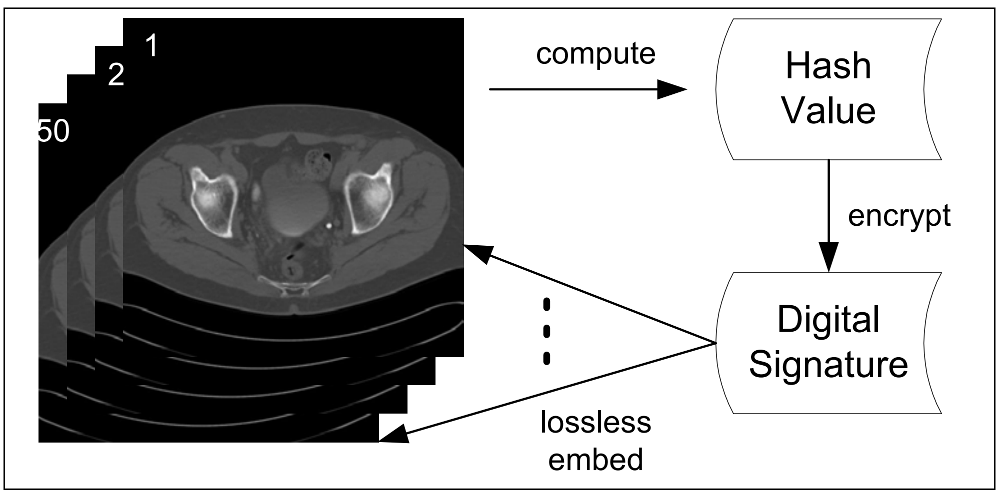
UNIVERSITY OF SOUTHERN CALIFORNIA

Image Processing and **Informatics Laboratory (IPI)**

Annual Progress Report

February 2005



3-D Lossless Digital Signature Embedding

IPI Laboratory,
Department of Radiology

4676 Admiralty Way, Suite 601
Marina del Rey, California, 90292
University of Southern California
Tel: (310) 448-9440
Fax: (310) 448-9441

IMAGE PROCESSING AND INFORMATICS LABORATORY

Department of Radiology
University of Southern California

2005 Annual Report

SUMMARY

The Image Processing and Informatics Laboratory (IPI) settled down in a new laboratory suite with 3,400 square feet at ISI (Information Science Institute), USC, Marina del Rey, CA in February, 2004 (See Pg. vi). Major imaging equipment has been reestablished in the server room and the workstation room as well as network and Internet2 connectivity (See Pg. vii & viii). Five major goals in 2004 were:

1. **Train more post-doctoral fellows and students in Imaging Informatics research.**
   IPI now has five post-doctoral fellows with a mix of M.D. and Ph.D. Six students are pursuing graduate degrees, of these, two are medical doctors with many years of clinical experience enrolling in the imaging informatics Ph.D. program.
2. **Develop the Data Grid Concept and implement a prototype system.**
   The first phase of a Data Grid prototype has been designed and implemented. It won the "infoRAD Certificate of Merit" Award in the 90th RSNA Scientific Exhibit (See Pg. xii). A U.S. patent has been filed.
3. **Research Image Integrity to comply with the HIPAA requirement.**
   A novel 2-D and 3-D "Lossless Digital Signature Embedding" concept has been developed (See Pg. ix).
4. **Embark on the final phase of the bone age assessment of children research project.**
   With the continuous support from NIBIB, the bone age assessment of children research project is entering its last phase. In this third and final phase, additional data will be collected to improve the statistical power of the CAD method in order to be ready for clinical evaluation. The progress report won the "Honorable Mention" Award in the 90th RSNA Scientific Exhibit (See Pg. x).
5. **Assist HCC2 (Healthcare Consultation Center) to implement a total solution PACS.**
   IPI staff assisted the new HCC2 Facility, USC Health Sciences Campus, in implementing a HIS/RIS/PACS/VR total solution for filmless operation, opened for clinical use in July 2004. Along with five BME undergraduate students, an ASP data backup model with Internet 2 connection has been established (See Pg. xi).

In addition, Dr. Huang's new book: "PACS and Imaging Informatics" was published by John Wiley & Sons in March 2004. A Special Issue in the Journal of Computerized Medical Imaging and Graphics: "Imaging Informatics", with Dr. Huang as the Guest Editor, has been completed and will be published in March 2005. Some of the papers contributed by the IPI staff are enclosed in the Selected Reprints and Preprints Section.

This Report also contains selected published and in-press papers during the year, as well as preprints to appear in the *Proceedings of the International Society for Optical Engineering SPIE Medical Imaging*, San Diego, California, February 12-17, 2005.

Our research has been supported by:

- US Army Medical R&D IPA Fellow 30-8034-150475
- NIH R01 EB 00298
- NIH R01 LM07606
- NLM Training Grant 1T15LM07358
- USC RA No. 3051-00
- USC Undergraduate Research Award No. 22-2149-6044
- Array Corporation, Japan

TABLE OF CONTENTS

PREPRINTS FROM SPIE 2005

SELECTED PEER-REVIEWED REPRINTS AND PREPRINTS

IMAGE PROCESSING AND INFORMATICS (IPI) LABORATORY AND COLLABORATORS

Faculty and Administration

Edward V. Grant, M.D., FACR.
Professor and Chairman, Department of Radiology

Michael C.K. Khoo, Ph.D.,
Professor and Chairman,
Department of Biomedical Engineering (BME)

H.K. Huang, D.Sc., FRCR(Hon.)
Professor of Radiology and BME
Director, IPI

Brent J. Liu, Ph.D.
Assistant Professor of Radiology and **BME**

Vicente Gilsanz, M.D.
Professor of Radiology and Pediatrics

Greg T. Mogel, M.D.
Assistant Professor of Radiology and BME

Ewa Pietka, Ph.D. D.Sc.
Professor, University of Silesia, Poland
Visiting Professor of Radiology

Jianguo Zhang, Ph.D.
Professor, Shanghai Institute of Technical Physics,
The Chinese Academy of Science
Visiting Professor of Radiology

James Sayre, Ph.D.
Professor of Biostatistics and Radiological Science
UCLA
Consultant

Maria YY Law, MPhil, BRS, Ph.D.
Assistant Professor
The Hong Kong Polytechnic University
Consultant

Cammy Huang, Ph.D.
Virtual Labs Project Director, SUMMIT
Director of Scientific Outreach
WGLN(Wallenberg Global Learning Network)
Wallenberg Hall

Mary Hall
Administrative Assistant

Postdoctoral Fellows

Paul Huang, M.D.

Nelson King, Ph.D.

Bing Guo, M.D.

Arek Gertych, Ph.D.

Lawrence Chan, Ph.D.,
Hong Kong Polytechnic University

Research Assistant

Michael Zhou, M.S., PhD Candidate

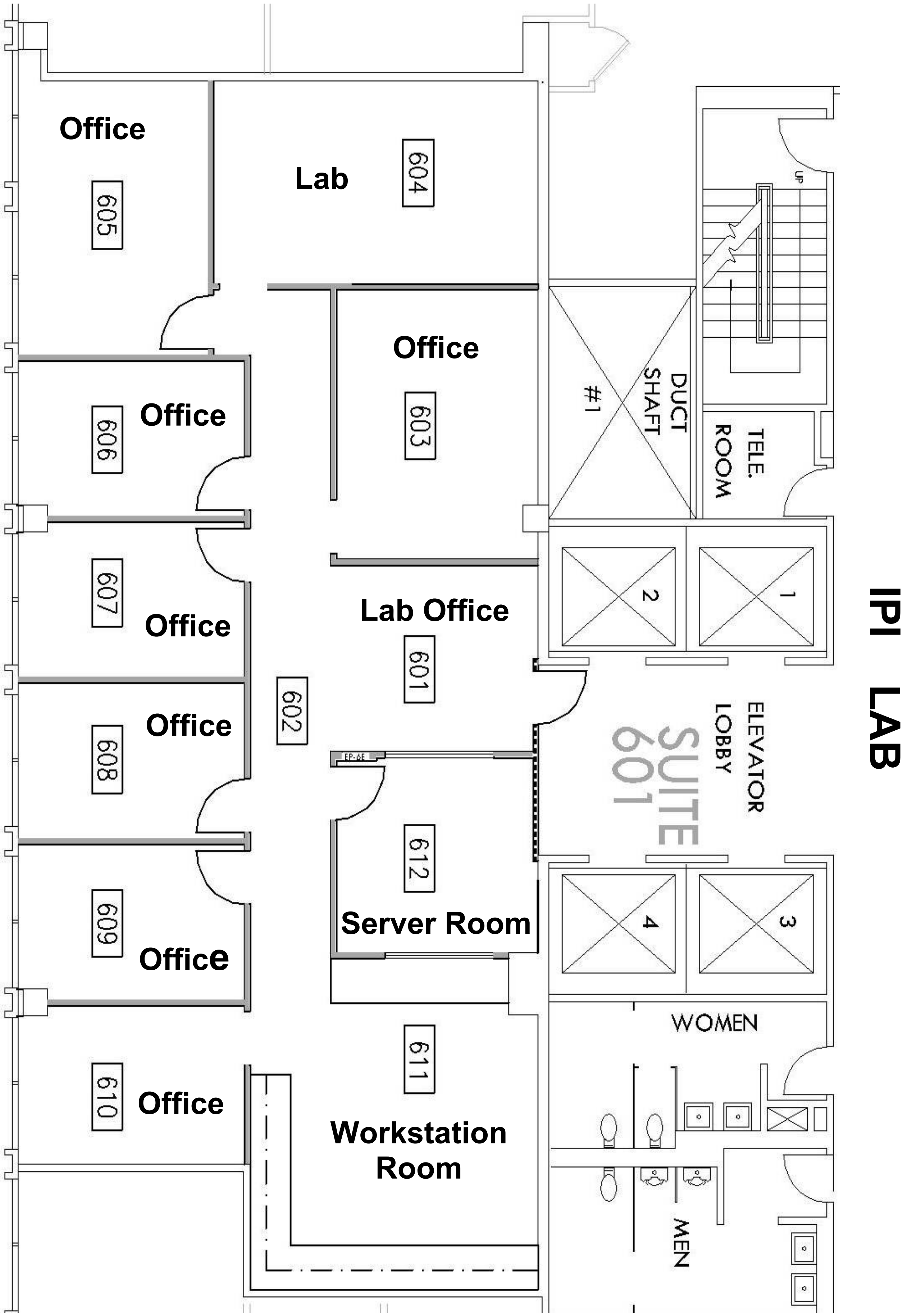
Aifeng Zhang, M.S., PhD Candidate

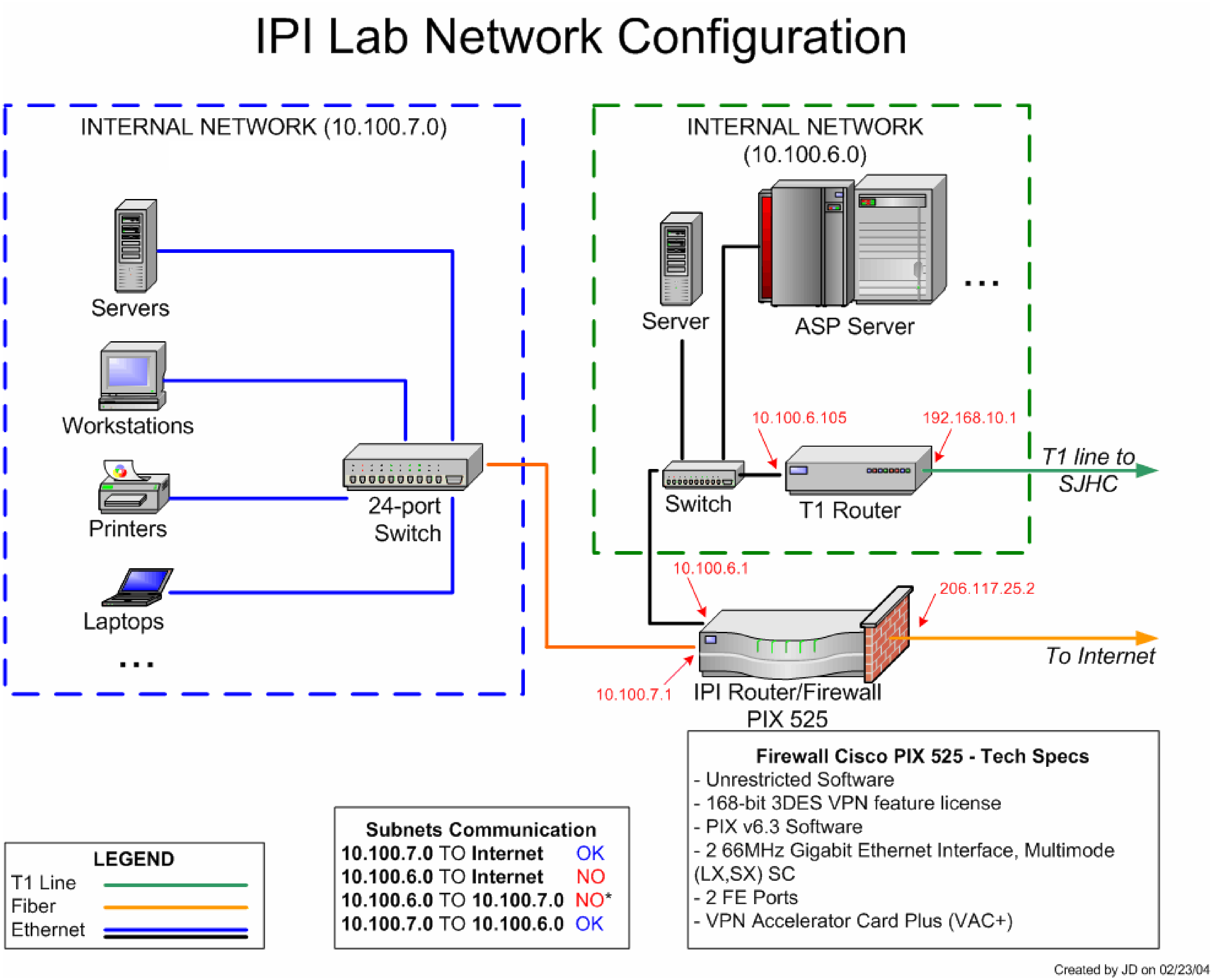
Jorge Documet, B.S.

