Secure Patient Information and Privacy in Medical Imaging

Ming YANG
Math and Computer Science Dept.
Jacksonville State University
Jacksonville, AL 36265, USA

Monica TRIFAS
Math and Computer Science Dept.
Jacksonville State University
Jacksonville, AL 36265, USA

Lei CHEN
Department of Computer Science
Sam Houston State University
Huntsville, TX 77341, USA

Lei SONG
Yahoo! Inc.
Sunnyvale, CA 94089, USA

Dorothy BUENOS-AIRES
Math and Computer Science Dept.
Jacksonville State University
Jacksonville, AL 36265, USA

Jaleesa ELSTON
Math and Computer Science Dept.
Jacksonville State University
Jacksonville, AL 36265, USA

ABSTRACT

In present times, identity protection is becoming increasingly jeopardized. Numerous ways of protecting one's personal, financial or medical information are therefore being utilized by individuals, businesses, and governments. When it comes to protecting patient information in medical images, we have developed an information hiding methodology that includes the RSA encryption algorithm and a Discrete Cosine Transform (DCT) based hiding technique. With this system, any medical image that will be electronically transferred (i.e. emailed, faxed, etc.) will have the patient's information hidden and embedded in the image outside of the Region of Interest (ROI). For example, an X-ray of a skull that is emailed will not have the patient information displayed during transmission, but will be readily available once it reaches its destination. This system is also unique in the fact that when a medical image is electronically delivered, the patient information and the picture are transferred in the same file, whereas now an image and the corresponding patient information are transmitted in two different files.

Keywords: Encryption, Privacy, HIPAA, Information Hiding, Medical Images.

1. INTRODUCTION

The Health Insurance Portability and Accountability Act (HIPAA) [8] requires that medical providers and insurance companies implement procedures and policies to protect patient's medical information. Areas to be specifically addressed include ensuring that confidential data is secured during electronic transmission, and that access is limited only to authorized personnel. In this project, our goal is to design a security scheme to protect patient information while making the information readily accessible when necessary. The design of our project was motivated by the following points:
(1) Nowadays, the patient information is usually printed in the corner of the medical images for viewing. Thus, it is easily accessible to everybody.
(2) The patient information prominently displayed may be intercepted by a third party during electronic transmission. Sometimes, this could cause a big lawsuit.
(3) For scenarios such as medical imaging research, the patient information should not be accessible either.
(4) For diagnosis purposes, the patient information needs to be readily accessible to the doctors.

One vital aspect of this project was research on the DICOM (Digital Imaging and Communication in Medicine) standard which specifies the format that all digital medical images should use in order to be compatible. [2] It was discovered that while the DICOM format does call for the inclusions of some personal information (patient's birth date, sex, and patient ID along with information about where the image was taken), the patient's name was not specifically included in the format. Since having the patient's name directly associated with the image is important from the patient's point of view, and keeping that information confidential is important for compliance with HIPPA, this project was implemented to create that link.

2. SYSTEM OVERVIEW

The basic idea of the methodology was utilizing a high bitrate information hiding technique to embed patient information (already encoded and encrypted) within the actual medical images. With this approach, the patient information was secured and concealed during electronic transmission. This methodology is robust to attacks such as compression, cropping, intrusion, cryptanalysis, etc.

In the system, patient information is not automatically visibly displayed in the corner of the medical image. Instead, the patient information is first encoded using ASCII character-encoding scheme and encrypted into a non-recognizable format using the RSA encryption algorithm.

Next the patient information is further secured by embedding it within a section of the image that is outside the ROI [7]. This ensures that the encoded and encrypted information is embedded in a location that will not affect the image quality and further diagnosis. The area outside the ROI is located by using image segmentation techniques.

After the area outside of the ROI is located, the information is embedded using a DCT domain methodology. This information hiding algorithm is very robust so that further attacks including cropping, noise, lossy compression, etc., will not remove the embedded information. This methodology effectively and securely protects patient information in scenarios such as electronic transmission and medical imaging research.

