Conceptual Framework of A New Secure Storage System for Medical Data Archiving

Sos S. Agaian, Okan Caglayan, Multimedia and Mobile Signal Processing Laboratory, University of Texas at San Antonio, Texas, USA

Abstract

The storage, retrieval, and manipulation of digital data, such as digital medical images, signals, and even documents (medical reports), and the analysis of the information held in these data are important requirements for the current and the next-generation medical archiving systems. The medical information is the most sensitive data, which requires strong security measures. In the medical archiving systems, the most of the security depends on the mandatory access controls, and the encryption of the vital data. The problems with these techniques arise when an unauthorized client has the knowledge regarding the existence of the secured data in the system; this can make the system vulnerable to alterations, extraction or destroying of the secured data.

This paper presents a new conceptual framework of secure storage system for medical data (digital medical images, documents, and signals, etc.) archiving through the use of steganographic and cryptographic techniques. The main objectives are: 1) To present a new conceptual framework of a multilayer database system for medical data archiving, 2) To provide highest level of security and privacy of the vital information in the database against unauthorized alteration or destruction by any personnel; and 3) To combine steganography and cryptography in the infrastructure of the database system to further increase the security of the crucial data.

Introduction

The significance and urgency of the security and privacy problems faced by medical archiving systems has been and endures as a prime concern due to the technological advances in the computer infrastructure. In the past, medical information had been physically stored in hospitals, laboratories, and doctors' office. Access to this sensitive data was limited, and it was protected by its physical isolation and ignorance of its existence. With the digitization of medical data, this information is becoming accessible through distributed systems, including the Internet, mobile communication. This consequently has increased the numbers of people that can potentially access medical information by orders of magnitude, often providing more efficient transfer of medical records and related information. Centralizing and sharing these electronically-managed data, which basically provides accessibility to the patients' personal information, eventually can cause the misuse of these records. Misuse of a person's medical and genetic data could potentially impact his/her ability negatively to be hired, and limit the career path and insurability [15]. Therefore, the medical information privacy assurance is a sensitive topic which ultimately requires strong security measures.

There are several ongoing research projects focusing on specialized medical archiving systems striving to maintain the security of stored vital information [1,2,3,6-9]. In [7], Bowen et. al. describes the design of access control methods, which mainly based on authentication and auditing of the personnel, for protecting the confidentiality of patient information. Fernandez et. al in [15] argues the general security models based on the mandatory access control. The authors briefly describe some of the requirements for the security models. In [5], Ateniese et. al. discusses security and confidentiality issues in the medical database. The author's emphasis on the database security is based on the crytographical approach. As we noted in [11], cryptography (encryption) and mandatory access control policies in medical database systems provide some security, yet the problems arise with these approaches when an unauthorized client has knowledge regarding the existence of the secured vital data in a system, these techniques cannot protect the vital information in its entirety.

In recent years, steganography (data hiding) has become a very popular means for securing data. Steganography is the secure communication of information by embedding a message into a 'cover' digital media in a manner that is undetectable by external observers [5]. The message may at any time be retrieved from the transmitted/stored digital data. The difference between steganography and cryptography is, though the intercepted message may not be decoded without knowledge of the necessary key, there is an obvious existence of some data transmission. On the other hand, in steganography, though the message itself may not be difficult to decode, the manner in which the data was inserted makes the message invisible to any outside source [11].

The focus of this paper is as follows:

1. To develop a new conceptual framework of a multilayer database system for medical industry.
2. To protect the confidentiality of the sensitive information.
3. To provide accessibility to the system based on the classification of the personnel.
4. To ensure the integrity of the data.

In this paper, we have presented an efficient system level approach to protect the sensitive medical information, which is stored electronically in the medical data archiving systems. The rest of the paper is organized as the following. In section II, we briefly discussed the medical database architecture of the proposed system; section III discussed the entire system; we concluded this paper in section IV by providing some open problems.

Background

The medical data is one of the most sensitive types of information, and requires strong security on the following three aspects [9].

- Confidentiality
- Integrity
- Availability

For example; an incorrect change in a medical record may result in a wrong prescription with damage to the patient, or leakage of information about a psychiatric treatment could ruin a career [15]. In the proposed system, the main task has been to provide the highest level of security by simply protecting the vital data's confidentiality, because it serves as a foundation for all the three aspects, as it is noted above in [15].

