumura, Yoshinori Yamashita, Haruhiko Hiramatsu, Naoto Kume, Atsushi Sato, *Non-member*

***Abstract*—To secure business continuity is indispensable for hospitals to fulfill its social responsibility under disasters. Although to back up the data of the hospital information system (HIS) at multiple remote sites is a key strategy of business continuity plan (BCP), the requirements to treat privacy sensitive data jack up the cost for the backup. The secret sharing is a method to split an original secret message up so that each individual piece is meaningless, but putting sufficient number of pieces together to reveal the original message. The secret sharing method eases us to exchange HIS backups between multiple hospitals. This paper evaluated the feasibility of the commercial secret sharing solution for HIS backup through several simulations. The result shows that the commercial solution is feasible to realize reasonable HIS backup exchange platform when template of contract between participating hospitals is ready.**

I. Introduction

Business continuity plan (BCP) is indispensable for hospitals to fulfill its social responsibility in case of disasters. On the other hand, to maximize cost-efficiency of social healthcare system is one of the critical social demands of any advanced countries.

Recent rapid introduction of the information communication technologies into hospitals pushes up the importance of the hospital information systems (HIS). HIS provides all health records, clinical guidelines, available laboratories and medicines, and other required information, and support all clinical activities through barcode enables medication administration (ABMA) systems, scheduling systems, and other advanced features. Therefore, the loss of HIS may drastically decrease the performance of clinicians in the organization with the advanced HIS. Thus, to secure data stored in HIS (HIS data) is one of the most important issues of today's healthcare BCP.

The mutual exchange of HIS data backup seems good solution to secure data within legal limitations [1][2]. After the Tohoku disaster, Ishinomaki city hospital, which lost all their storage servers by the Tsunami, could retrieve its HIS data from the backup at Yamagata city hospital stored under their mutual backup exchange agreement.

However, to establish such mutual backup agreement is not so easy. The patient record is "the ultimate privacy sensitive data." Therefore, the social public including laws urges us to make ultimate security countermeasure against the data storage to secure health records' secrecy in it. To provide properly secured storage for the backup of other organization requires a lot of cost.

As the cause of all the limitations is the fact that the target data is the ultimate privacy sensitive data, the key technology to enable more flexible treatment of the backup is to remove privacy sensitivity from the data. This paper proposes and evaluates a method to enable us to have cost-effective distributed backup by properly removing privacy-sensitivity from the original secret data.

II. Secret Sharing

Secret Sharing is a method to distributing a secret among multiple participants who shares part of the secret. Secret Sharing is also called as Secret Splitting, as the method is to split a original secret message up so that each individual piece is worthless, but putting sufficient number of pieces together to reveal the original message. The method is originally proposed as the electric tally, which is to distribute risk of losing and abusing of key among multiple administrators by Blackley [3] and Sharman [4] independently in 1979. The theory is now widely applied for data encryption methods such as Pretty Good Privacy (PGP).

(k, n) threshold scheme is an approach to realize secret sharing, which splits the original secret message into n pieces so as to reveal the original message from k pieces among them (k-out-of-n secrecy). For example, $(2,3)$ threshold scheme encrypt secret message M by equation (1) and denoted by three points A, B, and C on the line generated by equation (1). Finally, the three points are stored separately.

$$y - nx + M \tag{1}$$

As a straight line defined by two points, sets of two out of three pieces of data shown in equation (2) reveals original secret message S. Such data shown in equation (2) is called qualified set. On the other hand, sets of single piece of data shown in equation (3), called forbidden set, won't disclose the original secret message.

$$S_1 = \{x, y | x, y \in \{A, B, C\}, x \neq y\} \tag{2}$$

$$S_2 = \{x | x \in \{A, B, C\}\} \tag{3}$$

Unlike conventional computationally secure encryption methods, the security of (k, n) threshold scheme is based on

*This research is partly funded by the Strategic Information and Communications R&D Promotion Programme (SCOPE) of the Ministry of Internal Affairs and Communications (MIC) of Japan.

