Pixel-Based Dispersal Scheme for Medical Image Survivability and Confidentiality

Nelmiawati, Mazleena Salleh, Malek Najib Omar
Faculty of Computing
Universiti Teknologi Malaysia
81310, Skudai, Malaysia
nelmia.wati@gmail.com, mazleena@fsksm.utm.my, malek2omar@hotmail.com

***Abstract*—Medical image survivability and confidentiality is an important concern for digital medical image storing in healthcare institution. Survivability issue arises when a storage server goes down due to unexpected disasters whilst medical image confidentiality issue arises due to disclosed data by third party on a conventional storage. This study explores secret sharing threshold scheme: Rabin's IDA, and Shamir's SSA, as potential approaches to address these issues. Currently, these schemes are widely known for providing data survivability and confidentiality in other fields. However, the aim of this project is to apply these secret sharing schemes through MIaDPACS to provide survivability and confidentiality on digital medical image. Pixels are extracted from the image, encoded and dispersed using Rabin's IDA into distributed storages. In addition, the secret key used in Rabin's IDA is also dispersed through Shamir's SSA. To reconstruct back the original image, it only requires a subset of the dispersed files with numbers equals to a threshold defined during dispersal. Experimental results conducted have shown that MIaDPACS able to provide survivability and confidentiality for digital medical image.**

Keywords—medical image, Rabin's IDA, Shamir's SSA

I. INTRODUCTION

Digital data is widely used in various aspects of human life because it offers cost-efficiency and flexibility on data manipulation, storage, and transmission. Radiology areas in healthcare institution have made a significant revolution to adopt digital technology by using digital modalities such as X-Ray, CT (Computed Tomography), MRI (Magnetic Resonance Imaging), USG (Ultrasonography), etc. Digital imaging in medicine provides an instance access to radiologist, doctor, and patient, regardless their location and without afraid of degrading the image quality through time in a storage system as what happens on analog film imaging. Moreover, it also providesdynamic range of contrast, dynamic levels of gray, and several manipulations that could improve accuracy on image analysis during diagnose and treatment of diseases.

Digital medical imaging technology has become an important part in health care ecosystem. Several studies mentioned that a lot number of digital medical image used in health care institutions. Study by [1]found that there were significant rise on usage of CT and MR in one tertiary hospital at United States from 1984 to 1993, while there were a decrease on usage of conventional analog plain films and fluoroscopy. Besides, there were about 26 million of CT examinations were conducted at the United States in 1998, and then by 2008 it became more than 70 million[2]. Moreover[3], found that there has been also a growing number usage of digital medical imaging at developing countries.

Digital medical images require a quite long retention storing period, it could be decades or even forever to be kept in a storage system [4]. Those images have to be able to be retrieved at any time they are needed without any tolerate data corruption. This requirement makes data survivability become one of important capability to be provided by a medical image storage system. The ability to provide data confidentiality is another important feature that is required. Only limited authorized parties granted to have access to the images.

Storing digital medical images in local centralized repositories as what is offered by current PACS (Picture Archiving and Communication System) does not satisfy survivability and confidentiality requirements. Failure that is occurred in this storage makes its stored images lost their survivability. A recovery might be done from a backup system, but it could not recover everything. The backup system will fail to recover any images that are stored after the last performed backup, because they are not yet there in the backup system. An approach of using RAID (Redundant Array of Independent Disk) as a storage media in the local repositories still could not really answer this survivability issue. RAID is a centralized storage that is located in a single site. Damage occurred in that site could make RAID fail to provide survivability for medical images stored in it. Furthermore storing medical image in local centralized repositories is not sufficient enough to provide confidentiality. Any unauthorized parties that are success to find

978-1-4799-6444-4/14/$31.00 ©2014 IEEE

a way to get access into the repositories will be able to get all the medical images stored in it.