Manpreet Bedi, B.S.

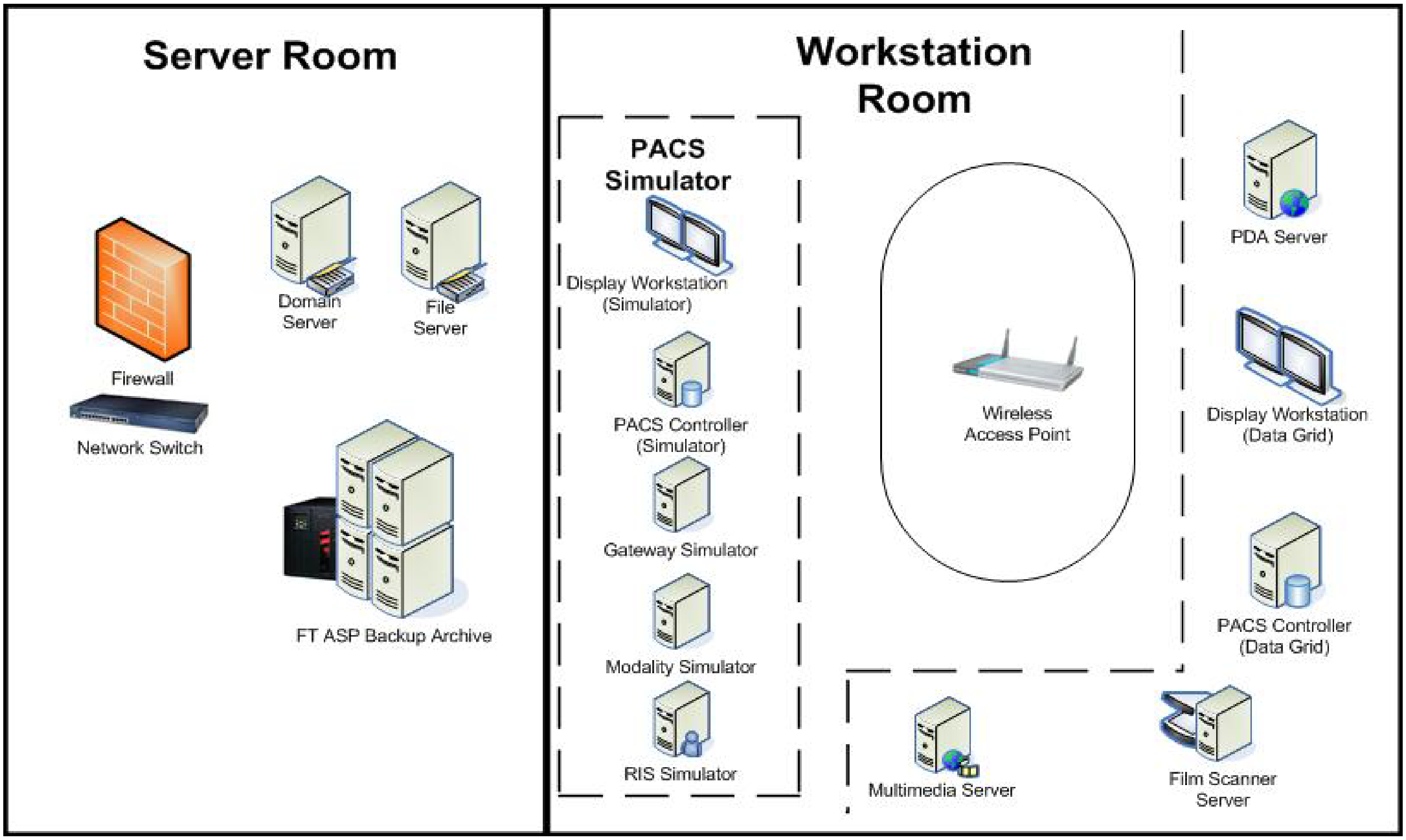
Heston K. Kwong, M.D., MBA, MS.,
Ph.D. Candidate, Hong Kong Poly Tech U.

Tao Chan, M.D.
Ph.D. Candidate, Hong Kong Poly Tech U.

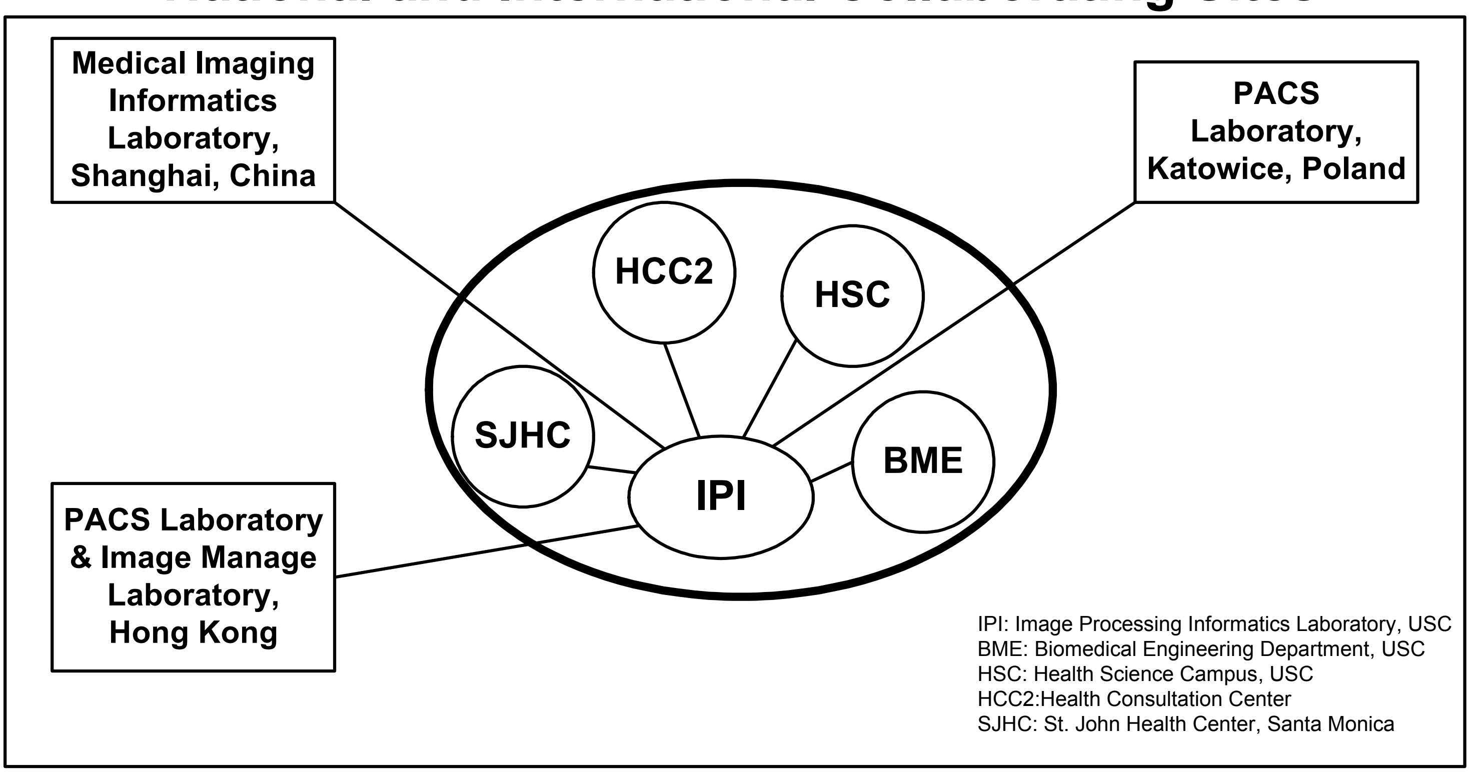

IPI LAB
Office
605
Lab
604
Office
603
Office
606
Office
607
Office
608
Office
609
Office
610
602
Lab Office
601
612
Server Room
611
Workstation Room
DUCT SHAFT
#1
TELE. ROOM
UP
ELEVATOR LOBBY
SUITE 601
1
2
3
4
WOMEN
MEN


IPI Lab Network Configuration
INTERNAL NETWORK (10.100.7.0)
Servers
Workstations
Printers
Laptops
24-port Switch
INTERNAL NETWORK (10.100.6.0)
Server
ASP Server
Switch
T1 Router
10.100.6.105
192.168.10.1
T1 line to SJHC
10.100.6.1
206.117.25.2
10.100.7.1
IPI Router/Firewall PIX 525
To Internet
LEGEND
T1 Line
Fiber
Ethernet
Subnets Communication
10.100.7.0 TO Internet OK
10.100.6.0 TO Internet NO
10.100.6.0 TO 10.100.7.0 NO*
10.100.7.0 TO 10.100.6.0 OK
Firewall Cisco PIX 525 - Tech Specs
- Unrestricted Software
- 168-bit 3DES VPN feature license
- PIX v6.3 Software
- 2 66MHz Gigabit Ethernet Interface, Multimode (LX,SX) SC
- 2 FE Ports
- VPN Accelerator Card Plus (VAC+)
Created by JD on 02/23/04

Imaging Equipment Layout at IPI



National and International Collaborating Sites



Lossless Digital Signature Embedding (LDSE) for Image Integrity in a Fault-Tolerant Clinical Off-Site Backup Archive

Michael Z. Zhou, H.K. Huang, Brent J. Liu

Image Processing and Informatics Laboratory (IPI)
Department of Radiology, University of Southern California

Goals

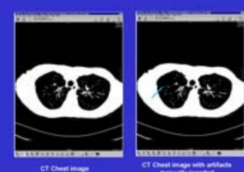
- To develop a lossless digital signature embedding (DSE) method for assuring medical image integrity after image archived or during image transmission.
- To integrate the LDSE method with a fault-tolerant clinical off-site backup archive to provide image integrity assurance.

Rationale

One of the new trends to protect the PACS image data against disaster situations is to store clinical images at an off-site backup archive. In order to support the mission-critical clinical PACS, the backup archive must be 24/7 continuously available (CA). We have developed a novel Data Grid for this purpose by using the grid computing technology, a prototype is shown in Figure 1 (See InfoRAD 9710 NT-i). With the federation of three PAC systems in a grid infrastructure, the Data Grid can provide the true CA (99.999%) image data backup for the three PAC systems.

In our previous work, we illustrated how easy it is to alter an image with an artifact shown in Figure 2, and have developed an image integrity assurance method using a digital signature embedding method (DSE) presented in RSNA last year. However, DSE method would permanently change the least significant bit (LSB) of the image pixels selected for embedding. This method is adequate for most current medical imaging modalities since their LSB contains mostly noise. However, due to the higher density resolution of newer imaging modalities in the horizon, the LSB may also contain relevant data. In this presentation, we presented a lossless digital signature embedding (LDSE) method, which can completely recover the original image pixels after extracting the embedded digital signature, and assure image integrity. We demonstrate the LDSE method by using a Data Grid prototype, because image integrity becomes a vital issue in the Data Grid development, in which image data are no longer under the security protection of local PACS anymore.

Figure 1. A fault-tolerant off-site backup archive using Data Grid.



CT Chest image | CT Chest image with artifacts manually inserted

Figure 2. An example of image integrity problems in Data Grid.

LDSE Process

LDSE method consists of two parts: 1) Sign & Embed operations at the sending site, and 2) Extract & Verify operations at the receiving site, as shown in Figure 3. The Sign & Embed operation includes a Generate Signature process, which is to generate digital signature (DS) of the input image "I", and a Lossless Embed process to embed "s" into I to get the signature embedded image "I*" using a lossless embedding method shown in Figure 4. The Extract & Verify operation is an inverse operation of Sign & Embed operation. It starts with a Extract Data process to extract the embedded signature s and get a recovered image " I' ". A Verify Signature process then verify the signature. Since the image is completely restored, the image integrity can be assured if the verification succeeds.

The Lossless embedding method starts with extracting image features such as randomly selecting a sequence of LSBs (least significant bits) from original image. The extracted LSBs is then losslessly compressed. When the available space defined by subtracting the binary length of the compressed LSBs from that of the original LSBs is sufficient to store the binary length of the DS, the LSB extraction and compression processes end. The concatenated bit stream of the compressed LSBs and the DS then replaces the extracted LSBs in the image pixels. We have developed two LDSE algorithms.