Unauthorized accesses to data and records in the military and commercial industries are likely to be used for criminal purposes, such as the sale of military secrets or fraud, respectively [6]. With medical information such breaches and uses can be more dangerous, and the damages are less evident. Information systems administrators in both military and financial institutions are given strong mandates to curb criminal use of the housed data; breaches are often followed by investigation to assign responsibility, and by disciplinary action [6]. The loss of credibility following a finding of gross negligence can be as damaging to the institution as the event itself.

As stated in [6], a different picture prevails in the medical field. Unlike commercial institutions, e.g., banks, health care institutions have avoided public degradation after breaches by blaming the personnel who violate the sacred principles of ethical behavior. The public and even many health care professionals, perhaps out of a lack of understanding of security principles and practices, assume that the high ethical standards expected of health care personnel are enough of a deterrent to the misuse of information in all but exceptional cases. This view is contradicted by the fact that medical records are routinely available to non-medical personnel for essential business functions such as claim payment processing. Moreover, medical information has concrete monetary value to other stakeholders than the health care provider. Until recently the prevalent view in the health care industry was that investing in security would hinder efficiency, decrease performance, and increase costs. It has been argued that we should learn as a society to accept some measure of risk to the security of our medical records as a better alternative to pricing health care beyond the reach of many. Embracing such stance shifts the cost of damages from the institution to the individuals who become victims of such breaches [6].

Medical Database Model Requirements and Specifications

While several authorization models have been proposed for general use, few models are specifically intended to represent access constraints in medical environments. One of the earliest discussions of unique security needs for medical systems is a paper by T.C. Ting discussing the requirements of mental health security [20]. J. Biskup did some significant work on privacy aspects of medical systems [21]. G. Pangalos developed several design models for medical database [22]. R. Anderson did a systematic work of identifying policies for general clinical records [23]. There are also several studies by medical informatics researches on the issues and requirements of patient records, including some actual implementations [24,25,26]. Their studies are valuable to understand experience in implementing and using medical information but they do not attempt to develop new security approaches [15].

The efficient implementation of the security specifications that apply to the design of strict secure database model interpreted as the following [15]:

- Necessary to apply "need to know" policy, providing only the information to the authorized medical users
- Access for the users of this system should be defined by their roles, e.g., patients can see their records and doctors can modify their patients' information.
- Privacy, which implies a large amount of control for the patients' information
- Closed system design, where the lack of an authorization rule implies no access.

Based on these general specifications, the authors have stated the assumption of the following security requirements:

a. *Attribute and credential-based authorization:* In an environment where not all the users that may need access to a document are known in advance, we need to have authorization models that can consider user attributes and credentials to determine access rights.

b. *Content-dependent authorization:* The granularity of access should be to the record level to separate individual information.

c. *Context-dependent access models:* There are occasions where the standard predefined authorization must be overridden. For example, if a patient is unconscious and needs immediate attention, it is possible that the authorized users of his/her record may not be present and someone must access the record to decide about the treatment.

d. *Delegation of rights:* Any authorization model must contain policies on how the rights of a subject are delegated to other subjects. This is specially important in models where privacy is a major objective.

e. *Administration of security:* Need to have traditional security administrators that define roles, assign users to roles, create groups, and perform similar global functions.

f. *Multimedia objects:* Medical records are a combination of text (medicines, treatments, annotations), audio, and images (X-rays, CAT scans, ultrasound images), as well as other documents related by hypertext links.

g. *Inference control:* Access to some information could allow one to infer other aspects and we need to control at least basic inferential associations.

h. *Explicit audit:* Audit is particularly important when we have context-dependent authorization because of the possible legal implications of overriding or adding authorizations.

It is natural that no single model can satisfy all these requirements. We need several related models at different abstraction levels [15].

Database Design

Relational databases have been used to store a variety of "structured" textual information and numeric values for business applications [17]. However, with the advent of Internet and web, the ever-increasing range of "unstructured" multimedia formats such as images and audio and video clips, require the database to have the ability to load, store, and access all the multimedia data in addition to the traditional structured data [17].