Cloud storage service start becoming an approach that is introduced to store digital medical images, also known as cloud PACS. It offers an economic advantage to reduce cost storage, and good quality of service to provide survivability on storing medical images[5]. However this approach has disadvantages in addressing confidentiality. Uploading data into a cloud service makes it no more resistance for disclosure or unauthorized usage [5]. Moreover, data survivability provided through a cloud service is still questioned. Consumer lock-in and bankruptcy issue on cloud provider could lead to lost control to the stored medical images in it.

One possible alternative to address survivability and confidentiality issue on digital medical image storage is to use pixel-based data dispersal approach. A digital medical image is dispersed into several pieces, and each could be stored into different distributed storages system. A subset number of pieces dispersed images that is not less than a particular defined threshold have already enough to reconstruct the original digital medical image. Having numbers of dispersed data less than the threshold, means nothing to reconstruct the original image. This method could provide a better survivability and confidentiality to digital medical image.

This work discusses a pixel-based dispersal scheme to answer storage issues on medical image survivability and confidentiality. The rest of this paper is organized as following. Section II discusses related works that have been done for providing data survivability and confidentiality. Subsequently, Section III describes a proposed approach to develop a pixel-based dispersal scheme. It is then continued with result and discussion on the scheme implementation towards digital medical image survivability and confidentiality experiments in the Section IV. Finally, conclusion of the work is discussed in the Section V.

II. Related Works

Secret sharing threshold scheme is common implemented in a distributed storage system solution as discussed in [6], [7], and [8]. Rabin's IDA (Information Dispersal Algorithm) is chosen to disperse and encode medical image into n several pieces to be stored in different n sites. Then a subset of the pieces with numbers are not less than a defined threshold m, have already enough to fully reconstruct the original medical image. It tolerates numbers of dispersed pieces up to $(n - m)$ to be unavailable or corrupted. Thus it provides survivability. Comparing with the conventional full backup, Rabin's IDA is also more efficient in the total size of storages require to store all the pieces as discussed[6]. Rabin's IDA also provides confidentiality since dispersal and reconstruction process requires a secret key. More confidentiality could be added by putting salt into a message before it is encoded through Rabin's IDA scheme [6].

A comprehensive framework data dispersal solution is introduced for securing query processing on relational data in a cloud system [6]. The solution offered is based on Rabin's IDA by adding such salt for improving encoding security. In addition, a B+-Tree binary index for giving an efficient relational database operation such as select, insert, update, and delete.

Another current works on information dispersal is proposed by [7] for providing a cloud data storage system. Relevant issues on existing information dispersal strategies such as Shamir's SSA (Secret Sharing Algorithm), Rabin's IDA, secret sharing made short, etc. are discussed in this work.

Furthermore, a Jigsaw secure distributed file system introduced to securely store and retrieve files on large scale networks [8]. It addresses confidentiality, integrity, and availability of a stored data by applying recursive Rabin's IDA and layered encryption. Each pieces of dispersed part of a file is sent into a distributed storage, and then these pieces are dispersed again with encoding based on a hashed-key chain algorithm derived from previous encoding key. This solution also provides a reasonable level of plausible deniability to provide users privacy and anonymity.

None of related works found that discuss details of technical implementation on applying secret sharing schemes specific to digital medical image storage system to answer issues on providing survivability and confidentiality. This work provides a comprehensive discussion on doing it.

III. Proposed Method

This work provides an end to end solution, named as Medical Image Dispersal PACS (MIaDPACS), for providing survivability and confidentiality based on secret sharing schemes that are Rabin's IDA and Shamir's SSA, for digital medical image storage system. MIaDPACS could be integrated to the existing medical storage system following PACS standard.

Images produced by medical imaging modality have to be sent to MIaDPACS instead of directly storing it into PACS. MIaDPACS then produces several pieces of the dispersed images and distribute each of them into different PACS servers. MIaDPACS is also installed in workstation such as diagnostic and clinical workstation, to retrieve pieces of the dispersed medical image from the distributed PACS servers and reconstruct back the original image for those pieces. The proposed architecture design of digital medical image dispersal infrastructure is described in Figure 1.