Tomohiro Kuroda and Naoto Kume are with Division of Medical Information Technology and Administrative Planning, Kyoto University Hospital, Japan. (Email: tomo@kuhp.kyoto-u.ac.jp)

Eizen Kimura is with Department of Medical Informatics, Ehime University Hospital, Japan

Yasushi Matsumura is with Department of Medical Informatics, Osaka University Hospital, Japan.

Yoshinori Yamashita is with Department of Medical Informatics, University of Fukui Hospital, Japan.

Haruhiko Hiramatsu is with Department of Medical Informatics, Hyogo College of Medicine, Japan.

Atsushi Sato is with NRI Secure Technologies Ltd., Japan.

information theory. Even a processer with unlimited calculation power and memory cannot disclose the original message from the forbidden set.

In (k, n) threshold scheme, the size of each piece cannot be smaller than the size of the original message. In order to decrease total size of data, Yamamoto [5] proposed (k, L, n) threshold scheme. Although sets of more than $k - L$ and less than k pieces, called ramp set, reveal part of original secret message in this scheme, the scheme makes the size of each piece into $1/L$ of original message. Therefore, total size of distributed data becomes n/L.

Matsumoto et al [6] evaluated best parameter setting for (k, L, n) secret sharing from performance point of view. They try to minimize the size of required data storage (and the size of transmitted data to store secret data) n/L, the size of transmitted data to retrieve the original data k/L, the number of connections to store data (and the number of storage) n, and the number of connections to retrieve the original data k under two mandatory conditions; one peace of data won't decrease the ambiguity of original data (4), and original data can be retrieved even when one piece is lost (5).

$$1 \leq k - L \quad (4)$$

$$k \leq n - 1 \quad (5)$$

In order to keep size of storage *n/L* small enough, such as twice as big as original data, formula (4) and (5) results in $n \geq 4$. As n should be as small as possible to reduce number of required storage, (3,2,4) secret sharing is the optimal solution. In this secret sharing, lost of one piece of data will not decrease secrecy and availability of the original secret data. Two times of the original data should be transmitted to store, and 1.5 times should be transmitted to retrieve.

III. Mutual HIS Backup Exchange Platform

Secret sharing using (k, L, n) scheme enable hospitals to backup the HIS data by removing privacy-sensitive nature from its distributed pieces. Additionally, the method provides $n - k$ redundancy. Once more than n hospitals start mutual exchange of backups, each hospitals has no need to prepare ultra-secure storage environment and ultra-strict management for exchanged data, which is not privacy-sensitive anymore. The method also enable hospitals to utilize low-cost cloud storage service for backup, and, consequently, to secure each piece of data by the redundancy of cloud storage services.

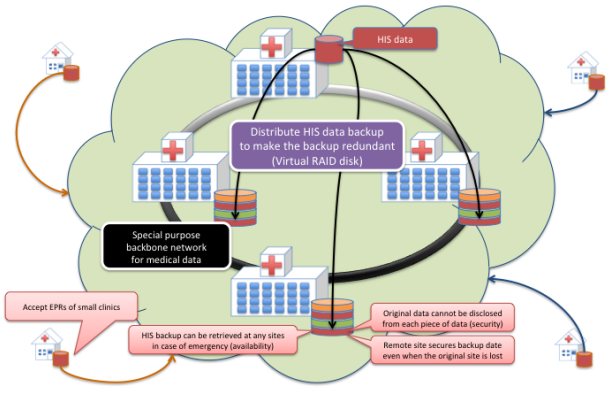

Figure 1. Sketch of HIS backup exchange platform

Figure 1 shows the conceptual sketch of the proposed HIS backup exchange platform. Each participating hospitals contribute certain amount of storage to the platform and distribute pieces to the platform. When a hospital lost their HIS data due to disaster or other reason, the data can be retrieved from any participating hospitals. Therefore, the patient of the lost hospital can be simply transferred to other sites to have continuous healthcare until the lost hospital recovers its function.

The required storage size in (k, L, n) secret share is n/L times bigger than the original data. Suppose total size of backups is X_{total} bytes and number of hospitals participating the platform is m, a participating hospital need to contribute $X_{tota;} \cdot n/L \cdot 1/m$ bytes storage. However, the secret share will not distribute the stored data equally to all participating hospitals. It splits the original data into n pieces. Therefore, in case the size of biggest dataset to backup is X_{max} bytes, each participating hospital need to provide $X_{max} \cdot 1/L$ bytes to create safer environment.