Figure 3. Data flow of Sign & Embed, and Extract & Verify operations in the LDSE method. I: Original Image, I*: Signature Embedded Image, I': the recovered image, s: Signature of the Original Image, u: Result of Verification.

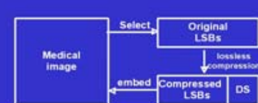


Figure 4. Lossless embedding process.

Two LDSE Algorithms

1. Random Pixels LDSE

Figure 5 shows an example using Random Pixels LDSE to embed the digital signature (DS) in an US image. Assume we can achieve 10% compression ratio for lossless compressing of selected LSBs. In order to embed a 512 bits digital signature, we need to extract about 6000 LSBs from the selected pixels. The compressed LSBs is then concatenated with the bit stream of digital signature and replaces the extracted LSBs. The extraction is a reverse process.

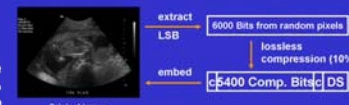


Figure 5 An example using Random Pixel LDSE to embed the DS in an US image. 6000 random pixels were needed to embed the DS losslessly. 'c' is the counter of the binary length of the bit stream.

2. RS Groups LDSE

The method starts with dividing the original image into disjoint groups of n adjacent pixels (x_1, ..., x_n). For example, we can choose n = 4. By applying an invertible flipping operation and a discrimination function on these groups, we can get three group types: R, S and U group. The embedding starts with the scanning of the image pixels to find R and S groups. We then assign 1 to every R group and 0 to every S group to obtain a bit stream, and lossless compress the 1 and 0 bit streams. Group U is not used in this algorithm. After compression, the bit stream of DS can be appended to the compressed bit stream. The newly formed bit stream then replaces the extracted bit stream by flipping the group of pixels. Figure 6 shows an example using RS Groups LDSE to embed DS in a CT image.

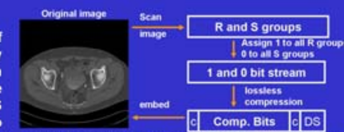


Figure 6 An example using RS Groups LDSE to embed the DS in a CT image. 'c' is the counter of the binary length of the bit stream.

Table 1. Time performance of RS Groups LDSE LDSE.

imaging modalities	Sign & Embed		Extract & Verify	
	Time (s) to generate DS	Time (s) to embed	Time (s) to extract	Time (s) to Verify DS
1 CR	2.1	0.35	0.38	2.0
1 CT	0.18	0.33	0.34	0.15
1 US	0.14	0.48	0.50	0.09
1 MR	0.1	0.3	0.3	0.05

Laboratory Evaluation of LDSE

Both methods have been evaluated in our laboratory. In these experiments, projection images (CR) and sectional images (CT, MR, and US) were used. Figure 7 shows some examples of test results from RS Groups LDSE method. We also recorded the time performance of RS Groups LDSE in Table 1. The time to generate DS and verify DS decrease with the size of image decreasing, while the time to embed and extract are similar to each type of image.

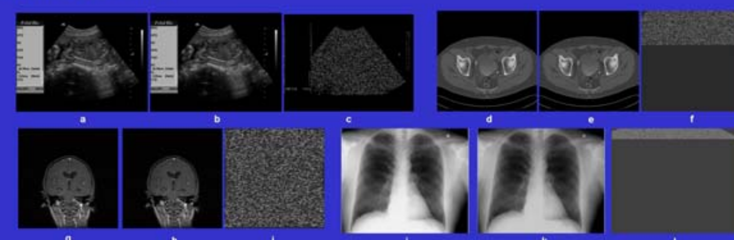


Figure 7 a: original US image, b: image signature embedded US image, c: subtracted image between b and a showing the embedded bits in the embedded image b; d: original CT image, e: signature embedded CT image, f: subtracted image between e and d; g: original MR image, h: signature embedded MR image, i: subtracted image between g and h; j: original CR image, k: signature embedded CR image, l: subtracted image between k and j.

Integration with Data Grid

Figure 8 shows the system integration of LDSE with the Data Grid. The LDSE Sign & Embed Operations can be installed in local PACS. Every image will be signed and embedded before it is sent to the Data Grid. The Extract & Verify Operations can be installed in every component of the Local PACS and in the Data Grid. Therefore, the image can be verified both locally and at the Data Grid, thus assuring the image integrity for the image data everywhere during data flow.

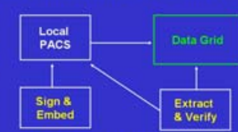


Figure 8 Integration of LDSE with Data Grid.

Summary

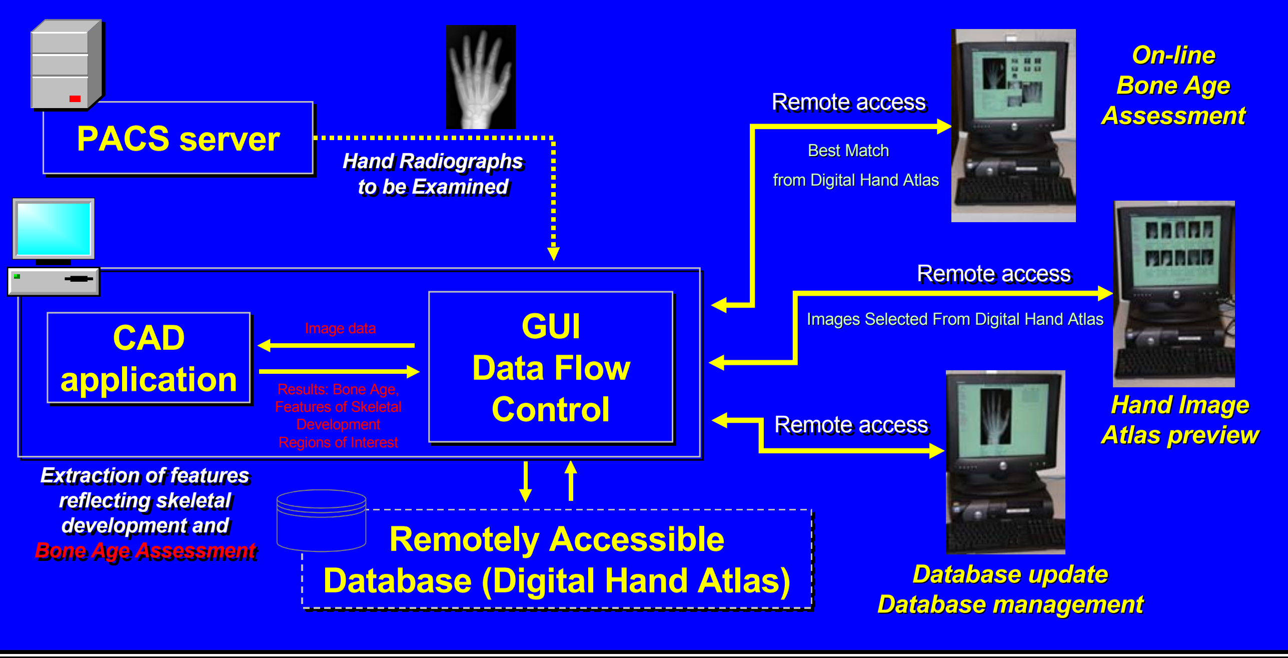
LDSE methods have been developed for assuring image integrity of the image data . We use an off site data grid backup archive as an example. We have also tested LDSE methods with all prevailing medical images at our laboratory. The system integration of LDSE with the Data Grid has been designed. Currently, we are in processing of system integration development.

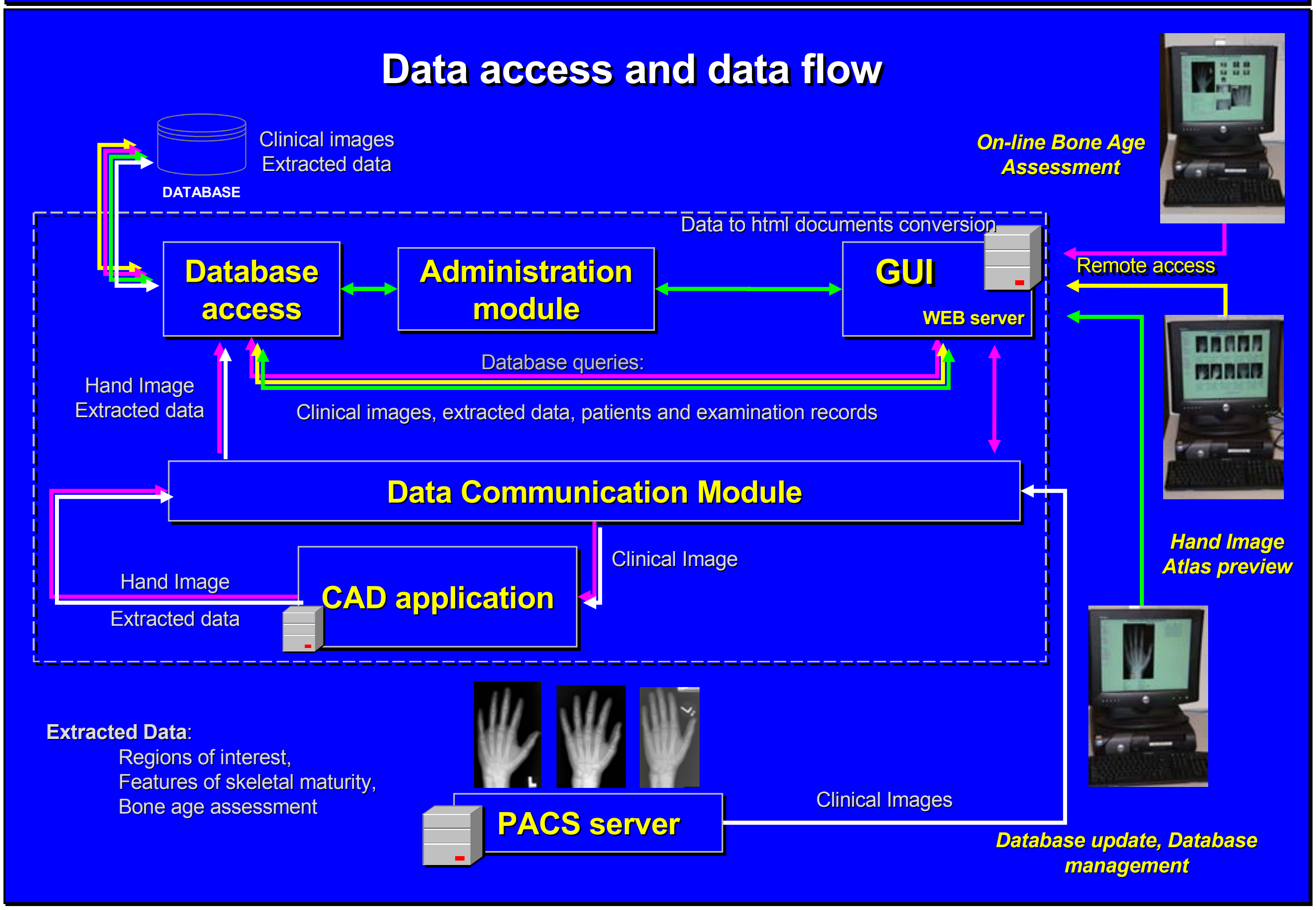
Acknowledgement

This research is partially supported by a NIH Grant No. R01-LM06270 from the National Library of Medicine.

REMOTELY ACCESSIBLE COMPUTER ASSISTED SKELETAL MATURITY ASSESSMENT

E. Pietka PhD[1,2], A. Gertych [1,3] PhD, K. Witko[2] MS.
1 - Department of Electronics, Silesian University of Technology, Gliwice, Poland; 2 - Central Teaching Hospital, University of Silesia, Poland
3 - Information Sciences Institute, University of Southern California, USA





One-Year Clinical Experience at Saint John's Health Center Utilizing the PDA as a Wireless Remote Control for Clinical Image Distribution

Jorge Documet, Brent J. Liu, Luis Documet

Department of Radiology, Keck School of Medicine/University of Southern California;
Imaging Department, Saint John's Health Care, Santa Monica, CA