A large database is likely to be multilayer – that is, composed of several different schemas, with physical (source) schemas at the bottom and virtual (view) schemas layered on top [11,28]. We have designed each database, with the combination of the two in mind, (public and classified) in order to store digital imaging, signal studies, and patient records. Digital imaging and signal studies have been referred to a set of X-ray, MRI, and CT images, and ECG signals respectively (shown in figure 4), and the patient records in text format, collected on a given single patient [12]. We have divided the information stored in the proposed system into three categories as digital image, textual, and signal data (shown in figure 1). The textual data that is subject to indexing and querying consists of metadata, patient data, and clinical data, such as patient information, examination technique, diagnosis, treatment, etc. In order to provide the users with the capability to search and retrieve information from the database, we have worked on developing a content-based information retrieval (CBIR) algorithm. It should be noted that the images can be stored in the database as BLOB, binary large object, data type, which can be used to perform analysis, data mining, and content-based information retrieval functions [13, 14]. In the proposed system, we have stored the images (color, and binary) and the textual data in BLOB format, and when an image is loaded into the database, the system generates *64 by 64* thumbnails of the image. These thumbnails are used for display purposes, preview of the original image, however once the original image is requested for analysis, all is required is to click on the given thumbnail to retrieve the original data.

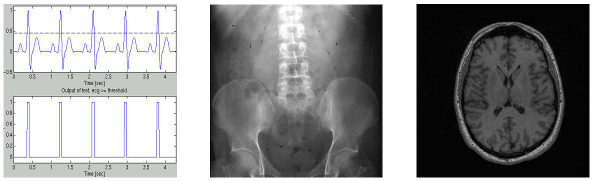
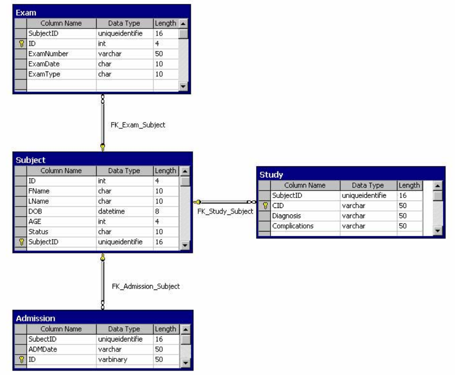


Figure 1: Example of stored imaging, and signal studies [12]



Exam

Column Name	Data Type	Length
SubjectID	uniqueidentifie	16
ID	int	4
ExamNumber	varchar	50
ExamDate	char	10
ExamType	char	10

Subject

Column Name	Data Type	Length
ID	int	4
FName	char	10
LName	char	10
DOB	datetime	8
AGE	int	4
Status	char	10
SubjectID	uniqueidentifie	16

Study

Column Name	Data Type	Length
SubjectID	uniqueidentifie	16
CID	varchar	50
Diagnosis	varchar	50
Complications	varchar	50

Admission

Column Name	Data Type	Length
SubectID	uniqueidentifie	16
ADMDate	varchar	50
ID	varbinary	50

Figure 2: Subset Entity-Relationship Diagram of the proposed system schema

Multimedia and Mobile Signal Processing Laboratory is providing a Dell Poweredge 2800 Server that allows us to implement and assess the proposed system. The database server side is implemented by using Microsoft SQL Server 2000 software, and the client side (front end graphical user interface) is tied in by using MFC application. The system has the capability of querying the public and the classified databases simultaneously, yet only the clients with the appropriate security clearances can access the data. Furthermore, the proposed system allows the user to consolidate an unclassified and classified database.

System Architecture

In the proposed system, the first task is to ensure the prevention of multi-users (clerks, nurses, doctors, etc.) with different access privileges from accessing data classified above their security level through steganographic techniques. The design of the system contains two separate databases, which will be integrated to operate as one database, as public and classified. The medical data depending on the filing classification levels (secret, confidential, top secret, etc.) will be stored accordingly in the proposed system. In the subsequent layers of the database, mainly in the classified layers, the digital media contains vital information that is invisible to users that do not have the security privileges. Steganographic techniques are the essential component of the classified layer, because of two reasons; 1) they provide the multi-user key security concept, and 2) The data hidden inside the cover media is invisible to the outside source in the event of intrusion to the system.