In order to secure at least one generations of backup, the platform need to provide double buffer architecture like video cards of computer. Otherwise, ta hospital loses whole backup data, if the hospital is crashed while it overwriting its backup.

Additionally, a temporary storage to retrieve data in case of disaster should be bigger than biggest backup data.

Therefore, a participating hospital needs to contribute storage of S bytes, as defined by equation (6).

$$(2n/mL + 1)X_{total} \leq S \leq (2/L + 1)X_{max} \quad (6)$$

Matsumoto et al [6] revealed that (3,2,4) threshold scheme secret sharing provides the best performance. Therefore, five hospitals are enough to startup mutual exchange platform. In case all participating hospitals equally have 40 Giga bytes of HIS data, each hospital need to contribute 80 Giga bytes of storage.

The platform may accept HIS backup from small clinics as shown in figure 2. As the small clinics cannot contribute storage resources, another participating scheme to contribute the platform by paying management fee should be required. To design membership variations like FON or other shared networks [7] is required.

IV. Feasibility of The Platform

A. Performance of the platform

To confirm the feasibility of the proposal, the authors evaluate the performance of a commercially available secret sharing solution.

The authors first evaluated the performance through simulation.

The simulation performed under the scenario shown in figure 2. In the first step, namely "before disaster," five hospitals with 40 Giga bytes of HIS data distributes HIS data each other. In the second step, namely "under disaster," a certain disaster hits a hospital (hospital A) and the hospital lose its own HIS and its data center for mutual backup exchange. In this condition, another hospital (hospital B) may retrieve the HIS data of hospital A to treat patients of hospital

A. At the same time all the distributed data must be reorganized to recover redundancy of the distributed data. After a while, the hospital A recovers from the damage of the disaster and back to normal operation. In this stage, namely "after disaster", all the data again needed to be reorganized to have better distribution.

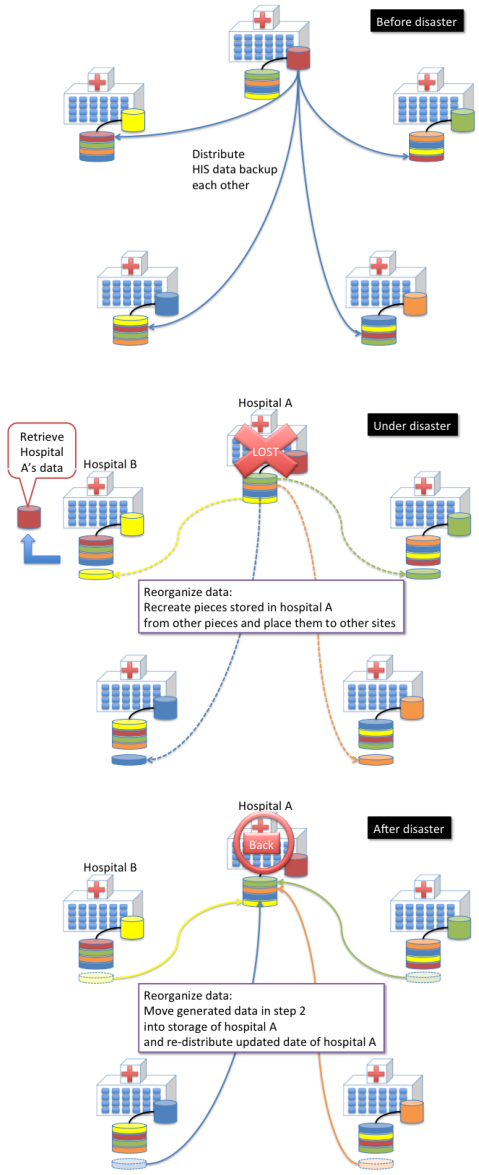

Figure 2. Simulation scenario

The authors performed the simulation under two setups.

In the first setup, the authors have deployed SecureCube / Secret Share of NRI Secure Technologies Ltd. [8] into StarBED3 [9][10], the large-scale network experiment environment of National Institute of Information and Communications Technology (NICT). In this setup, the participating hospitals connected by the network of 1Giga bit per second and almost no communication latency, and each of them has 40 files of 1 Giga byte data as HIS data.