The image can be viewed in one of two forms. If the viewer does not have the authority to access the patient's personal information, for example a medical or computer researcher (or network hacker for that matter), the image is viewed with no data displayed in connection to the image. If the viewer has the authority to access the personal information, such as the patient's doctor, the information can be extracted, decrypted, decoded, and displayed upon the image with the entry of the correct encryption/decryption key. The above procedure can also be combined with a fragile watermark to validate data integrity. Our approach is illustrated in Fig.1.

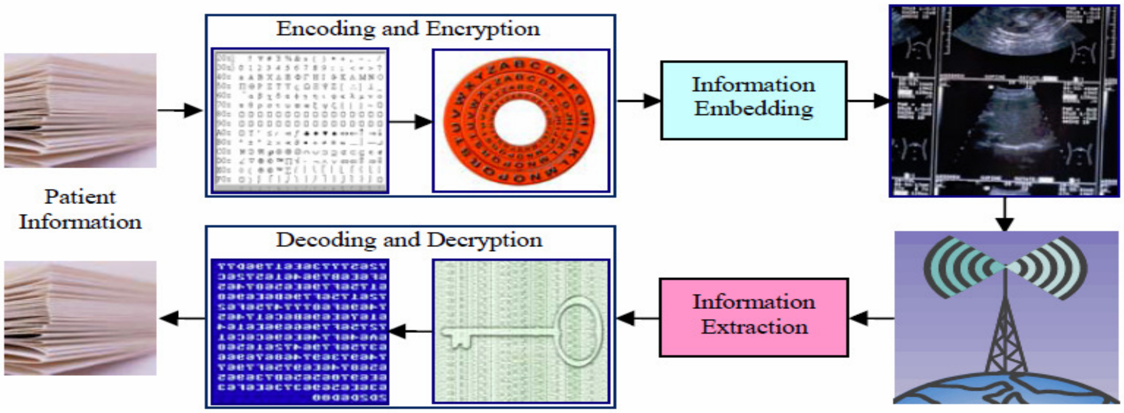

Fig. 1 Flow Chart of Proposed Methodology

3. IMAGE SEGMENTATION

Image segmentation is the process of dividing an image into sections, regions, or parts [6]. This process has numerous applications, such as automated inspection. Gonzalez gives the example that in the automated inspection of electronic assemblies, image segmentation is used to find defects, such as a missing or broken path [6]. In respect to our project, we needed to identify the ROI, or where the actual picture is on the medical image. For example, if we have an X-ray of an elbow, our region of interest would be the elbow, NOT all of the black space surrounding it. Clearly, image segmentation is a vital part of this project. This process identifies the region of interest of an image, and draws a boundary inside which the patient information cannot be placed. This was first achieved using MATLAB's built-in contour functions, and later done using a program in Java. Earlier quests to identify the ROI used edge detection methods [1].

Edge detection is a process of "detecting meaningful discontinuities in gray level," or, in other words, finding discontinuities in pixels of an image [6]. The implementation of edge detection methods was also done in MATLAB. The edge detection methods were not used in the final implementations because the methods revealed much more detail on the image than what was necessary.

4. INFORMATION HIDING

Throughout the ages various methods have been devised to conceal information in transit. Tactics in previous times ranged from tattooing the message on a shaved head then waiting for the hair to re-grow before sending the message [10] to placing microfiche with the information under the postage stamp on a letter. With the creation of the Internet and other electronic data transmission mediums, steganography or the art of hiding information, has become even more important and commonplace.

While past motivations for hidden messages may have been primarily military, today there are many civilian justifications for employing the available technology such as embedding watermarks or fingerprints for copyright protection, safeguarding product information and trade secrets, electronic transaction protection during e-commerce, confidential communications, operation tracking, signature fields, and authentication of documents. [5]

Information hiding is one branch of a well developed information security plan. The other branch is cryptography. [5] The use of the two technologies together provides a double layer of information protection. Ge, Jiao, Tian, and Wang [5] proposed a four phase process model of information hiding which includes pretreatment, embedding, transmission, and extraction phases. Encryption is included as part of the pretreatment phase.