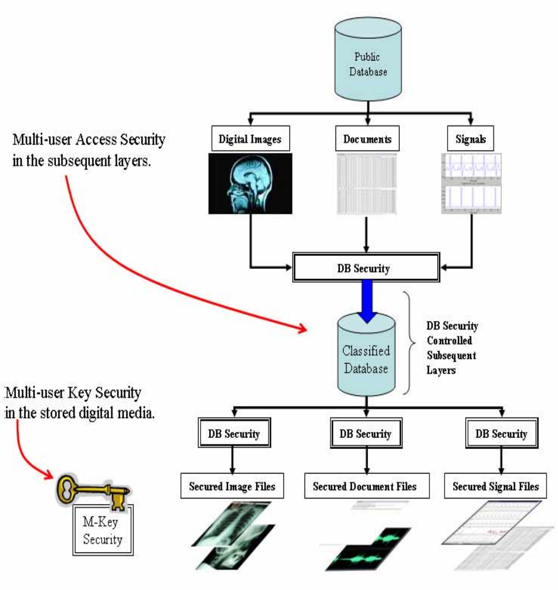
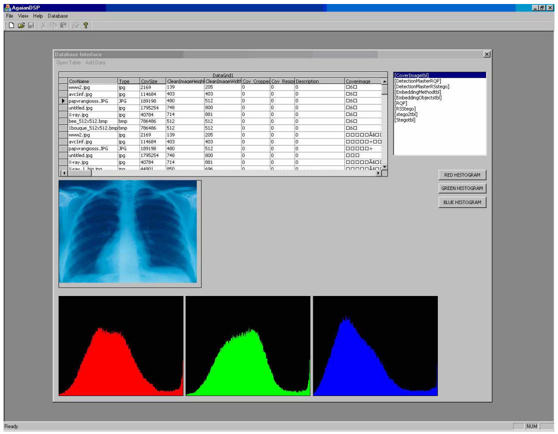


Figure 3: Proposed archiving system

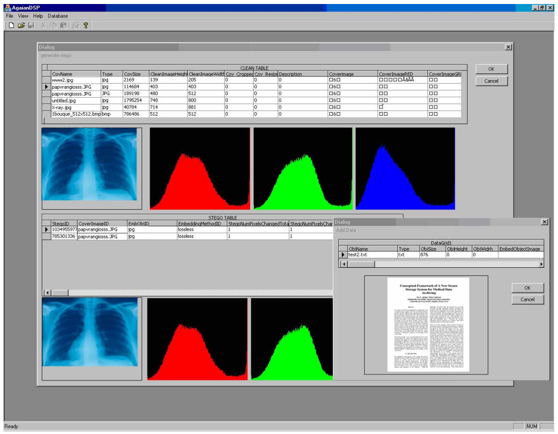


DataGrid1

CovName	Type	CovSize	CleanImageHeigh	CleanImageWidth	Cov_Croppec	Cov_Resize	Description	Coverimage
www2.jpg	jpg	2169	139	205	0	0	0	□6□
avc1inf.jpg	jpg	114684	403	403	0	0	0	□6□
papvrangiosss.JPG	JPG	189198	480	512	0	0	0	□6□
untitled.jpg	jpg	1795254	748	800	0	0	0	□6□
X-ray.jpg	jpg	40784	714	881	0	0	0	□6□
bee_512x512.bmp	bmp	786486	512	512	0	0	0	□6□
1bouque_512x512.bmp	bmp	786486	512	512	0	0	0	□6□
www2.jpg	jpg	2169	139	205	0	0	0	□□□□□Ãã□□
avc1inf.jpg	jpg	114684	403	403	0	0	0	□□□□□÷□□
papvrangiosss.JPG	JPG	189198	480	512	0	0	0	□□□□□÷
untitled.jpg	jpg	1795254	748	800	0	0	0	□□□
X-ray.jpg	jpg	40784	714	881	0	0	0	□□□□□Ãã□□
X-ray_1_big.jpg	jpg	44901	850	696	0	0	0	□□□□□Ãã□□

[CoverImagetbl]
[DetectionMasterRQP]
[DetectionMasterRSstego]
[EmbeddingMethodtbl]
[EmbeddingObjectstbl]
[RQP]
[RSStego]
[stego2tbl]
[Stegotbl]