In the second setup, the authors evaluated same scenario on the commercial secret sharing service using SecureCube/ Secret Share using the Internet. The pseudo hospital with 10 files of 1Giga byte data is implemented on a computer connected to the local area network of Kyoto University Hospital, and connected to the multiple data centers through the https proxy of the Kyoto University [11]. In this test, no reorganization process is performed.

The Table I shows the result of the simulation. The result of the test over the Internet connections is scaled to fit 40 files of 1 Giga byte of data.

TABLE I. SIMULATION RESULT

	Before	**Under disaster**		**After**
	Backup	***Retrieve***	***Reorganize***	***Reorganize***
StarBed	03:05:35	02:20:03	03:29:47	03:07:43
Internet	04:10:47	06:24:00	--	--

The detailed analysis tells that the bottleneck of the performance was in the file IO. The solution utilized the file IO of the operating system, and it stores temporary data into the files instead of memories for implementation stability reason. Thus, the consumed bandwidth was just up to 29.5Mbps. However, the performance of the test over the Internet was clearly low. The current commercial solution upload the data in five parallel threads, and retrieves the data in single process, on the other hand. The result tells that the latency of the network affected the performance of whole system. As a matter of fact, when the authors try to backup the data in single process, it took 1.5 times longer.

Theoretically, once the file IO is optimized, the performance of file IO depends on the performance of the disk device. The performance of the conventional consumer products is about 30 Mega bytes per second, eight times faster than the result. Additional optimization is required.

B. Data size of HIS data

The authors measured the data size of HIS of five university hospitals participating this project.

Kyoto University Hospital (KUHP) has approximately 50TB including about 40TB of picture archiving and communication system (PACS) data stored since 1993. Among the gigantic data, the main component of HIS data stored in IBM DB2 database, including CPOE, laboratory results, various types of summaries with compressed key images, and EPR is about 1TB. IBM CIS, the main component of the HIS of KUHP, backs up the data by flash copy feature of DB2, and stores it to its backup storage everyday. Although KUHP does not applying, the backup function of DB2, which compresses the data in 1/3, can make the backup data about 350GB. Both the university of Fukui Hospital and Ehime University, which also uses IBM CIS, has 1.3TB of data. Thus, the three university hospitals have the same size of HIS data.

Osaka University Medical Hospital (OUMH), which uses NEC MegaOak HIS, has 4.2TB of HIS data with 2.1TB of document archive in the document archiving and communication system (DACS) [12]. Hyogo College of Medicine Hospital (HCMH), which uses Fujitsu HOPE/EGMAIN HIS, has 19TB of HIS data. Thus, the HIS software differs the size of stored data so drastically.

V. Discussion

A. Performance

The result of the simulation and data size of KUHP tells that about two days are required to backup, and that about four days are required to retrieve the main component of HIS data over the Internet. In case of OUMH, it goes up to 23 days, and it goes up to 70 days. Thus, the compression of the initial backup data is key issue to make it realistic. The authors will evaluate compression performance for HIS data.

If the secret sharing performance is optimized up to 240Mbps, the required times becomes 4 hours for KUHP, one day for OUMH, and a week for MCMH.

The experiences of Kobe disaster in 1995 and Tohoku disaster in 2011 told us that the required clinical information for the disaster medicine is quite limited. Therefore, if the time to retrieve whole HIS data backup is short enough, we can just retrieve whole HIS data instead of paying lot effort to develop subset of the clinical data. If not, we should design sub-dataset for quick reference such as DACS.

Although the required bandwidth, 240Mbps in the worst case, is far less beyond the current network setup, the damage on the network infrastructure and the flood of communication massage following the disaster may harm the performance of the information network dramatically. Therefore, the authors need to simulate various conditions. As StarBED3 [9] enable us to simulate various networking environment.

B. Storage size and cost

The required storage to participating the platform using (3,2,4) threshold secret sharing is two times of the original data. As the current unit price of storage is about $0.1/GB per month, the whole payment for the storage goes $840 per year. Even for MCMH, it goes $3890 per year. As the secret share enable us to use cloud storage service by deleting the privacy sensitivity from the HIS data, it increases the possibility to have offsite backup within feasible cost.