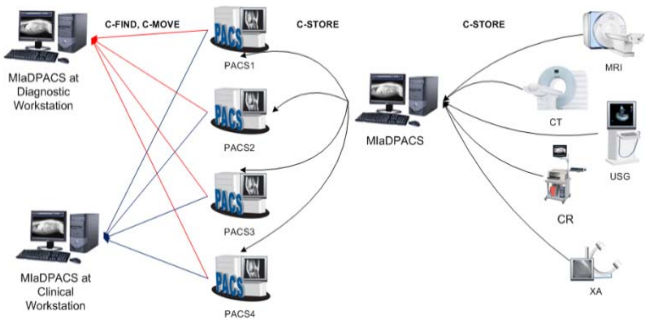


Fig. 1. Proposed architecture design to integrate MIaDPACS with PACS servers

Digital medical image follows DICOM (Digital Imaging and Communication in Medicine) standard. DICOM provides standard for both medical image file system as well as network protocol communication on transferring it from one entity to another entity. A DICOM image comprises of pixels for representing the image digitally and metafile information providing information about patient, modality, pixel interpretation, general information about the image, etc. [8].

A. Key Generation

Secret key for dispersing DICOM image is generated randomly with following rules:

i. Length equals to number of targeted PACS server to distribute the dispersed image.
ii. Each byte of the secret key has to be unique from each other.

This generated secret key is then used for generated a Vandermonde matrix secret C that is used in Rabin's IDA computation. The matrix that has dimension number of rows n equals to number of targeted PACS server to disperse the image, and number of column m equals to number of PACS servers threshold for image reconstruction. Taking an example, $n = 4$, $m = 3$, random generated secret key has to be 4-byte length, for example: {0x03, 0x0F, 0x06, 0x02}. These secret key generate the following Vandermonde matrix secret:

$$C = \begin{bmatrix} 3^0 & 3^1 & 3^2 \\ F^0 & F^1 & F^2 \\ 6^0 & 6^1 & 6^2 \\ 2^0 & 2^1 & 2^2 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 5 \\ 1 & F & 55 \\ 1 & 6 & 14 \\ 1 & 2 & 4 \end{bmatrix}$$

All the computation used in MIaDPACS based on Galois Field GF(2^8).

The secret key {0×03, 0×0F, 0×06, 0×02} is then split into several shared secret keys by using Shamir's SSA. The generated shared secret keys are to be stored later on each of generated new dispersed DICOM image files, meanwhile the secret key is destroyed and to be kept secret from anyone.

B. Image Dispersal

Image dispersal process begins with extracting pixels and header from a DICOM file to be dispersed. Sequences of pixels are grouped into a block before it continues to encoding process. Number of pixel in one block equals to m number of threshold dispersed image to reconstruct back the original image. Each pixel comprises of one or more band, for example monochrome has one band, and RGB color has three bands (red, green, and blue), and allocated size of each band in DICOM medical image is always one byte. Thus GF(2^8) is suitable in MIaDPACS.

A block retrieved from the original image is transformed into matrix D and to be encoded later by using Rabin's IDA. First pixel in the block becomes the first row in matrix D, the second pixel becomes the second row, and the m^{th} pixel becomes the m^{th} rows. Meanwhile, the first band of the first pixel becomes element $e_{1,1}$ in the matrix D, the second band of first pixel become element $e_{1,2}$ and the x^{th} band of first pixel become element $e_{1,x}$. Taking an example of sequence of first 5 pixels of a RGB DICOM image = {211DF1, A7FF17, 130703, 10100F, FF2323}. Suppose that $m = 3$, then the first 3 pixels {211DF1, A7FF17, 130703}, is considered as one block. It could be transformed into matrix $D = \begin{bmatrix} 21 & 1D & F1 \\ A7 & FF & 17 \\ 13 & 07 & 03 \end{bmatrix}$.

MIaDPACS enhances the Rabin's IDA by adding salt into the image before the encoding process to increase data confidentiality. Salt factor is generated by using PRNG (Pseudorandom Number Generator) function with seed equals to Σa_i of the Vandermonde matrix secret.