(a)



CLEAN TABLE

CovName	Type	CovSize	CleanImageHeight	CleanImageWidth	Cov_Cropped	Cov_Resize	Description	CoverImage	CoverImageRED	CoverImageGR
www2.jpg	jpg	2169	139	205	0	0	0	□6□	□□□□□ĀāĀĀ	□□
papvrangiosss.JPG	jpg	114684	403	403	0	0	0	□6□	□□	□□
papvrangiosss.JPG	JPG	189198	480	512	0	0	0	□6□	□□	□□
untitled.jpg	jpg	1795254	748	800	0	0	0	□6□	□□	□□
X-ray.jpg	jpg	40784	714	881	0	0	0	□6□	□	□□
1bouque_512x512.bmp	bmp	786486	512	512	0	0	0	□6□	□□	□□

STEGO TABLE

StegoID	CoverImageID	EmbObjID	EmbeddingMethodID	StegoNumPixelsChangedTota	StegoNumPixelsChar
1034955977	papvrangiosss.JPG	jpg	lossless	1	1
785301336	papvrangiosss.JPG	jpg	lossless	1	1

DataGrid1

ObjName	Type	ObjSize	ObjHeight	ObjWidrh	EmbedObjectImage
test2.txt	txt	876	0	0	

(b)

Figure 4: Graphical user interface view of the proposed system

System Security

In this section, we briefly discuss the new steganographic methods that we have developed in order to utilize them in the security of the vital informations.

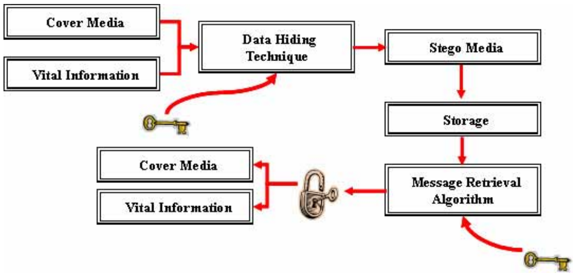


Figure 5: Flow diagram of the steganographic system

Figure 5 has shown the flow diagram to show the structure of the steganographic system. The main difference between the proposed algorithms and the existing ones is that reconstruction of the cover media and vital information is crucial for our purposes.

Furthermore, figure 5 describes the multi-key security within the classified data. The vital data based on the sensitivity level will be accessed only by the personnel that have the privilege of having the right key to be able to retrieve the crucial data.

Storage of the digital medical images (color, and binary, such as X-ray, MRI, CT scan, etc.), documents (patient reports, etc.), signals (ECG), video (surgical educational applications, etc.) in a secure environment is the essential part of the system. We have developed several algorithms in order to apply to the proposed system design. The key properties of some of the algorithms are as the following:

1. *Binary Image or Text Steganography:* Agaian et al. [19] introduced a run length based steganographic technique for binary images, which ultimately secures the vital data by altering the pixels of the embeddable blocks of the cover media depending on the run length characteristics and characteristics values of the block. This technique has also been applied to text format media, such as signatures, documents.
2. *Image Steganography:* In [18], authors have described the capacity of embedding into a given cover media, and proposed a new adaptive technique that is able to overcome embedding capacity limitations, and reduce the revealing artifacts that are customarily introduced when applying other embedding tools.
3. *1-D Signal Steganography:* In [27], we have developed a lossless adaptive digital audio steganography algorithm. This adaptive algorithm featured choosing the best blocks for embedding perceptibly inaudible stego-information. Embedding of stego information was carried out in the transform domain and a pseudo-noise sequence was added to the carrier. This method of embedding did not require the original signal information to reconstruct the secret message. Furthermore, a capacity measure was introduced to select the audio carriers that presented the minimum distortion after undergoing the embedding process.
4. *Video Steganography:* In [29], a novel data hiding algorithm and system design for high quality digital video has been developed. The authors argues that instead of targeting on a single degree of robustness, which results in overestimation and/or underestimation of the noise conditions, they applied multi-level embedding to digital video to achieve more than one level of robustness-capacity tradeoff.