Information can be hidden with success in text, image, audio/image, and protocol file formats. [10] There are two main groups of information hiding techniques: techniques in the spatial domain and techniques in the transform domain. Spatial domain techniques generally involve manipulation of pixel intensity. Lossless image formats are most suited for spatial domain techniques [10]. The most well-known technique of information hiding in the image domain is Least Significant Bit (LSB) algorithm. In this algorithm, the least significant bit, which will affect pixel color the least, is examined and either changed or not changed so that it matches the bit that is to be embedded. The algorithm can either adjust every least significant bit or, for greater security, adjust only every n^{th} bit with n being known only to the message recipient thus providing and additional level of security. [10]

Hiding information in the transform domain is a multi step process. First an appropriate transform must be applied. The most popular are the Discrete Fourier Transform (DFT) and the DCT. The transform changes the signal to a frequency representation from a spatial representation. This has the effect of spreading the pixel values over a part of the image, usually an 8x8 block. The next step is to modify the transform coefficients to embed a desired bit. This has the effect on the image of modifying the brightness of the image. [10] The modified DCT coefficient blocks are then transformed back to the spatial domain to obtain the stego-image. [14]

5. EXPERIMENTAL RESULTS

The patient information securing system was implemented using MATLAB, a high-level language and interactive numerical computing environment. MATLAB has a wide variety of image processing capabilities and can process DICOM, BMP, JPG as well as other image formats.

Patient data, delimited by commas and spaces, was read in from a test file for the first version of the system. In a later version the data was entered interactively by the user through a Graphical User Interface (GUI), written in Java and called by MATLAB.

The procedure to convert the text data to ASCII format in MATLAB resulted in seven bit character strings instead of the expected eight bits. These then needed to be broken apart and individually converted back into integer data types in order to perform the necessary mathematical operations for encryption and embedding.

The RSA encryption method was used to encrypt the patient's information. This particular method was chosen due to the simplicity of its algorithm. RSA is an asymmetric or public key algorithm, meaning it has both a public key and a private key. [11] The advantage of an asymmetric encryption is that it is more secure. The main disadvantage is that it has a slower run time and does not handle large amounts of data as well as symmetric algorithms. As the amount of data being encrypted would not be particularly large, this was determined to not be an important consideration.

Several different algorithms were experimented with to embed the data within the image. The first implementation was written with no attempt to "hide" the data. This procedure added an additional padding to the image that included the data. This algorithm was studied because of its simplicity and the fact that it does not alter the original image data, but rather adds the information in an additional area. A drawback of this method is that the viewer clearly sees where the information is located and if motivated would only need to break the encryption to access the data. This method would also be very vulnerable to attacks by cropping. An example of the results of this implementation is seen in figure 2.

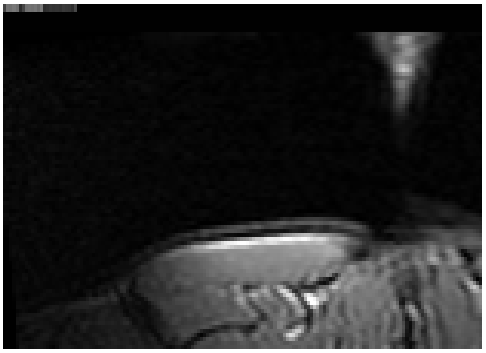

Figure 2- A sample image with patient information added in the top left corner.

The second method used was embedding the data directly in the image using the LSB technique. With this method the addition of the data itself could no longer be detected with the human eye. However, a flag in the first column was used to indicate the number of additional columns in that row which held data. This flag was visible upon close inspection. It was determined that this technique could be too vulnerable to attack and therefore attention was shifted to using a technique in the transform domain. Figure 3 shows an example of an image using LSB algorithm.