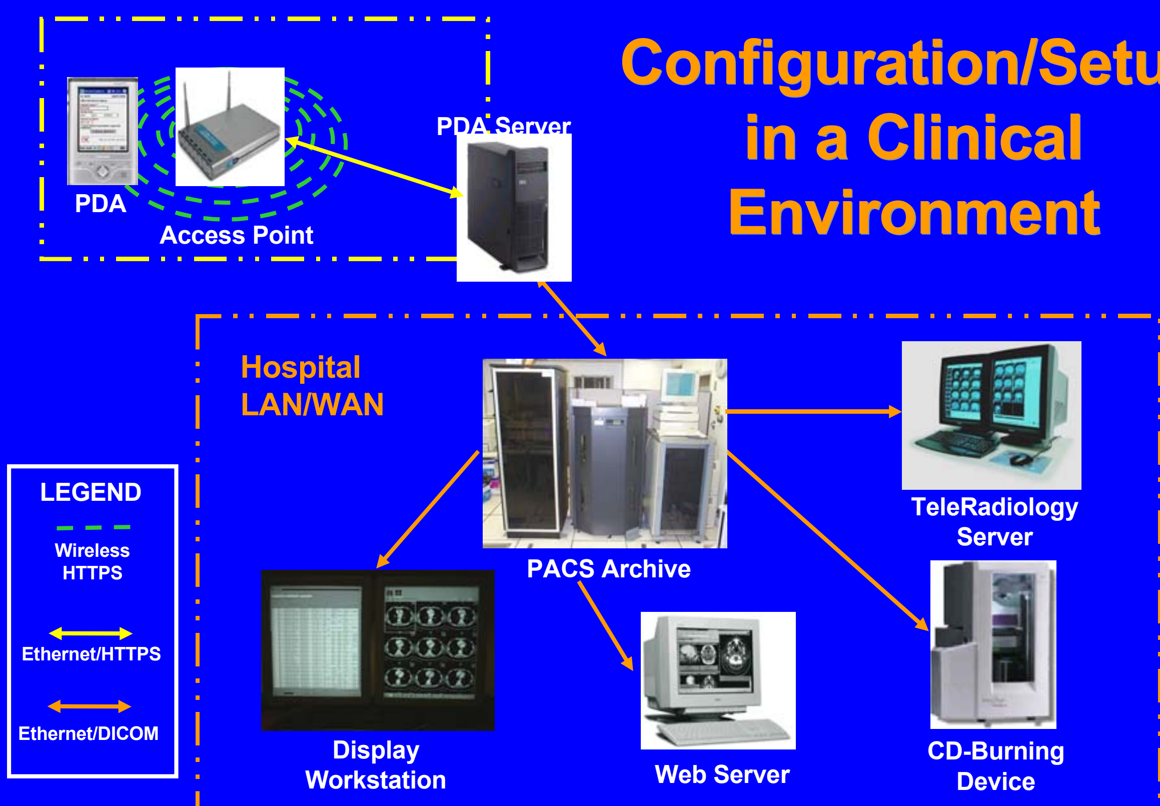
IN THIS PRESENTATION…

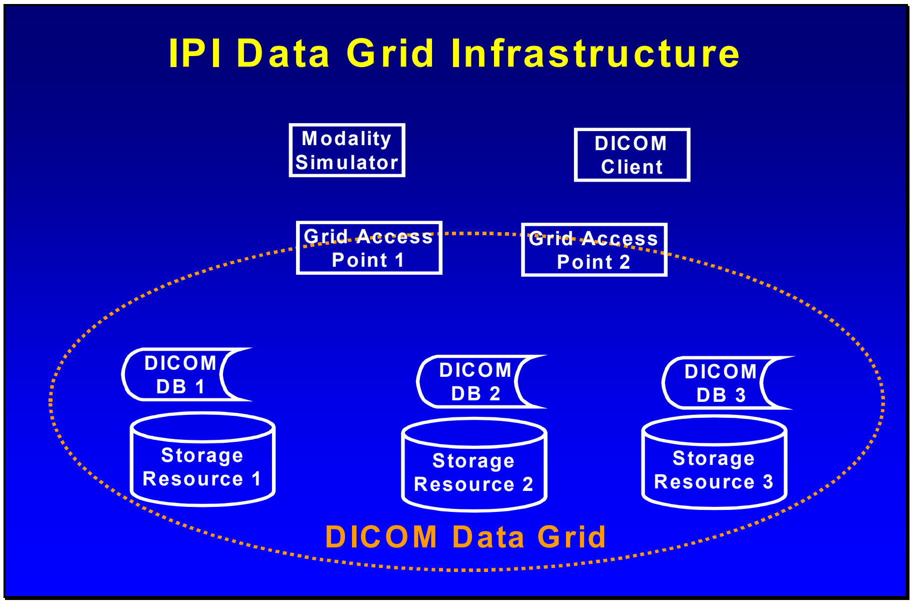
WE Demonstrate: Remote Control Clinical Workflow Management Using a Wireless Device

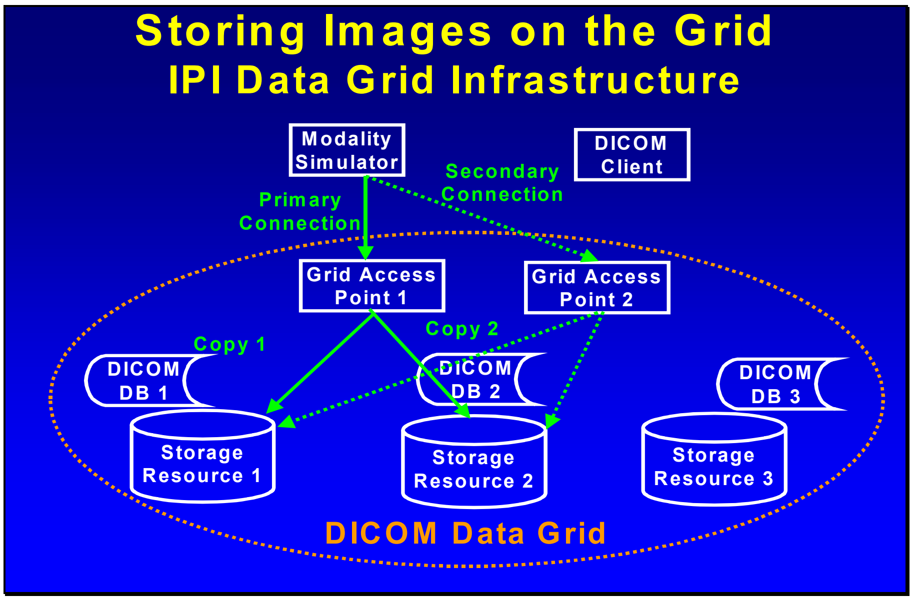
Learning Objectives:

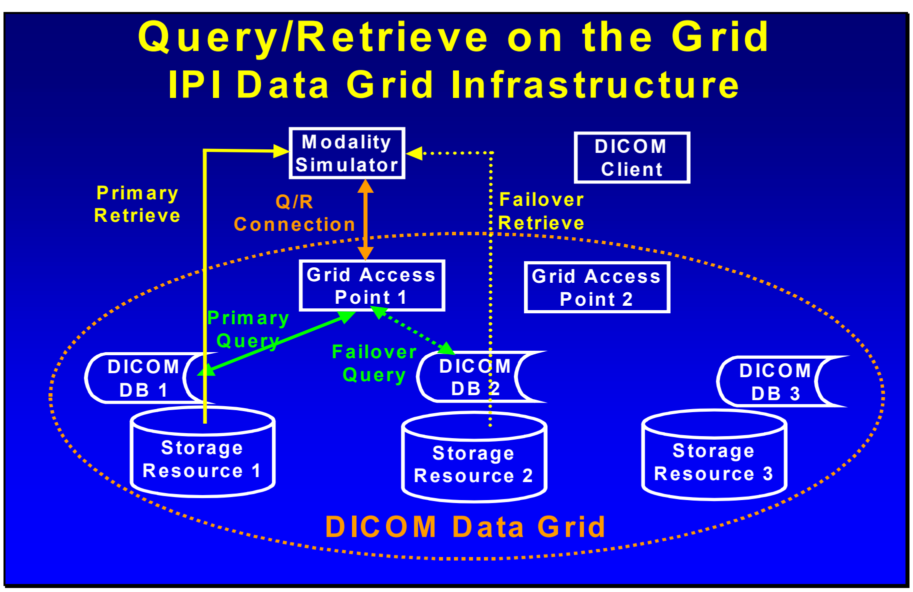
- Demonstrate the capability to manage and distribute PACS image data from a PDA with a wireless network connection.
- Provide a thumbnail capability of the examinations in the PDA as a preview before distribution.
- Provide clinical review from clinical users.

Configuration/Setup in a Clinical Environment




IPI Data Grid Infrastructure
Modality Simulator
DICOM Client
Grid Access Point 1
Grid Access Point 2
DICOM DB 1
DICOM DB 2
DICOM DB 3
Storage Resource 1
Storage Resource 2
Storage Resource 3
DICOM Data Grid


Storing Images on the Grid
IPI Data Grid Infrastructure
Modality Simulator
DICOM Client
Primary Connection
Secondary Connection
Grid Access Point 1
Grid Access Point 2
Copy 1
Copy 2
DICOM DB 1
DICOM DB 2
DICOM DB 3
Storage Resource 1
Storage Resource 2
Storage Resource 3
DICOM Data Grid


Query/Retrieve on the Grid
IPI Data Grid Infrastructure
Modality Simulator
DICOM Client
Primary Retrieve
Q/R Connection
Failover Retrieve
Grid Access Point 1
Grid Access Point 2
Primary Query
Failover Query
DICOM DB 1
DICOM DB 2
DICOM DB 3
Storage Resource 1
Storage Resource 2
Storage Resource 3
DICOM Data Grid

PREPRINTS FROM SPIE 2005

Data mining and visualization of average images in a digital hand atlas

Aifeng Zhang*, Arkadiusz Gertych, Brent J. Liu, H.K. Huang
Image Processing & Informatics Lab, Dept. of Radiology, ISI/USC, Marina Del Rey, CA 90292

ABSTRACT

We have collected a digital hand atlas containing digitized radiographs of normally developed children grouped accordingly to age sex and race. A set of features stored in database reflecting patient's stage of skeletal development has been calculated by automatic image processing procedures. However due to improper hand position and poor image quality, almost 20% of cases in early stage of development were failed for feature extraction. The purpose of this paper is to propose a methodology to fill up the missing features by means of regression. Based on regression results, average feature vectors (AFVs) for each age group were obtained. Then a least median squares (LMS) classifier was trained by AFVs and the final model was applied to predict the bone ages. The accuracy of prediction achieved by this model was compared with fuzzy logic approach and two radiologists' readings. The results prove AFV is an efficient way to predict values of missing data for a given reference hand image.

Keywords: Bone age assessment, data mining, digital hand atlas, least median squares

1. INTRODUCTION

Bone age assessment (BAA) is a procedure performed in pediatric radiology to evaluate a stage of skeletal maturation. A most commonly used method in clinical practice is the book atlas matching assessment developed by Greulich and Pyle[1]. The reference set includes hand images of normally developed children and their demographical data. This method is based on visual comparison of the patient's hand image with images collected in the atlas. The closest match is subjectively selected by the radiologist and yields the bone age of the patient. A difference between the assessed bone age and the patient's chronological age indicates the degree of abnormalities in skeletal development. The book atlas remains unchanged from its initial publication in the early 1950s with data collected entirely from only upper middle class Caucasian populations in the mid west of the US. Due to changes both in population diversity and nutrition, an updated data collection becomes crucial in improving the bone age assessment process. In addition, different radiologists experience and background may lead to different diagnosis performed for the same case thus this approach is subjective by nature. Our study found that two radiologists' readings on the 137 African American girls from newborn to 13 years old have the mean difference at 0.53 years and it could be up to 2.50, which confirm results achieved by other researches[2]. Double reading of hand images (consecutive reading by two radiologists) may increase the accuracy, but at high costs. So, an automatic bone age assessment tool is highly desirable in assisting the radiologists to achieve high efficiency and effectiveness.

Recent works on the field of computerized approach to BAA have already been completed [4, 5, 10]. This hand atlas consists of 1080 cases grouped accordingly to sex and age. Thus there are 19 clusters (newborn, 1 – 18) in eight categories (i.e. Caucasian, African-American, Hispanic, and Asian). These images have been subjected to fully automatic procedure of image processing yielding a vector of features for each successfully processed region of interest [6, 7]. To perform objective evaluation of the bone age based on extracted features, a fuzzy logic classifier has been applied [4]. Using this procedure an average difference of evaluated bone age and chronological age is equal to 1.13 yr for girls and 0.94 boys respectively [3]. In order to evaluate and compare other findings, two readings were performed by radiologists for all images collected in database. Image data collection, extracted regions and classifications results are stored in database[10].

However, due to improper hand position and poor image quality, almost 20% of cases in early stage of development were failed to feature extraction. The purpose of this paper is to propose polynomial regression to solve the missing features problem and study least median square (LMS) [8] approach to predict the bone age. The accuracy of prediction achieved by this model was compared with fuzzy logic approach and two radiologists' readings. We observe a good accuracy in bone age assessment while using new-obtained features value.

*azhang@usc.edu; phone 1 310 448 9436; fax 1 310 448 9441; www.ipilab.org

2. DATA COLLECTION

2.1 Hand atlas

We have acquired a total of 1,080 digitized hand images of normally developed children, 5 images for each group of pre-pubertal children and 10 images for children during puberty from the Childrens Hospital Los Angeles (CHLA) from boys and girls of European, African, Hispanic and Asian descent [3, 4, 5]. This image collection has been subjected to fully automatic image processing including: background removal, hand and regions of interest detection, regions segmentation and features extraction [10, 15]. Then the objective evaluation of bone age is achieved based on features dataset and fuzzy logic algorithm [13].

We asked two radiologist experts for independent readings of our collected image data. The discrepancy between radiologists' readings and chronological age was calculated for each image and we found only 2% of our collections are beyond the standard division of the chronological age. A good image collection is a key issue in further processing and a designing of a classifier which is based on assumption that that the chronological age is equal to the bone age in reference data set.

Due to inaccuracies of hand position in the radiographs almost 10% of cases remain unprocessed. This value mainly includes X-rays performed in newborns and kids below 3 years of age, when up-right hand position is difficult to achieve. The number of not processed images is also increased by additional 10% of cases where the image processing methodology fails because of poor image quality. Hence, the bone age is not calculated by BAA software in those images. Figure 1 is an overview of number of images for each age group with completed features extracted in African American girls at the age ranging from newborn to 12 years old. This category has been selected to perform overall analysis described in this paper.