Conclusion

The need for a medical image repository is growing at a rapid pace as more and more healthcare professionals utilize imaging for diagnosis and research. Although it is a complex task to build an enterprise class system that is reliable and robust, but the benefits it would bring to the research community is unimaginable [17].

In this paper we have presented a concept of a new secure multilayer database system for archiving digital medical data (X-ray, MRI, CT scan, medical reports, ECG, surgical educational applications) based on the steganographic and cryptogaphic techniques. The key components of the proposed multilayer system are: the confidential data is only visible to the personell with the granted privilieges (classification levels), data integrity and availability with a secure approach. The main advantage of the steganographic techniques is that the embedded information (data) within the cover image is invisible to an outside source. We use the both types of steganographic techniques lossy and lossless. In addition, the integration of two database systems, simply appearing as one system to an outside source, will ultimately save the consumer time, energy, and the money it costs to operate and maintain two separate systems.

Acknowledgement

This research was partially funded by the Center for Infrastructure Assurance and Security at University of Texas San Antonio, under the contract with the Air Force Research Laboratory (US Air Force), and Air Force Information Warfare Center.

References

[1] Chiang, Y.C., Hsu, T.S., Kuo, S., Wang, D.W., “Preserving Confidentiality When Sharing Medical Data”, Proceedings Asia Pacific Medical Informatics Conference (APAMI-MIC), 2000.

[2] Rogulin, D., Estella F., Hauer, T., McClatchey, R., Solomonides, T., “A Grid Information Infrastructure for Medical Image Analysis”, Proceedings of the Distributed Databases and processing in Medical Image Computing (DiDaMIC) Workshop, MICCAI 2004, Rennes, France, 2004

[3] Lozano, C.C, Kusmanto, D., Chutatape, O., “Web-based Design for Medical Image Database”, Seventh International Conference on Control, Automation, Robotics, and Vision (ICARCV’02), Dec 2002, Singapore

[4] Tsai, C.L, Fan, K.C., Chung, C.D., Chuang, T.C., "Reversible and Lossless Data Hiding with Application in Digital Library", IEEE 38th Annual 2004 International Carnahan Conference on Security Technology, 11-14 Oct. 2004, pp: 226-232

[5] Neil Johnson, Zoran Duric, Sushil Jajodia; "Information Hiding: Steganography and Watermarking – Attacks and Countermeasures", Kluwer Academic Publishers, Second Printing 2001.

[6] Ateniese, G., Curtmola, R., Medeiros, B., Davis, D., "Medical Information Privacy Assurance: Cryptographic and System Aspects", 3rd Conference on Security in Communication Networks 2002 (SCN 2002), Sept. 12-13 2002, Amalfi, Italy

[7] Al-Salqan, Y.Y., "Security and Confidentiality in healthcare informatics", 7th IEEE International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises Proceedings, pp. 371-375, 17-19 June 1998

[8] Bowen, J.W, Klimczak, C., Ruiz, M., Barnes, M., "Design of Access Control Methods for Protecting the Confidentiality of Patient Information in Networked Systems", American Medical Informatics Association (AMIA) Proceedings, pp: 46-50, 1997

[9] Saffron, C., Ring, D., et al., "Protection of Confidentiality in the Computer-Based Patient Records", M.D. Computing, vol 12, No. 3, May/June 1995, pp. 187-192

[10] Ponniah, P., "Database Design and Development", IEEE Press, Wiley-Interscience, 2003

[11] Sos S. Agaian, Okan Caglayan and Natalie Granado, "Secure Multilayer Database System for Digital Image Archiving", IS&T Archiving Conference, Proceedings pp. 165-169, April 26-29, 2005, Washington, DC.

[12] Computer Graphics Group, University of Erlangen, Germany. http://www9.cs.fau.de/Persons/Roettger/library/

[13] Zhu, X., Lee, K., Levin, D.L., Wong, S.T.C., Huang, H.K., Hoo, K.S., Gamsu, G., Webb, W.R., "Temporal Image Database Design for Outcome Analysis of Lung Nodule", Computerized Medical Imaging and Graphics, 20(4):347-356, Aug. 1996

[14] H. Muller, N. Michoux, D. Bandon, A. Geissbuhler, "A Review of content-based image retrieval applications-clinical benefits and future directions", International Journal of Medical Informatics, 73:1-23, 2004