The technique that was finally selected was one discussed by Yang and Bourbakis [14]. They proposed using a DCT on 4x4 blocks of pixels and then modifying the eight lowest frequency coefficients, which represent the luminance of the background of the image, to embed one bit of information.

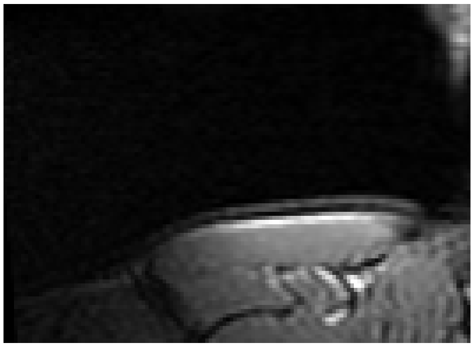

Figure 3- A sample image with patient information embedded using the least significant bit algorithm.

In the embedding procedure a simple image segmentation algorithm was first employed to identify the ROI. Figure 4 below is an x-ray of a skull that has been analyzed by image segmentation and has only the contour lines showing. This was done using MATLAB's contour function.

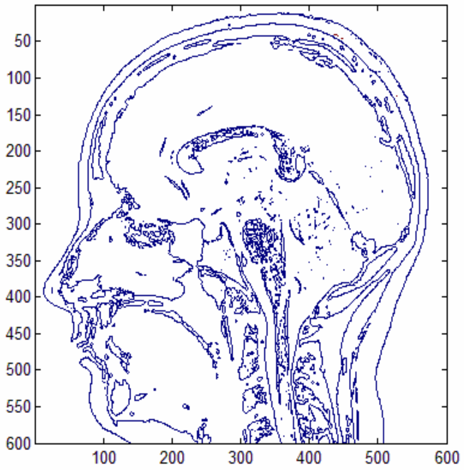

Figure 4- Image with Region of Interest boundary.

After segmentation, the encoded/encrypted patient information is embedded in the part of the background area that was determined to be outside of the ROI to preserve the quality of the host image [7]. A high bitrate transform domain information hiding algorithm was designed to enable data embedding. In the proposed algorithm, 1 bit is hidden within each 4x4 DCT coefficient block by means of vector quantization. Low-frequency coefficients are chosen for information hiding due to their relatively large amplitudes and the corresponding small step sizes in the quantization matrix. The proposed algorithm is very robust to lossy compression, according to the experimental results. After the data is embedded the transform was reapplied to convert the data back to the image domain and then the image was ready for transmission.

At the receiving end, the patient information can be extracted, decrypted and decoded upon entry of the correct passkey. The extraction, decryption, and decoding of the data are simply the reverse of the above procedures, with the addition of the display of the patient data below the image in the receiver's GUI. A copy of the program was placed on a remote computer and the full procedure was tested. The data file was encrypted and embedded into the image and transmitted by email. The image was retrieved at the second computer, extracted, and decoded.

6. CONCLUSIONS AND FUTURE WORK

While protection of a patient's personal data is very crucial it is also important that the patient be reassured that the data being viewed is that of themselves. This system has been designed to integrate these two goals. Additional implementations could include some or all of the following:

- Auto display of original image in GUI at embedding stage;
- Display of image with no data option in end display of image (no key needed);
- Fragile Watermark – to ensure that data has not been compromised.

7. ACKNOWLEDGMENTS

This project is partially funded by the Computing Research Association Committee on the Status of Women in Computing Research (CRA-W) and the National Science Foundation (NSF). Opinions expressed are those of the authors and are not necessarily of the Foundation.

This project was designed to facilitate participation for two women students in the undergraduate student research program, Collaborative Research Experience for Undergraduates (CREU) organized by the Computing Research Association's Committee on the Status of Women in Computing Research (CRA-W) and the Coalition to Diversify Computing (CDC). The goal of this program is to provide positive research experience and to increase the numbers of women and minorities who continue on to graduate school in computer science, engineering, and allied disciplines.