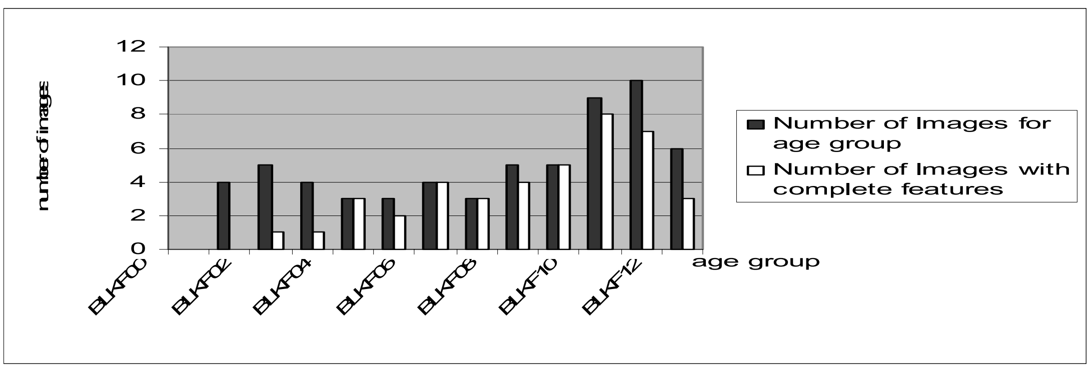


Figure 1: The overview of number of image collection for African American girls

2.2 Features of skeletal development

Image processing software has been successfully developed to segment out a hand image, locate regions of interest[6, 7], and perform detailed analysis of 6 epi-metaphyseal regions of interest (Figure 2) to extract features of skeletal development. A feature vector consists of 11 values calculated based on radiological findings followed by means of region of interest (ROI) segmentation (Figure 3). Distal regions of interest from all hand images have been automatically extracted and subjected to this procedure. At the early stage of skeletal development, the epiphyses are separated from the metaphyses (Figure 3). Due to the bone's growth, they change its shape from a single deposit of calcium to a disc-like shape and move toward the metaphyses.

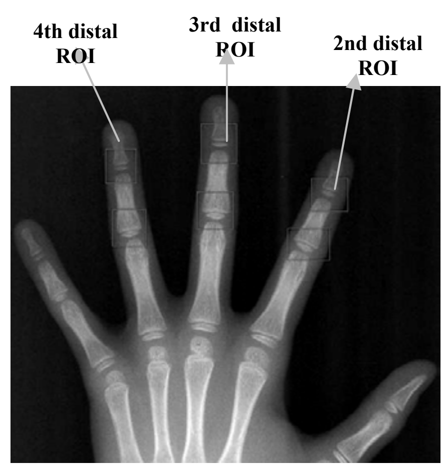


Figure 2: Epi-metaphyseal ROIs marked on 3 fingers.

*azhang@usc.edu; phone 1 310 448 9436; fax 1 310 448 9441; www.ipilab.org

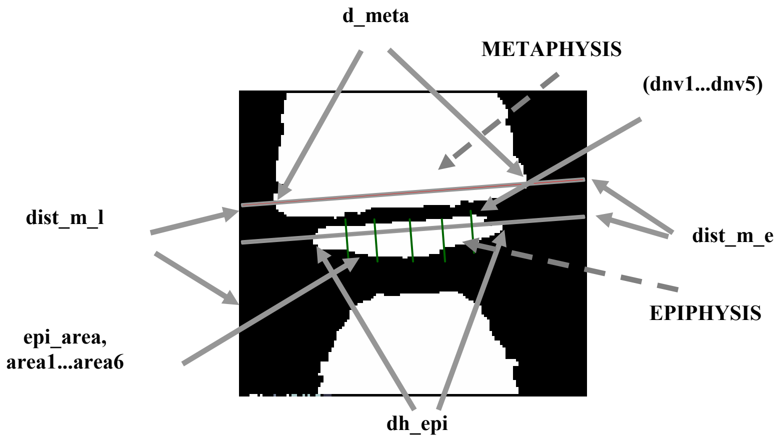


Figure 3: Radiological findings superimposed on the thresholded ROI.

Features describing the size and shape of epiphysis are of high discrimination power in this case. Distance-related features are only available in the children under 12-14 years of age. In a latter stage, the gap between epiphysis and metaphysis disappears and another type of information is gathered. Features reflecting the size of bones are no longer pertinent at this developmental stage. The feature vector consists of 9 distance-related and 2 contrast-related features. The first set of features is defined as follows:

$$F1 = \frac{dh\_epi}{d\_meta} \qquad F2 = \frac{d\_meta}{dist\_m\_l} \qquad F3 = \frac{dist\_m\_e}{dh\_epi}$$
$$F10 = \frac{dh\_epi * dist\_m\_e}{dist\_m\_l * d\_meta} \qquad F8 = \frac{dh\_epi * dist\_m\_et}{dist\_m\_l * d\_meta} \tag{1}$$

Area-related features are defined as:

$$F4 = \frac{area1 + area6}{dnv3 * d\_meta} \qquad F5 = \frac{area3 + area4}{dnv3 * d\_meta}$$
$$F7 = \frac{epi\_area}{dh\_epi * d\_meta} \qquad F6 = \frac{area1 + area6}{(dnv1 + dnv5) * d\_meta} \tag{2}$$

And contrast features are defined as:

$$F11 = \frac{\bar{\mu}_{bony} - \bar{\mu}_{soft}}{\bar{\mu}_{soft} + \bar{\mu}_{bony}} \bar{\mu}_{bcg} \quad (normalized\ contrast - for\ a\ whole\ region\ of\ interest)$$
$$F9 = \frac{\bar{\mu}_{bony} - \bar{\mu}_{soft}}{\bar{\mu}_{soft} + \bar{\mu}_{bony}} \bar{\mu}_{bcg} \quad (normalized\ contrast - for\ an\ epiphyseal\ sub-region) \tag{3}$$

3. METHODOLOGY

This section describes the procedures of the methodology. First, polynomial regression is performed in order to fill the missing feature data. Then from regression results, the average feature vectors were calculated. Finally least median squares classifier was trained by AFVs and the final model was used to predict the bone age for a given hand image with unknown bone age.

*azhang@usc.edu; phone 1 310 448 9436; fax 1 310 448 9441; www.ipilab.org

As described above, 33 (11*3) features were extracted from each image. To evaluate the importance of the features in predicting the bone age, the correlation between feature and chronological age was obtained. The correlation coefficient for each feature *r,* were calculated as

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}} \tag{4}$$

Where: x - is the chronological age, y - denotes the feature variable, $\bar{x}$ and $\bar{y}$ denotes mean values of variables respectively.

Figure 4 shows the absolute value of correlation coefficients for 33 features, from F1_2d, F1_3d … to F11_4d.

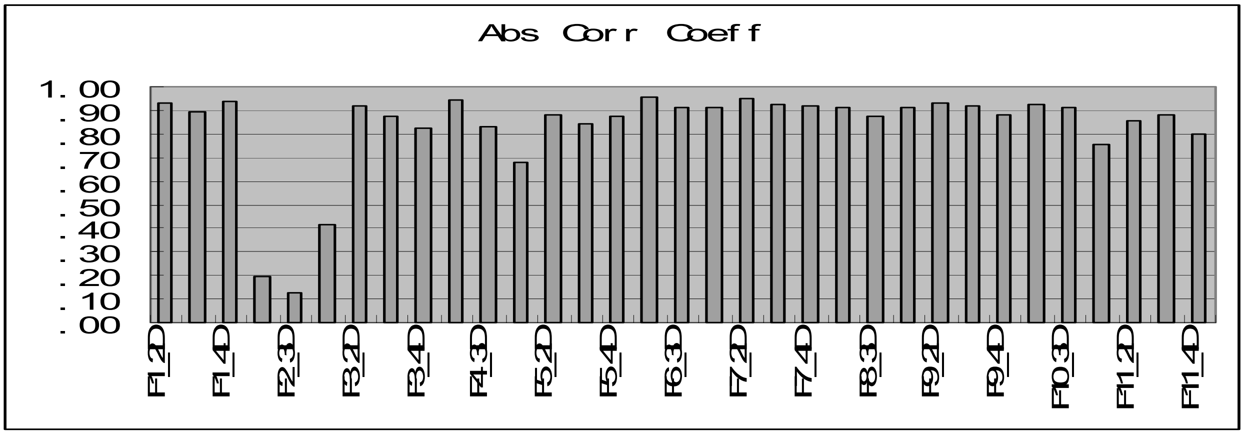


Figure 4 Absolute Correlation Coefficients for 33 features (F1_2d denotes Feature no 1 of 2nd distal ROI)

All the coefficients, except for F2_2d, F2_3d and F2_4d, are within the scope of 0.7 to 1.0, which indicated that the 30 out of 33 features are well correlated with the chronological age. Feature F2_2d, F2_3d and F2_4d by themselves don't reflect correlation between its values and chronological age, since measurements are not related to size of epiphyses. This feature is only used to distinguish early and late stages of development.

3.1 Missing data processing

As we can see in Figure 3, about 20% of the images have missing features. In order to fill the missing features value and complete the database, polynomial regression was performed. As we mentioned in the data collection section, based on our normal collection, we can assume that the chronological age is equal to the bone age. The high correlation coefficient between each of the 30 features and chronological age verifies the accuracy to predict the missing feature values by the chronological age. Polynomial regression with degree of 3 for each feature is described as follows:

$$Y = p_0 + p_1 \cdot X + p_2 \cdot X^2 + p_3 \cdot X^3 \tag{5}$$

Where p_0, p_1, p_2, p_3 are the estimated polynomial coefficients, X denotes the chronological age as input variable and Y denotes the feature as output variable.

This regression procedure was repeated for every feature. Two example results for Feature no. 1 of 2nd ROI and Feature no. 7 of 3rd ROI respectively are shown in Figure 5.

*azhang@usc.edu; phone 1 310 448 9436; fax 1 310 448 9441; www.ipilab.org

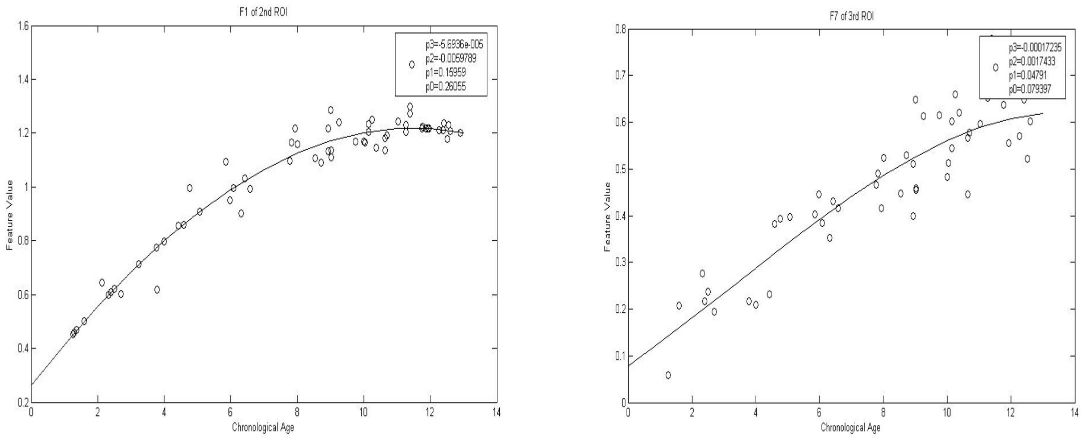


Figure 5 Polynomial regression results for feature F1 of 2nd ROI and feature F7 of 3rd ROI

Considering computation complexity, the degree of 3 was chosen in the polynomial regression. It gives a good approximation on 30 out of 33 features. The regression results with degree 1 and 2 were proved to have bigger error than degree of 3, especially at the age 9 to 12 because more cases were collected at this range.

3.2 Average feature vector (AFV) calculation

From the regression result, the average feature vector (AVF) was obtained by resampling of regression curve with 1 year interval. We have tried to adjust the better interval by testing the methodology with three different intervals equal to 0.2, 0.5 and 1 year respectively. This yields a number of 70, 28, and 14 AVFs obtained for our sample data set. Each AVF consist of 33 average features.

3.3 Bone age assessment by least median squares (LMS)

In the polynomial regression procedure, the chronological age is used to estimate the missing feature values. In reverse, we could predict the age based on the feature values since, the feature and age has strong correlation. But how to combine 33 features to predict is a multi-variant regression problem. In this paper, we use the least median of squares (LMS) method to predict the bone age[8]. In contrast to least square (LS) algorithm, which finds the minimum of sum of the square errors, it estimates the parameters by solving the nonlinear minimization problem:

$$E = \min \underset{i}{med} \{ e_i^{\,2} \} \tag{6}$$

Where $e_i^{\,2}$ is the square error between predicted result and real data and *i* stands for the instance number.

A series of AFVs obtained above were taken as training set. In this study, for every set of AFVs (dimension equal to 33) the regression entity is computed and the least median of square errors of all the other AFV is calculated. When the smallest median of square error is computed for all the possible combinations, the set which has the smallest median of square error will be chosen as the final model. Both LMS and LS were tested on African American girls' category and LMS achieved better prediction accuracy.