[15] E.B. Fernandez, M.M. Larrondo-Petrie, T. Sorgente, "Security models for medical and genetic information", Proceedings of the IADIS International Conference (e-Society 2004), Avila, Spain, July 2004, 509-516

[16] E.B. Suh, S. Warach, H. Cheung, S. A. Wang, P. Tangiral, M. Luby, R.L. Martino, "Web Based Medical Image Archive System", Proceedings of SPIE Medical Imaging 2002, PACS and Integrated Medical Information Systems: Design and Evaluation, vol. 4685, pp. 31-41, May 2002

[17] National Institutes of Health – Center for Information Technology, Division of Computational Bioscience Research Studies, "Design Considerations for Medical Image Archive System",http://dcb.cit.nih.gov/research/repository/repository.pdf

[18] S.S. Agaian, R.R. Sifuentes, R. Cherukuri, "T-order Statistics and Secure Adaptive Steganography", Proceedings of SPIE Mathematical Methods in Pattern and Image Analysis, vol. 5916, Sept. 2005

[19] S.S. Agaian, R.C. Cherukuri, "Run Length Based Steganography for Binary Images", Lecture Notes in Computer Science, Pattern Recognition and Machine Intelligence: First International Conference, PREMI 2005, Kolkata, India, pp. 481-484, Dec. 2005

[20] Ting, T.C., "Application information security semantics: A case of mental health delivery", In Database Security III, Status and Prospects, Elsevier Science Publishers, IFIP 1990, pp. 1-12

[21] Biskup, J., "Protection of privacy and confidentiality in medical information systems: Problems and guidelines", In Database Security III, Status and Prospects, Elsevier Science Publishers, IFIP 1990, pp. 13-23

[22] Pangalos, G.A., Pomportsis, L., M. Khair, "Development of secure medical database systems", In Procs. of DEXA, pp. 680-689, 1994

[23] Anderson, R., Security in Clinical Information Systems, Computer Laboratory, Univ. of Cambridge, ver. 1.1, 1996

[24] Chalmers, J., Muir, R., "Patient privacy and confidentiality", British Medical Journal, vol. 326, pp. 725-726, 2003

[25] Denley, I., Smith, S.W., "Privacy in clinical information systems in secondary care", British Medical Journal, vol. 317, no. 1794, pp. 1328-1331, 1999,

[26] Dugas, M. et. al., Impact of integrating clinical and genomic information, Univ. of Munich, 2001 http://www.bioinfo.de/isb/gcb01/talks/dugas/main.html

[27] Sos S. Agaian, David Akopian, Okan Caglayan and S. A. D'Souza, "Lossless Adaptive Digital Audio Steganography", IEEE 39th Asilomar Conf. on Signals, Systems and Computers, 2005, Monterey, CA

[28] Arnon Rosenthal, Edward Sciore; "Administering Propagated Metadata in Large, Multi-Layer Database Systems", *IEEE Workshop on Knowledge and Data Exchange*, 1999

[29] Min Wu, Hong Heather Yu, "Video access control via multi-level data hiding", IEEE International Conference on Multimedia and Expo, ICME 2000, vol. 1, pp. 381-384, Aug 2000

Author Biography

***Sos S. Agaian** is the Peter T. Flawn Distinguished Professor, College of Engineering, The University of Texas at San Antonio and an Adjunct Professor in the Dept. of Electrical Engineering, Tufts University, Medford, Massachusetts. He has authored more than 300 scientific papers, four books, and holds 13 patents. He is an associate editor of the Journal of Real-Time Imaging, the Journal of Electronic Imaging, and an editorial board member of the Journal Pattern Recognition and Image Analysis. His current research interests lay in the broad area of Signal/image processing and transmission, Information security, and Mobile and Medical Imaging.*

***Okan Caglayan** holds a Master of Science degree in Electrical Engineering from the University of Texas at San Antonio and is currently pursuing Doctor of Philosophy degree in Electrical and Computer Engineering at University of Texas at San Antonio. He is currently affiliated with the Multimedia and Mobile Signal Processing laboratory research team. His research is in the areas of digital signal/image and video processing with hardware implementations, fast transform algorithms, steganography, cryptography and steganalysis.*