Given Vandermonde matrix where the second column of these matrix, $C_{:,2}$, is considered as secret key that is summed in GF form then finally create secret seed ss value. Before encode and disperse the dispersed files into distributed servers, each row of D is sum up with different salt factor in every block. The reason of salt factor on each row is that every row is referred to 1 pixel and different band number per pixel. Thus, salt factor randomize each pixel of medical image. Therefore, it removes any characteristics of corresponding original medical image in every dispersed files generated. Equation (1), (2), and (3) describes formula to generate and use salt in MIaDPACS.

$$ss = \text{GF}\ (C_{1,2} + C_{2,2} + \ldots + C_{n,2}) \quad (1)$$

$$\text{fs}(ss);\ \text{fs} = \text{PRNG},\ ss = \text{secret seed} \quad (2)$$

$$\text{D} = \begin{bmatrix} a & b & c \\ 1 & 2 & 3 \\ \ldots & \ldots & \ldots \\ x & y & z \end{bmatrix} + \begin{bmatrix} fs.nextRandom \\ fs.nextRandom \\ \ldots \\ fs.nextRandom \end{bmatrix} \quad (3)$$

The encoding process is looped from the first block subsequently until the last block of pixel. Padding could be added for the last block if number of remaining pixels is less than m. For each of block encoding, it is produced a scrambled matrix with dimension rows equals to n number of PACS servers and

column equals to number of band. Each row at the result matrix is one pixel to be dispersed into their respective new DICOM file.

Next step of dispersal process is to copy all the DICOM tags, except 0×7FE0 (pixel data), from original DICOM file into each of new DICOM files to be dispersed. Following private tags need to be added as well into each of new DICOM file:

i. (0×0029, 0×0010) Creator, with value equals to MIaDPACS, to indicate that these new dispersed DICOM file was created by MIaDPACS application.

ii. (0×0029, 0×1000)Dispersal index, with value equals to image dispersal index, to indicate index for each of new dispersed DICOM file. The index value is in the range of $0 \leq i < n$ number of targeted PACS server.

iii. (0×0029, 0×1001)Shared secret key value, with each of shared secret keys value stored in this location.

iv. (0×0029, 0×1010) Original width, with value equals to width of DICOM original image.

v. (0×0029, 0×1011) Original height, with value equals to height of DICOM original image.

The last step of dispersal process is to store the new generated dispersed DICOM file to their respective distributed PACS servers.

C. Image Reconstruction

Image reconstruction is a reversed of image dispersal process. It tolerates up to (n–m) absent of dispersed files. Image reconstruction starts from retrieving numbers of dispersed files stored on the distributed servers as much as the defined threshold during dispersal. The retrieved dispersed files are then decoded by using Rabin's IDA. The shared secret keys are extracted from the dispersed images to reconstruct the original secret key through Shamir's SSA.

Pixel decoding is also done as reversed of dispersal process. $m$number of available dispersed files are chosen for image reconstruction, For each of i^{th} pixel from each of m dispersed file, a matrix E is reconstructed. A pixel from the first dispersed image become the first row at the matrix E, a pixel from the second dispersed image become the second row at the matrix E, and a pixel from the m^{th} dispersed image become the m^{th} row at the matrix E. Similar to dispersal process, bands for each pixel are also distributed into their respective matrix column. Inverse of sub-matrix secret C^{*-1} as explained by [6] is then multiplied with matrix E to re-compute matrix D following Rabin's IDA. Matrix D has dimension row equals to m and column equals to number of band for pixels in the original DICOM images, thus m pixels of original DICOM image is produced for each of loop. Before writing it to construct the original image, this matrix D needs to be subtracted with its respective salt factor.

IV. Result and Discussion

Survivability and confidentiality of data can be also implemented by introducing redundancy backup with cryptographic mechanism namely encryption. However this type of implementation will cost extra storage as computation time. For example if a system have b backup systems, the total space for storing medical images with size F pixels is given by Equation (4) and image need to be encrypted before dispersing to the backup storage.

$$[b \times F] \tag{4}$$

Meanwhile, through MIaDPACS with $n = b$ = number of distributed servers, m = threshold number where $m < b$, the total space for storing medical images with size F pixels is given by Equation (5).