4. RESULTS

The quantitative evaluation of methods was performed by applying the algorithm to African American Girls' category. This study did tests with different age intervals, 0.2, 0.5 and 1 year, in the AFV set which trains the LMS classifier, the prediction results vary. It's easy to conclude that the smaller interval, the better prediction. But in this study, decreasing the interval for average feature vector won't improve the predication accuracy. We present the errors between chronological age and predicted bone age in Table 1. For the smallest interval more AFVs contribute to classifier structure and model becomes more complicated and follows the input data, thus larger maximal error is received. For higher intervals value we observe that model adapts to trend of input data and mean absolute value error decreases. Based on this calculation we decided to use 1 year interval which achieves small mean absolute error.

*azhang@usc.edu; phone 1 310 448 9436; fax 1 310 448 9441; www.ipilab.org

Table1 Prediction performance with 3 different intervals

Interval of AVFs	Min absolute error	Max absolute error	Mean absolute error	Standard Deviation
0.2 yr	0	3.20	1.0578	0.79571
0.5 yr	0	2.90	0.8096	0.68717
1 yr	0	2.96	0.7644	0.68351

Table 2 presents the accuracy accomplished by fuzzy logic classifier, radiologists' readings and least median square classifier. Results marked with * indicates that the numbers are obtained from the good images which have all features correctly extracted.

Table 2 Prediction error of different approaches

Error statistics \ Assessment method	Radiologist 1	Radiologist 2	Fuzzy Logic	LMS
Mean absolute error	0.5264	0.5854	0.6108*	0.7644
Standard deviation	0.45262	0.50624	0.42670*	0.68351

LMS makes use of only distal ROI features and provides similar prediction results for all the images in sample test as fuzzy logic tool. We could improve the performance if the 3 middle ROI features were also taken into consideration.

The fuzzy logic algorithm cannot yield the bone age for images with missing features, since no information about skeletal development is provided to this procedure. On one hand it makes it more robust, reliable and protects the final result from heavy errors, but on the other some images remain unprocessed in database. However, LMS can estimate the bone age by filling the missing features with polynomial regression. Table 3 gives a case of the bone age prediction of LMS comparing with two radiologists' readings. Features for one ROI are missing for this particular hand image.

Table 3 Prediction results for 1.3 years-old African American girl

Sample case \ Assessment method	Radiologist 1	Radiologist 2	LMS
Chr Age = 1.30	2.00	1.25	1.52

5. CONCLUSION

In this paper we present an improvement of existing methodology of computerized bone age assessment approach. It provided a solution of the missing feature problem with polynomial regression. A collection of regression curves was determined for all feature types based on sample data set. They serve as a basis for average feature vectors calculations. They can be further applied to diminish the number of unprocessed cases when automatic procedures of image processing fail or when radiographs quality don't allow extracting the correct data. This will improve the overall performance of BAA since more complete feature data will be provided to both fuzzy classifier and LMS algorithm as input.

This study found the optimal interval average feature vector as training set. The advantages of least median squares methods were proved in automatic bone age assessment. And because of the nature of LMS, this methodology has the robustness to outliers in the feature extraction.

6. FUTURE WORK

Better prediction performance could be achieved in two ways. The 3 middle ROI features will also be taken into consideration in the bone age prediction. And, we will refine our collections by excluding the images which are not

*azhang@usc.edu; phone 1 310 448 9436; fax 1 310 448 9441; www.ipilab.org

within the normal distribution in terms of age. The images which have big gap among radiologists' readings and chronological age will be regarded as abnormal.

In older children, calculation of features is based on the horizontal component and wavelet angle of the wavelet decomposition[13]. Further work will be done based on wavelet features.

Statistical evaluation of our previous work indicates the need for more cases during the fast period of rapid maturation. For this reason, we will collect 400 additional cases in the age group of 5-12, male and female. As future data is added to the digital hand atlas as well as future techniques, this system can also grow to provide clinicians an aid in the bone age assessment process.

ACKNOWLEDGMENT

This work has been supported by NIH R01 EB 00298.

REFERENCES

1. W. Grelich, S. Pyle, *Radiographic Atlas of Skeletal Development of Hand Wrist*, Stanford, CA: Stanford University Press, 1959.
2. D. Kirks, *Practical Pediatric Imaging*, Diagnostic Radiology of Infants and Children, (Little, Brown & Company, Boston/Toronto), 1st ed., 1984, chap. 6, pp. 198-201.
3. J.M. Tanner, R.H Whitehouse, *Assessment of Skeletal Maturity and Prediction of Adult Height (TW2 Method)*, Academic Press, London, 1975.
4. E. Pietka, S. Pospiech, A. Gertych, F. Cao, *Integration of Computer Assisted Bone Age Assessment with Clinical PACS*, Computerized Medical Imaging and Graphics, 1-12, 2002.
5. F. Cao, HK. Huang, E. Pietka, V. Gilsanz, *Digital Hand Atlas and Web-based Bone Age Assessment: System* Design and Implementation, Computerized Medical Imaging and Graphics, 297-307, 2000.
6. HK Huang, *Picture Archiving and Communication Systems: Principle and Applications*, Wiley & Sons, NY, p. 521, 1999.
7. A. Zhang, et al. *Data Mining for Average Images in a Digital Hand Atlas*. Proceedings of SPIE Medical Imaging, Vol. 5371, pp. 251-258, February 2004.
8. P.J. Rousseeum, *Least Median of Squares Regression*, Journal of the American Statistical Association, 79, 871-880, 1984.
9. E. Pietka, A. Gertych, S. Pospiech, F. Cao, HK. Huang, V. Gilsanz, *Computer Assisted Bone Age Assessment: Image Processing and Epiphyseal/Metaphyseal ROI Extraction*, IEEE Trans. Medical Imaging, 20, 715-729, 2001.
10. E. Pietka, S. Pospiech, A.Gertych, F. Cao, H.K. Huang, V. Gilsanz, *Computer Automated Approach to the extraction of epiphyseal regions in hand radiographs*, Journal of Digital Imaging, 14, 165-172, 2002.
11. F. Cao, HK. Huang, E. Pietka, V. Gilsanz, P. Dey, A. Gertych, S. Pospiech, *An Image Database for Digital Hand Atlas*, SPIE Medical Imaging, Vol 5033, 461-470, 2003.
12. M. Antonie, R. Osmar, A. Coman, *Application for Data Mining Techniques for Medical Image Classification*, Proceedings of the second International Workshop on Multimedia Data Mining, 94-101, Aug. 26, 2001.
13. S. Pospiech-Kurkowska, E. Pietka, F. Cao, H.K. Huang, *Fuzzy System for the Estimation of the Bone Age from Wavelet Features*, Proc. BIOSIGNAL 2002, Brno, 441-443, 2002.
14. E. Pietka, A. Gertych, Witko K, *Remotely Accessible Computer Assisted Skeletal Maturity Assessment*, Proc. RSNA Chicago, p.807, 2004, Scientific Merit Award.
15. E. Pietka, A. Gertych, S. Pospiech-Kurkowska, F. Cao, HK. Huang, V Gilsanz, *Computer Assisted Bone Age Assessment: Graphical User Interface for Image Processing and Comparison*, Journal of Digital Imaging, 17(3), 175-188, 2004
16. A.Gertych, E.Piętka, *An automated segmentation and features extraction from hand radiographs*, Archive of Applied and Theoretical Informatics, Vol. 15, Book 3, pp 315-326, 2003, ISSN 0867-2121.
17. K. Simmons, *Brush Foundation Study of Child Growth and Development II: Physical Growth and Development*, Periodicals Service Co, 1972.

*azhang@usc.edu; phone 1 310 448 9436; fax 1 310 448 9441; www.ipilab.org

A DICOM-RT BASED ePR RADIATION THERAPY INFORMATION SYSTEM BASED ON THE CYBERKNIFE FOR BRAIN TUMORS

Brent J. Liu[1], Maria Law[2], H.K. Huang[1], C.S. Zee[1], Lawrence Chan[2]
[1]Image Processing & Informatics Laboratory, Dept. of Radiology,
Keck School of Medicine, USC, Marina del Rey, 90292
[2]Hong Kong Polytechnic University, Hong Kong, China

ABSTRACT

The need for comprehensive clinical image data and relevant information in image-guided Radiation Therapy (RT) is becoming steadily apparent. Multiple standalone systems utilizing the most technological advancements in imaging, therapeutic radiation, and computerized treatment planning systems acquire key data during the RT treatment course of a patient. One example are patients treated for brain tumors of greater sizes and irregular shapes that utilize state-of-the-art RT technology to deliver pinpoint accurate radiation doses. One such system, the Cyberknife, is a radiation treatment system that utilizes image-guided information to control a multi-jointed, six degrees of freedom, robotic arm to deliver precise and required radiation dose to the tumor site of a cancer patient. The image-guided system is capable of tracking the lesion orientations with respect to the patient's position throughout the treatment process. This is done by correlating live radiographic images with pre-operative, CT and MR imaging information to determine relative patient and tumor position repeatedly over the course of the treatment. The disparate and complex data generated by the Cyberknife system along with related data is scattered throughout the RT department compromising an efficient clinical workflow since the data crucial for a clinical decision may be time-consuming to retrieve, temporarily missing, or even lost. To address these shortcomings, the ACR-NEMA Standards Committee extended its DICOM (Digital Imaging & Communications in Medicine) Standard from Radiology to RT by ratifying seven DICOM RT objects starting in 1997. However, they are rarely used by the RT community in daily clinical operations. In the past, the research focus of an RT department has primarily been developing new protocols and devices to improve treatment process and outcomes of cancer patients with minimal effort dedicated to integration of imaging and information systems. Our research, tightly-coupling radiology and RT information systems, represents a new frontier for medical informatics research that has never been previously considered. By combining our past experience in medical imaging informatics, DICOM-RT expertise, and system integration, we propose to test our hypothesis using a brain tumor case model that a DICOM-RT electronic patient record (ePR) system can improve clinical workflow efficiency for treatment and management of patients. This RT ePR system integrated with clinical images and RT data can impact the RT department in a similar fashion as PACS has already successfully done for Radiology. As a first step, the specific treatment case of patients with brain tumors specifically patients treated with the Cyberknife system will be the initial proof of concept for the research design, implementation, evaluation, and clinical relevance.

Keywords: DICOM-RT, ePR, Cyberknife, Brain Tumors

1. INTRODUCTION

Stereotactic radiosurgery is the use of high intensity radiation to ablate a targeted area without much damage to its surrounding normal tissue. It has been used for over 30 years for treating brain tumors and other disorders. Throughout the years, different equipment has been invented to facilitate such treatment, ranging from the Gamma Knife to X-knife to Cyberknife. Though accurate, the design of the Gamma Knife limits its use to brain only and requires an invasive immobilization device for framing. The X-knife operates using conventional linear accelerator with modifications making it less accurate to treat tumors. Both the Gamma Knife and the X-knife have been utilized by the RT department for many years now. The Cyberknife system is a newer radiosurgery system that utilizes two distinct innovative technologies. First, it tracks and verifies tumor location to enable automatic compensation for tumor movement utilizing a proprietary image-guidance system. This image-guidance system eliminates the need for an invasive immobilizing device making it frameless. Second, it utilizes a multi-jointed robotic arm with six degrees of freedom to deliver radiation to previously unreachable tumors, reducing damage to surrounding critical structures [1]. This allows for treatment of brain tumors greater than 3.5 cm, irregular shaped tumors, and multiple lesions in the same treatment session with minimal damage to surrounding tissue. For both the Gamma Knife and the Cyberknife treatment planning systems, currently CT and MR PACS studies are utilized to plot the treatment plan.

In order to track the progress of the treatment of brain lesions of patients, clinically relevant data needs to be retrieved from various sources. For example, treatment records for a single RT session would either be stored on paper or within an RT information management system. Another source of data originates from the RT modalities described above that are used to treat brain tumors. Most of these systems can receive DICOM PACS studies which are used to perform treatment planning. However, the treatment planning systems generate image-related data such as isodose curves and structure contours overlayed on the clinical PACS studies. Each RT modality contains a separate proprietary treatment planning system with no standard for retrieving the pertinent clinical data. Another source of data is the original clinical PACS studies used for the treatment planning which reside in a Radiology PACS. Finally, each RT modality may have verification images acquired specifically for the treatment session that are proprietary in format. In order to review a patient's case with brain lesions, the physician will have to interface with all these various sources of data which will increase time and inefficiency in the RT department.

With the emergence of PACS as an imaging informatics tool, it has improved the workflow efficiency within the Radiology Department. [2] RT and the utilization of image-guided RT Systems to treat tumors have benefited from PACS and the DICOM standard by utilizing clinical images from Radiology. However, the real benefit to an RT department is to extend the experience and knowledge gained from system integration of PACS and modalities to the various sources of clinical data dispersed within the RT department. To date within the RT department, one of the biggest challenges towards an effective and efficient clinical workflow is integrating pertinent image and image informatics data into one source point for all clinical users. [3] These include a lack of formal methodology to define the clinical workflow of tumor cases, lack of system integration, and insufficient IT experts in RT-related applications.

The need for an integrated solution utilizing a DICOM-based ePR server becomes apparent. In particular, the treatment of patients with brain tumors involve radiation treatment with the Cyberknife, which adds another challenge since the data derived from the treatment planning system can be quite complex to incorporate into the DICOM-RT data model. Currently, CT and MR clinical images are primarily utilized by the treatment planning systems. The redesign of workflow to incorporate the information will help to enrich the electronic patient records. Given this backdrop, the implementation of a DICOM ePR server would be an effective and efficient one-stop-shop source for tracking the treatment progress of patient's with brain tumors by merging the data from the various sources and present them in a user interface that will be designed for ease-of-use.