C. Contract

Many lawyers agree that each piece of data generated by secret sharing is not privacy sensitive information. Therefore, all the hospitals can participate the platform without hard negotiation related to data treatment.

Although the required storage to participating the platform is two times of its own HIS data size theoretically, the wide variety of the size of HIS data and the limitation of secret sharing platform cannot make the storage size calculation so simple, especially when the number of participating hospitals is limited. Additionally, in order to let small clinics with limited computational facilities to participate the platform without contributing storage, the platform should provide different type of contract. Designing template of several reasonable contracts is important question to start up the platform.

We have to keep in mind that this solution will make us face new challenges. If a hospital is lost, the medical professionals at another hospital may need to retrieve the EPRs of the lost hospital to provide medical support to the lost hospital. We should carefully design the agreement on the condition that when and how to allow the medical professionals of another hospitals to retrieve the data.

VI. Conclusions

This paper proposed the HIS data backup exchange platform using secret share technique, and studied its feasibility. The results show that secret sharing lowers the barriers to have redundant offsite backups for hospitals and to establish social platform to secure healthcare data under collaboration of multiple hospitals.

Although the simulation indicates the proposed platform is realistic, further detailed analysis using real HIS data is required to clears up remaining questions. The authors are developing mutual HIS backup exchange platform among five university hospitals in Japan for further evaluation. Under designing the mutual HIS backup exchange platform, the authors also design the required agreement to maximize the meanings of such social platform.

This research opens up the possibility of our proposal. The result of following project may provide further findings for creating really working social platform.

Acknowledgment

The authors would like to thank precious support of Hokuriku StarBED Technology Center, National Institute of Information and Communications Technology (NICT), NRI Secure Technologies Ltd., and IBM Japan.

References

[1] The Ministry of Healthcare, Labour and Welfare of Japan, *The Guideline for the Security Management of Medical Information Systems,* ver. 4.1, 2009.

[2] The Ministry of Internal Affairs and Communications of Japan, T*he Guideline Regarding Safety Control when ASP/SaaS Business Workers Handle Medical Information,* ver. 1.1, 2009.

[3] GR. Blackley, "Safeguarding Cryptographic Keys," *Proc. Nat. Comput. Conf*., p.313, 1979.

[4] A. Shamir, "How to Share a Secret," *Comm. ACM*, pp.612-3, 1979.

[5] H. Yamamoto, "Secret Sharing System Using (*k, L, n*) Threshold Scheme," *Electronics and Communications in Japan*, vol. 69, No. 9, pp.46-54, 1986.

[6] T. Matsumoto, T. Seito, A. Kamoshida, T. Shingai, A. Sato, "High-speed Secret Sharing System for Secure Data Storage Service," *Proc. Symp. Crypto. Inform. Security*, 1E2-4, 2012. (Japanese)

[7] C. Middleton, A. Potter, "Is it good to share? A case study if the FON and Meraki approaches to broadband provision," *Proc. ITS Biennial Conf.*, http://www.canavents.com/its2008/, 2008.

[8] NRI Secure Technologies, *SecureCube / Secret Share*. http://www.nri-secure.co.jp/service/cube/secretshare.html.

[9] T. Miyachi, T. Nakagawa, K. Chinen, S. Miwa, Y. Shinoda. "StarBED and SpringOS Architectures and Their Performance," Lec. N. Inst. Comput. Sci., Soc. Inform. Telecom. Eng., vol. 90, no. 1, pp.43-58, 2012.

[10] *StarBED Project*. http://www.starbed.org/.

[11] H. Takakura, Y. Ebara, S. Miyazaki, A. Sawada, M. Nakamura, Y. Okabe, "Structure and Security Strategy of a Secure Gigabit Network System, KUINS-III," *IEICE Trans. Comm*., vol. J86-B, no. 8, pp.1491-1501, 2003. (Japanese)

[12] H Takeda, "Updated Topics in Healthcare Informatics," *Proc. E-Health*, pp.1-4, 2010.