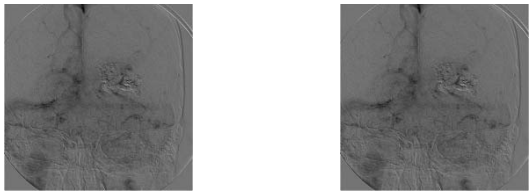
$$\left[\frac{n}{m} \times F\right] \equiv \left[\frac{b}{m} \times F\right] \tag{5}$$

Since the image has already been encoded through Rabin's IDA, explicit encryption is not necessary.

To analyze the implementationspecifically to address the issue on survivability and confidentiality, the experiments were conducted by dispersing and reconstructing DICOM images with $n = 4$ and $m = 3$.

A. Survivability

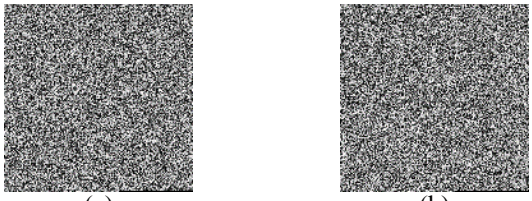
Fig. 2. shows original and its reconstructed image as result of survivability testing. It proved that pixels on the reconstruct image are exactly the same as the ones at the original image. PSNR between them is infinity that is they are exactly the same.



(a) (b)

Fig. 2. Image reconstruction (a) original, (b) reconstructed

Each pieces of the dispersed images produced during dispersal of original images are shown in Fig. 3.



(a) (b)

(c) (d)

Fig. 3. Image dispersal (a) dispersed file 1, (b) dispersed file 2, (c) dispersed file 3, (d) dispersed file 4

All the pieces in Fig. 3. looks scramble and different from each other.TABLE Ishows that PSNR between them are a quite small which implies that their differences are significant.

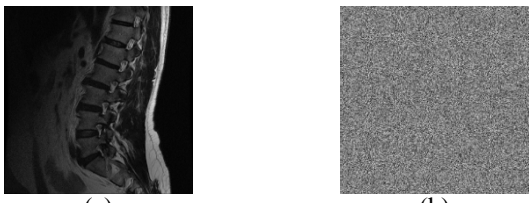
TABLE I. PSNR (DB) OF THE DISPERSED FILES

	A	B	C	D
A	-	7.7	7.74	7.77
B	7.77	-	7.77	7.82
C	7.74	7.77	-	7.81
D	7.77	7.82	7.81	-

B. Confidentiality

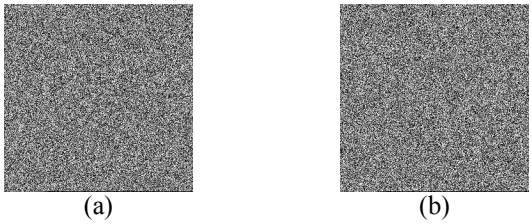
Confidentiality test proves that the system could protect original image from disclosure, if wrong IDA secret key is used during reconstruction. This prevents unauthorized entities to get access into medical image.

The test was simulated by dispersing a medical image into four dispersed images with a defined threshold for reconstruction equals to three. Suppose that a hacker steals two of those dispersed images successfully. He/she also has knowledge to interpret the stolen dispersed images perfectly, and could guess the pixels value to construct the third dispersed image correctly. However he/she is not able to get the correct shared secret key for that third dispersed image.



(a) (b)

Fig. 4. Image reconstruction and wrong shared secret key (a) original, (b) reconstructed



(a) (b)

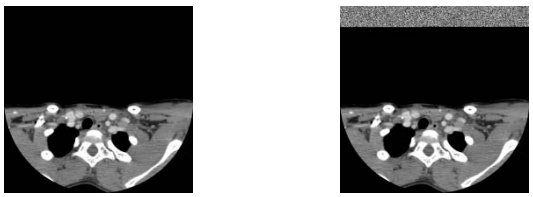
(c)

Fig. 5. Image dispersal (a) dispersed file 1, (b) dispersed file 2, (c) dispersed file 3 with wrong shared secret key

Fig. 4. shows that medical image reconstruction failed to produce the original image, with wrong shared secret key in one of the dispersed file used for image reconstruction as shown byFig. 5.