Various generic RT information/management systems feature the availability of necessary clinical data within the RT department. However, the most complete clinical data model is from the proposed DICOM-RT based ePR system of this research. Furthermore, the DICOM-RT based ePR system features open system integration based on the DICOM standard instead of proprietary like other RT information/management systems. In summary, the DICOM-RT based ePR system has the following superior features:

- Complies with DICOM-RT Object definitions
- Global data distribution
- Global Treatment Updates
- Open System Integration

In this research, a general RT workflow for brain tumors will be introduced along with the design criteria for the DICOM ePR system including the DICOM data model, the data collection plan, and integration within the RT department. Medical image and informatics data from one brain tumor patient treated by the Cyberknife system will be presented in the results section.

2. METHODS AND MATERIALS

2.1 Workflow Model for RT of Brain Tumors

One of the most important first steps for system integration of clinical image and information systems is to research the workflow model of the clinical operations. Since the focus of this research will be on patients with brain lesions, the workflow related to these particular treatment cases will be studied.

The clinical workflow model for brain tumor cases for the Departments of Radiology and Oncology is shown in Figure 1. It shows a preliminary workflow model for treatment of brain tumors including the special clinical workflow for the Cyberknife system as well as other systems such as the Gamma Knife. The treatment begins with the patient diagnosed with brain lesion or multiple brain lesions. The Radiologist reviews the diagnostic MR and determines not only whether to treat the tumor(s) but also what type of Radiotherapy will be performed. If conventional RT or surgery is prescribed, then a follow up MR may be acquired or the original MR is used if the study is acceptable. For treatment on the Cyberknife system, a CT study is acquired. Since the Cyberknife is frameless it utilizes the bony landmarks in the CT study to plan the treatment. However, sometimes the tumor(s) are not well delineated in the CT acquisition and an additional MR study may be ordered to assist in the delineation. Once the diagnostic studies are acquired, they are used to plan the treatment. Multiple treatment sessions may be required. The Radiologist reviews the treatment plan along with the Radiation Oncologist. Once the treatment session(s) are completed, the corresponding RT data are recorded in the treatment planning systems of the RT modalities as well as in the Oncology Information System. Follow up MR studies are acquired at one, three, and six months respectively after the treatment sessions are completed to track progress.

The utilization of a DICOM-RT ePR system can occur in four clinical points-of-decision within the general RT workflow.

1) The treatment planning area: Information is needed to design the treatment plan for the particular session.
2) The radiation oncologist's office: Information is needed for a clinical decision and treatment plan approval both before and after the treatment session.
3) The RT unit or modality: Information is needed to assist during the actual treatment session at the radiation treatment unit.
4) Referring Physician's office: Treatment summary is needed for referring physician follow up of the patient.

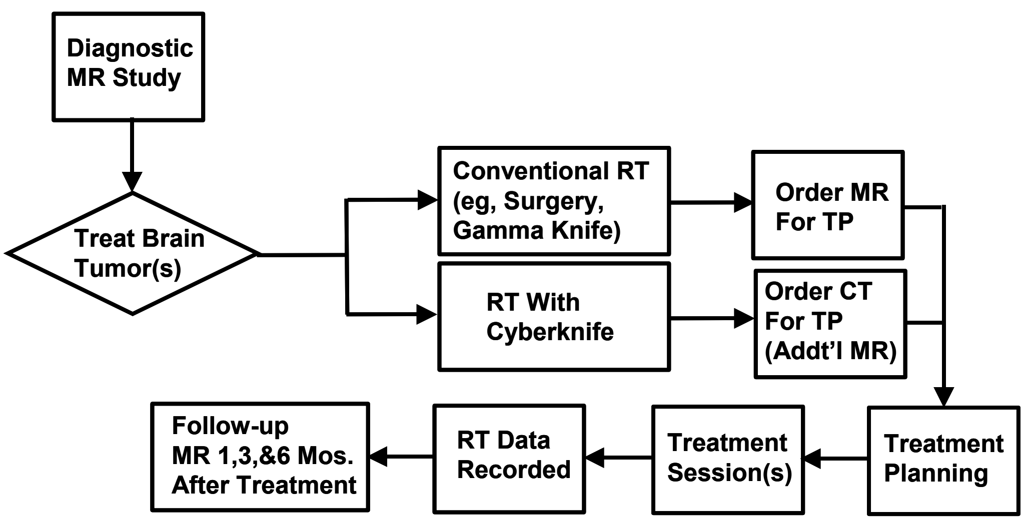


Figure 1: General Clinical Workflow for RT of Brain Tumors Including Cyberknife and Gamma Knife. TP: Treatment Plan

Since there are a variety of brain tumor types, the treatment paths can differ. It is important to research and develop a more robust workflow model that can accommodate the various treatment paths. Not only would this enhance the design of the DICOM-based ePR System, but also serve as the foundation for a methodology to investigate other tumor types in the future.

2.2 DICOM-RT Data Model Development

The DICOM (Digital Communication in Medicine) standard has been well established and widely successful for clinical imaging systems in Radiology, in particular PACS (Picture Archiving and Communication System). Image data acquired from equipment from different vendors can readily communicate with each other and integrate into a system through the DICOM standard. In 1997, the DICOM standard was extended to include radiotherapy information and further updated in the latest version released in 2003. [4,5] Seven DICOM radiotherapy (DICOM-RT) objects have been included by the DICOM standards committee for transmission and storage of radiotherapy images and related information. These DICOM-RT objects are: 1) RT Image, 2) RT Plan, 3) RT Structure Set, 4) RT Dose, 5) RT Beams Treatment Record, 6) RT Brachy Treatment Record, and 7) RT Summary Record as shown in Figure 2. [6] Generally, the sources for these data comes from a treatment planning system (TPS), a RT information system, and both RT and Radiology modalities. The DICOM-RT object information models can be utilized to develop the data structure for the electronic patient record.

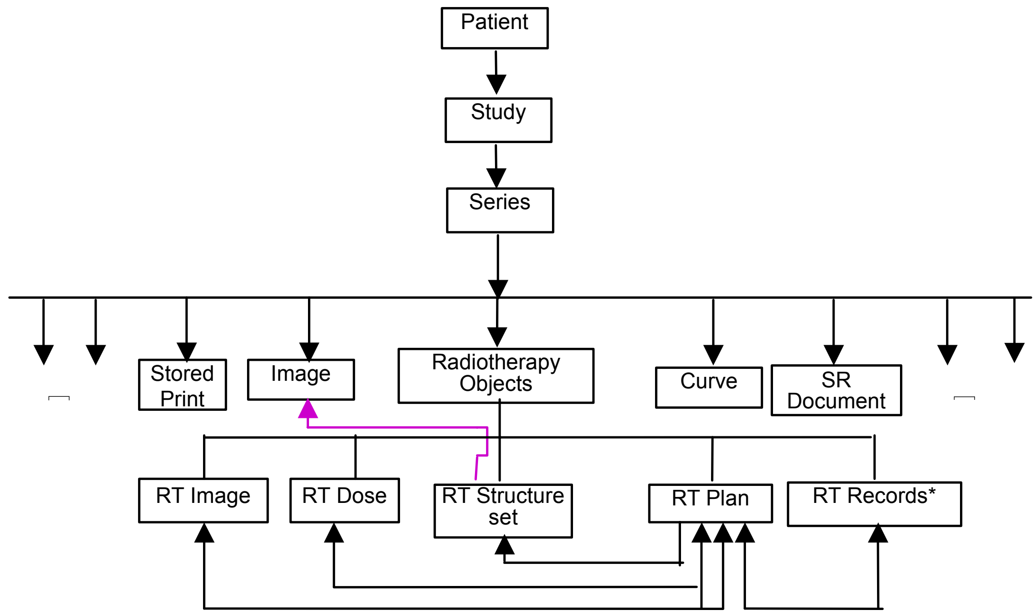


Figure 2: Portion of the DICOM Model of the real world. Note that the three records in DICOM-RT (Beam, Brachy, and Summary) are grouped under RT Records as denoted with the asterisk.

To develop a conceptual data model, the RT workflow must be reviewed to define the data required. Additionally, clinical user input is needed as well. With these input sources, a conceptual model can be developed for an RT electronic patient record. Table 1 shows an example of a conceptual data model that can be further refined with more details and input sources. This model includes a timeline showing the occurrence of DICOM images and DICOM objects for a radiation treatment of a patient with brain cancer.

2.3 Data Collection

A data survey was performed to track patient cases utilizing the clinical information systems at USC/HSC. Patient cases that exhibit brain tumors was tracked to determine the treatment path and outcome. These results were implemented into the clinical workflow model. Brain tumor cases from the USC Oncology department will be tracked. From these brain tumor cases, a total of 50 cases treated by the Gamma Knife and Cyberknife will be extracted for a specific workflow model and later conversion into the DICOM-RT data model. The preliminary data collection survey was performed to determine the feasibility of data collection for the treatment of brain tumors at USC/HSC. The brief survey was performed using clinical information systems to track historical patients and their records. Table 2 summarizes this survey. During the past 6 months, a total of ninety-four brain tumor cases were identified. The data survey was performed under the HIPAA Regulations and Compliance Guidelines set forth by USC. In addition, one data set from the Cyberknife System was reviewed and determined to be feasible to convert the non-DICOM data into DICOM-RT compliant objects.

2.4 System Integration

In June 2004, a fully filmless and paperless environment was implemented at the Health Care Consultation Center II (HCCII) located on the Health Science Campus, University of Southern California. The PACS implemented at HCCII stores Radiology clinical images and was designed with long-term storage capable of supporting the entire Health Science Campus' clinical image data. The Radiation Oncology department features the latest state-of-the-art

Patient Name	Jane Doe																
Patient ID	A123456																
Sex	Female																
Age	45																
Telephone	310-867-5309																
Address	1111 S Figueroa Street, Los Angeles, CA 90015																
Course	1																
Region	Brain																
	1 Aug 02	2 Aug 02	3 Aug 02	4 Aug 02	5 Aug 02	6 Aug 02	7 Aug 02	8 Aug 02	9 Aug 02	10 Aug 02	11 Aug 02	12 Aug 02	13 Aug 02	14 Aug 02	15 Aug 02	16 Aug 02	17 Aug 02
DICOM Image																	
CT/CT Sim	x																
MR																	
RT Image																	
Simulation image		x															
DRR																	
Portal image							x										
Treatment Plan (RT Structure set, RT Plan, RT Dose)					x												
RT Beams Record					x												
No. of treatment							1	2	3			4	5	6	7	8	
Brachy Record																	
Treatment Status									Continue RT							Continue RT	
Treat. Comment																	
Treatment Summary Record	x	x			x		x	x	x			x	x	x	x	x	

Table 1: Conceptual ePR Data Model w/ Time.

Year of Surveyed Cases	No. of Months Surveyed	No. of Brain Tumor Cases Found	No. of Cases Treated w/ Gamma Knife	No. of Cases Treated w/ Cyberknife
2004	6	94	41	8

Table 2: Summary of Prelim Data Survey of Brain Tumor Cases at USC/HSC

image-guided RT systems, including the Cyberknife, which was installed at the Norris Cancer Center and has been in clinical use since October 2002. [7] The entire HSC has connectivity to the two systems, which makes the integration of a DICOM ePR system possible. For the DICOM RT ePR system, a three-tier architecture was developed [8]: 1) The RT archive server manages, archives and distributes DICOM images and DICOM-RT objects, 2) The RT web-based application server processes patient planning and treatment data, and 3) the RT web-based client application presents the RT data. The database schema reflects this three-tiered system by physically representing the data as well as providing data structures, file organizations and mechanisms for system operation as well as data storage. In the design of the RT workflow, there are two database schemas developed; one for the RT archive server and the second for the RT web-based ePR application server. Because there is more RT data presentation at the web-based application server level, the latter database schema is much more complex as compared to the RT archive server. Based on the Data Model and the Clinical Workflow Model, the data workflow was designed based on Figure 3 for system integration [8]. Data from the Oncology Information System and Cyberknife treatment planning systems will be converted into DICOM-RT objects and sent to the DICOM RT Gateway. The diagnostic images will be sent from the USC PACS Server into the DICOM RT Gateway as well. Once the DICOM-RT objects have been received by the DICOM RT gateway, they will be sent to the Archive server. A database schema will be developed for the archive server so that the DICOM RT objects can be archived and distributed to the web-based application server. This archive server is a Continuous Available (CA) server design with 99.999% uptime that has been utilized for a variety of clinical applications. [9,10]

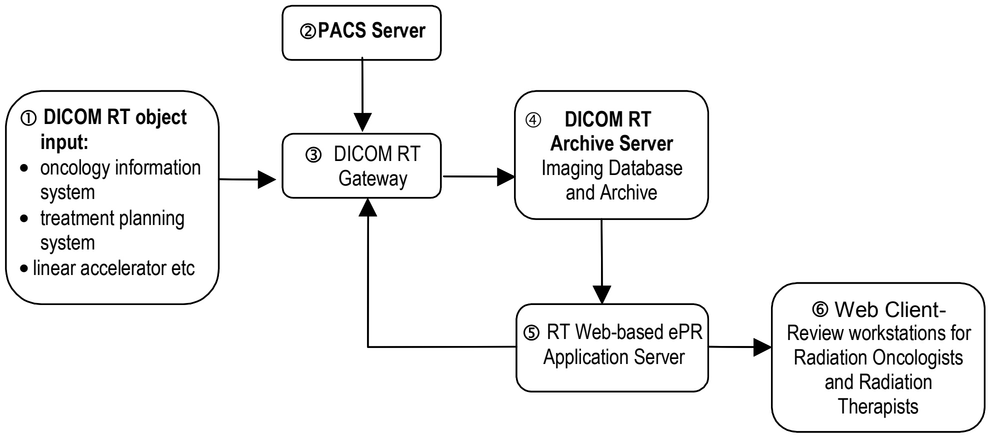


Figure 3: The RT Data Workflow. 1) RT data input from various different RT sources; 2) Diagnostic images from Radiology PACS; 3) Conversion into DICOM-RT objects by the DICOM-RT Gateway; 4) RT archive server stores, manages, and distributes RT data; 5) RT web-based ePR application server further manages and prepares patient planning and treatment information; 6) Web-based client review workstation displays RT-related data for clinician review.