8. REFERENCES

1. E. R Davies, **Machine Vision: Theories, Algorithms, Practicalities**, Amsterdam, Boston: Elsevier, 2005.

2. DICOM standard website - http://medical.nema.org/

3. C. Fei, D. Kundur, and R. Kwong, "The choice of watermark domain in the presence of compression", **Proceeding of International Conference on Information Technology: Coding and Computing**, Las Vegas, NV, 2001, pp. 79-84.

4. G. A. Francia, III and R. R. Francia, "An Empirical Study on the Performance of Java/.Net Cryptographic APIs", **Information Security Journal: A Global Perspective**, Vol 16:No. 6, November, 2007, pp. 344-354.

5. X. Ge, R. Jiao, H. Tian, and J. Wang, "Research on Information Hiding," **US-China Education Review**, USA, Volume 5, No.3, Serial No.18, May 2006, pp.77-81.

6. R. C. Gonzalez, and R. E. Woods. Digital Image Processing. Upper Saddle River: Prentice-Hall, 2002.

7. X. Guo, and T.G. Zhuang, "A Region-Based Lossless Watermarking Scheme for Enhancing Security of Medical Data," (abstract) **Journal of digital imaging: the official journal of the Society for Computer Applications in Radiology,** Jul 10, 2007 [Epub ahead of print]. http://www.ncbi.nlm.nih.gov/pubmed/17619929.

8. "Health Insurance Portability and Accountability Act (HIPAA) and Its Impact on IT Security," **Regulatory Compliance Series 3 of 6**, **Apani Networks White Paper Compliance Series**. May 12, 2005. http://www.apani.com.

9. D. Kundur, "Implications for high capacity data hiding in the presence of lossy compression", **Proceeding of International Conference on Information Technology: Coding and Computing**, March, 2000, pp. 16-21.

10. T. Morkel, J.H.P. Eloff, and M.S. Olivier, "An Overview of Image Steganography," **Proceedings of the Fifth Annual Information Security South Africa Conference. *(ISSA2005)***, Sandton, South Africa, June/July 2005. (Published electronically).

11. W. Stallings, **Cryptography and Network Security: Principles and Practice,** Upper Saddle River: Prentice Hall, 1999.

12. A.B. Watson, "DCT quantization matrices visually optimized for individual images", **Human Vision, Visual Processing, and Digital Display IV**, Bernice E. Rogowitz, Editor, Proc. SPIE 1913-14.

13. M. Yang, S. Li, and N. Bourbakis, "Data-Image-Video Encryption", **IEEE Potentials Magazine**, Aug/Sept. 2004, pp.28-34.

14. M. Yang, and N. Bourbakis, "A High Bitrate Multimedia Information Hiding Algorithm in DCT Domain", **Proceeding of World Conference of Integrated Design and Process Technology (IDPT 2005)**, Beijing, China, June 13th-17th, 2005.

15. M. Yang, and N. Bourbakis, "A High Bitrate Information Hiding Algorithm for Digital Video Content under H.264/AVC Compression", **Proceedings of IEEE International Midwest Symposium on Circuits and Systems 2005 (MWSCAS 2005)**, Cincinnati, Ohio, USA, August 7th-10th, 2005.

16. M. Yang, and N. Bourbakis, "An Overview of Lossless Image Compression Techniques", **Proceedings of IEEE International Midwest Symposium on Circuits and Systems 2005 (MWSCAS 2005)**, Cincinnati, Ohio, USA, August 7th-10th, 2005.

17. M. Yang, M.Trifas, N. Bourbakis, and C. Cushing, "A Robust Information Hiding Methodology In Wavelet Domain", **Proceedings of 12th International Conference on Signal and Image Processing**, Honolulu, Hawaii , August 2007.

18. M. Yang, M.Trifas, C. Truitt, and G. Xiong, "Wavelet Domain Video Information Embedding", **The 12th World Multi-Conference on Systemics, Cybernetics and Informatics**, Orlando, Florida, June 29th - July 2nd, 2008.