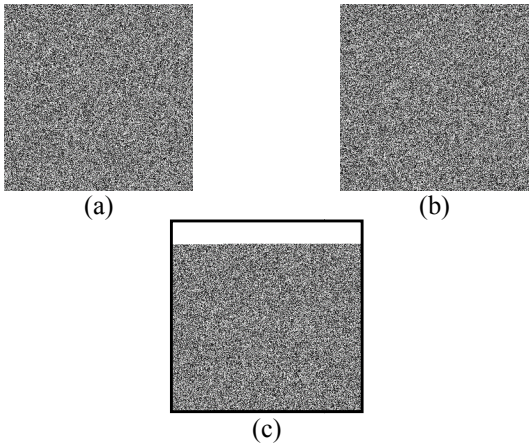
C. Integrity

To test for integrity, pixels of the image were corrupted in one of dispersed files before reconstructing the medical image. Result shows that the reconstructed image is corrupted as much as $t\times$ number of corrupted bytes, where t equals to threshold number. However, if the corruption is in the header of image, then reconstruction of the original imagefailed. This happens since DICOM header contains medical image information to interpret each of byte in medical image. Fig. 6. shows an original image and the reconstructed image, meanwhile Fig. 7. shows dispersed files used for the reconstruction, given that one of them is corrupted.



(a) (b)

Fig. 6. Image reconstruction and corrupted dispersed image (a) original, (b) reconstructed



(a) (b)

(c)

Fig. 7. Image dispersal(a) dispersed file 1, (b) dispersed file 2, (c) corrupted file 3

V. CONCLUSION

This study proposes pixel-based dispersal scheme through MIaDPACS implementation to address medical image survivability and confidentiality. Experimented result shows that MIaDPACS could be an alternative solution in storing medical images and keep their survivability and confidentiality. For further study, this study could be possible for leveraging implementation in a cloud computing environment.

ACKNOWLEDGEMENT

This work is supported by Ministry of Education (MOE), Malaysia and UTM under Vote No.(4L108).

REFERENCES

[1] R. Khorasani, P. K. Goel, N. M. Ma'luf, L. A. Fox, S. E. Seltzar and D. W. Bates, "Trends in the Use of Radiology with Inpatients: What has Changed in a Decades?," *American Journal of Radiology,* vol. 170, no. 4, pp. 859-861, April 1998.

[2] L. Watson and T. G. Odle, "Patient Safety and Quality in Medical Imaging: The Radiologic Technologist's Role," ASRT, Albuquerque, 2013.

[3] R. Mohd-Nor, "Medical Imaging Trends and Implementation: Issues and Challenges for Developing Country," *Journal of Health Informatics in Developing Countries,* vol. 5, no. 1, pp. 89-98, 2011.

[4] M. A. Boyle, "Bursting at the Seams: Storage is Growing Problem for EHR Images," *Medical Economics,* 14 September 2011.

[5] L. S. Ribeiro, C. Costa and J. L. Oliveira, "Current Trends in Archiving and Transmission of Medical Images," in *Medical Imaging*, O. F. Erondu, Ed., InTech, 2011, pp. 89-106.

[6] S. Wang, D. Agrawal and A. E. Abbadi, "A Compression Framework for Secure Query Processing on Relational Data in the Cloud," in *Proceeding SDM'11 Proceedings of the 8th VLDB international conference on Secure data management*, Seattle, 2011.

[7] D. Slamanig and C. Hanser, "On Cloud Storage and the Cloud of Clouds Approach," London, 2012.

[8] J. Bian and R. Seker, "The Jigsaw Secure Distributed File System," *Computers & Electrical Engineering,* vol. XXXIX, no. 4, p. 1142–1152, 2013.