Integration of the DICOM-RT ePR System within the clinical environment includes four major points-of-decisions and will be described in the following paragraphs:

1) Treatment Planning in the Treatment Planning Area: In planning the treatment session for a patient with brain tumor(s), pertinent clinical information is needed. For example, the patient's previous treatment history from the treatment planning systems as well as patient files from the oncology information system is needed to plan the current treatment session. In addition, the Cyberknife treatment planning system uses diagnostic CT and MR exams to plan the image-guided treatment. Traditionally, the diagnostic studies are transmitted directly from Radiology Imaging Modality, retrieved from Radiology PACS, or stored on data media and then loaded into the treatment planning system. However, treatment planning systems were not designed for robust DICOM transmissions between either a Radiology modality or a PACS and transferring imaging data through data media is time-consuming and inefficient. The DICOM-RT server can receive images directly from a Radiology modality or PACS, then based on the workflow, either automatically transfer the imaging data to the respective treatment planning systems or perform an on-demand DICOM send from the DICOM-RT server. System integration research and development between the DICOM-RT server and the respective treatment planning systems will insure a more robust clinical workflow when diagnostic CT and MR exams are required. In addition, the availability of important historical clinical imaging and information at the point-of-decision will help to improve the treatment planning process. At this point-of-decision, a user-interface will be designed tailored to the specific needs and requirement for viewing pertinent information. A web-based client application was implemented for the users to assist in their decision-making process.

2) Treatment Planning Verification Before and After Treatment Session at the Radiation Oncologist Office: This major point-of-decision in the RT clinical workflow is the verification of the treatment planning and follow-up by the Radiation Oncologist. Therefore, it is crucial for Radiation Oncologists to have every available clinical imaging and information data to make as accurate a decision as possible. Because the existing clinical workflow has limitations on both time and availability to retrieve the necessary data, the DICOM-RT ePR system must be researched and developed to address these needs. For example, development of the database schema to include all necessary clinical data and user interface design for the RT Web-based application server are two very key milestones for the successful integration of the system at this point-of-decision. Because image-guided treatment planning systems data can be complex, the presentation of the image and information data may reflect that same level of complexity. Based on input from both Radiation Oncologists and Radiation Therapists at USC/HSC married with existing data and workflow models, a database schema and user interface design was developed to meet the clinical needs. This was implemented in the Web-based application server as well as the web client. Figure 4 shows a preliminary user interface workflow model showing the different functions that a clinician end-user would perform. [8] This model will be expanded and modified to accommodate the complex nature of the

Cyberknife RT data. Once the user-interface design as well as the database schema was developed for this most important point-of-decision, it can be extended to the other points-of-decision more readily since the requirements of the Radiation Oncologist include the largest amount of available clinical image and information data.

3) Treatment session at the RT Unit: Prior to executing the treatment plan, it may be necessary to review the treatment plan or reference pertinent clinical imaging and information data. At this point-of-decision, a user-interface was designed tailored to the specific needs and requirement for viewing pertinent information. A web-based client application was implemented for the users to assist in their decision-making process.

4) Case Review at the Referring Physician's Office: Based on input from both Radiation Oncologists and Referring Physicians at USC/HSC married with existing data and workflow models, the user interface design will be developed to meet the clinical needs. This was implemented in the Web-based application server as well as the web client. Specific developments include the UI design for a patient summary overview and an ePR timeline data model as shown previously in Table 1.

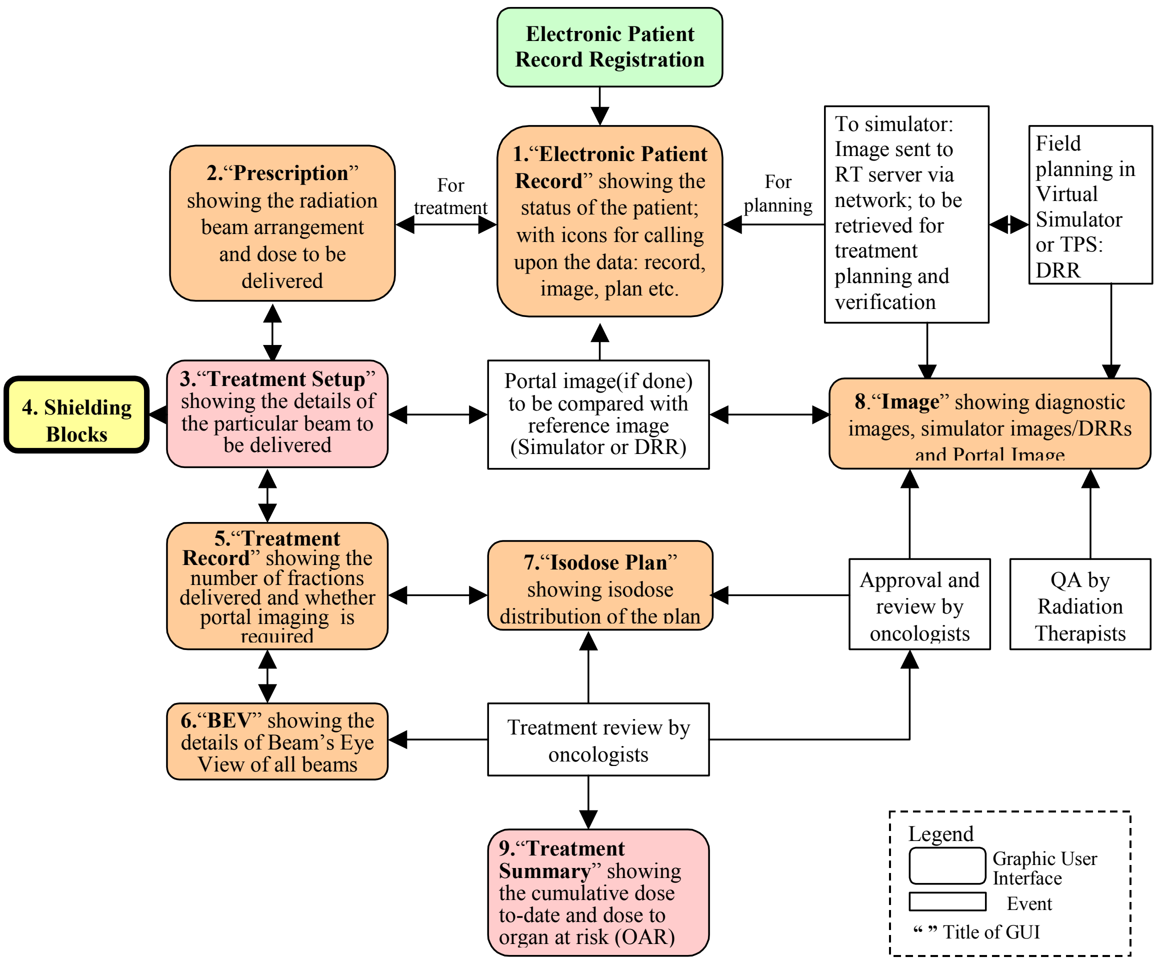
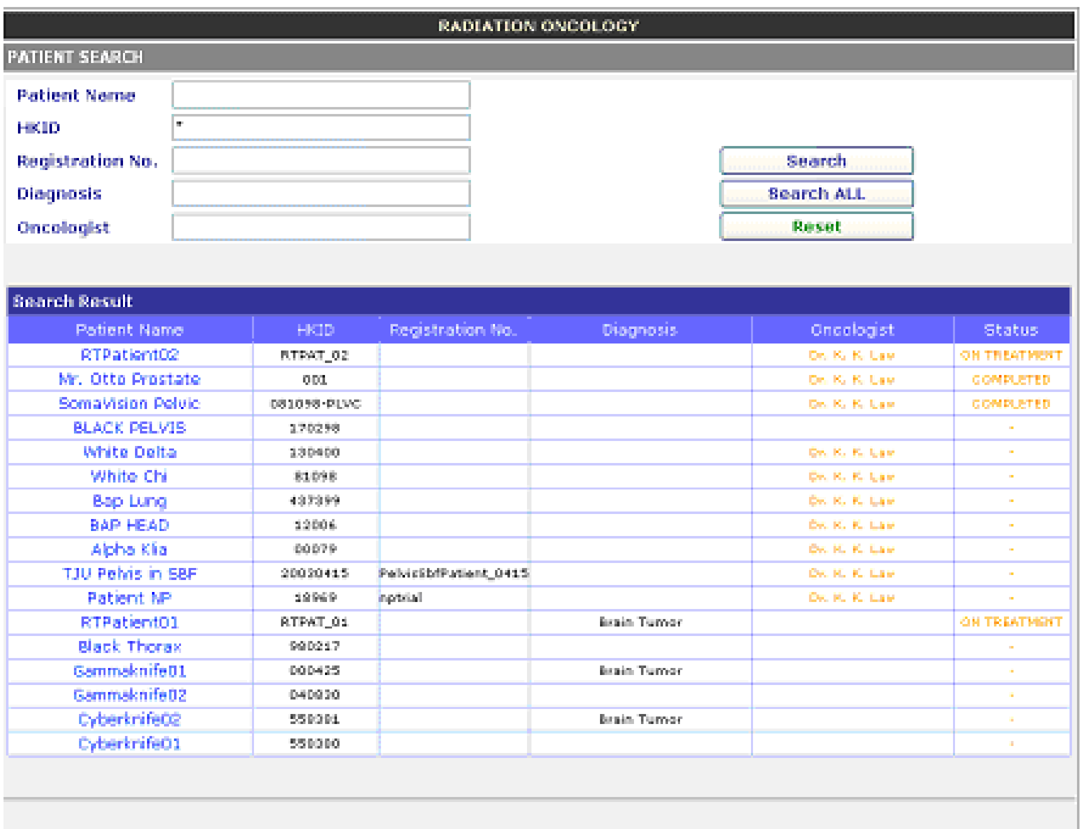


Figure 4: User Interface Workflow Describing the Different Functions

3. RESULTS AND DISCUSSION

This section demonstrates the results obtained from a sample brain tumor case treated by the Cyberknife RT system integrated within the proposed DICOM-RT based ePR system. The end result is a comparison between what clinical information is displayed by a conventional RT information system provided by a manufacturer versus that of the richer database of the preliminary DICOM-RT based ePR system which can provide more information in the display.

Figures 5, 6, and 7 show initial results of this sample brain tumor cases displayed by the GUI design of the web-based client. Figure 5 shows the results of a patient-based search in the DICOM-RT based ePR system. With this ePR system, the users can perform queries based on the DICOM data fields (eg, Patient name, Medical Record Number, Diagnosis) similar in clinical workflow to PACS and Radiology. Figure 6 shows CT diagnostic images overlayed with the DICOM-RT structure set that was acquired within the Cyberknife treatment planning system. These overlays represent the critical structures outlined within the Treatment Planning System for the Cyberknife. The blue colored contour encloses the tumor targeted for treatment. The red colored contour encloses the optic chiasm. The green colored contour encloses the left nerve. All three contours are displayed in the upper lefthand CT image slice.



Patient Name	HKID	Registration No.	Diagnosis	Oncologist	Status
RTPatient02	RTPAT_02			Dr. K. K. Law	ON TREATMENT
Mr. Otto Prostate	001			Dr. K. K. Law	COMPLETED
SomaVision Pelvic	081098-PLVC			Dr. K. K. Law	COMPLETED
BLACK PELVIS	170298				-
White Delta	130400			Dr. K. K. Law	-
White Chi	81098			Dr. K. K. Law	-
Bap Lung	437399			Dr. K. K. Law	-
BAP HEAD	12006			Dr. K. K. Law	-
Alpha Klia	00079			Dr. K. K. Law	-
TJU Pelvis in SBF	20020415	PelvicSbfPatient_0415		Dr. K. K. Law	-
Patient NP	19969	nptrial		Dr. K. K. Law	-
RTPatient01	RTPAT_01		Brain Tumor		ON TREATMENT
Black Thorax	990217				-
Gammaknife01	000425		Brain Tumor		-
Gammaknife02	040930				-
Cyberknife02	550301		Brain Tumor		-
Cyberknife01	550300				-

Figure 5: Patient Worklist from a Screenshot Display of the DICOM-RT based ePR system showing the ability of a patient search engine. In this case, a wildcard "*" search was performed displaying all the patients. Note, the two anonymized Cyberknife patients listed at the bottom.

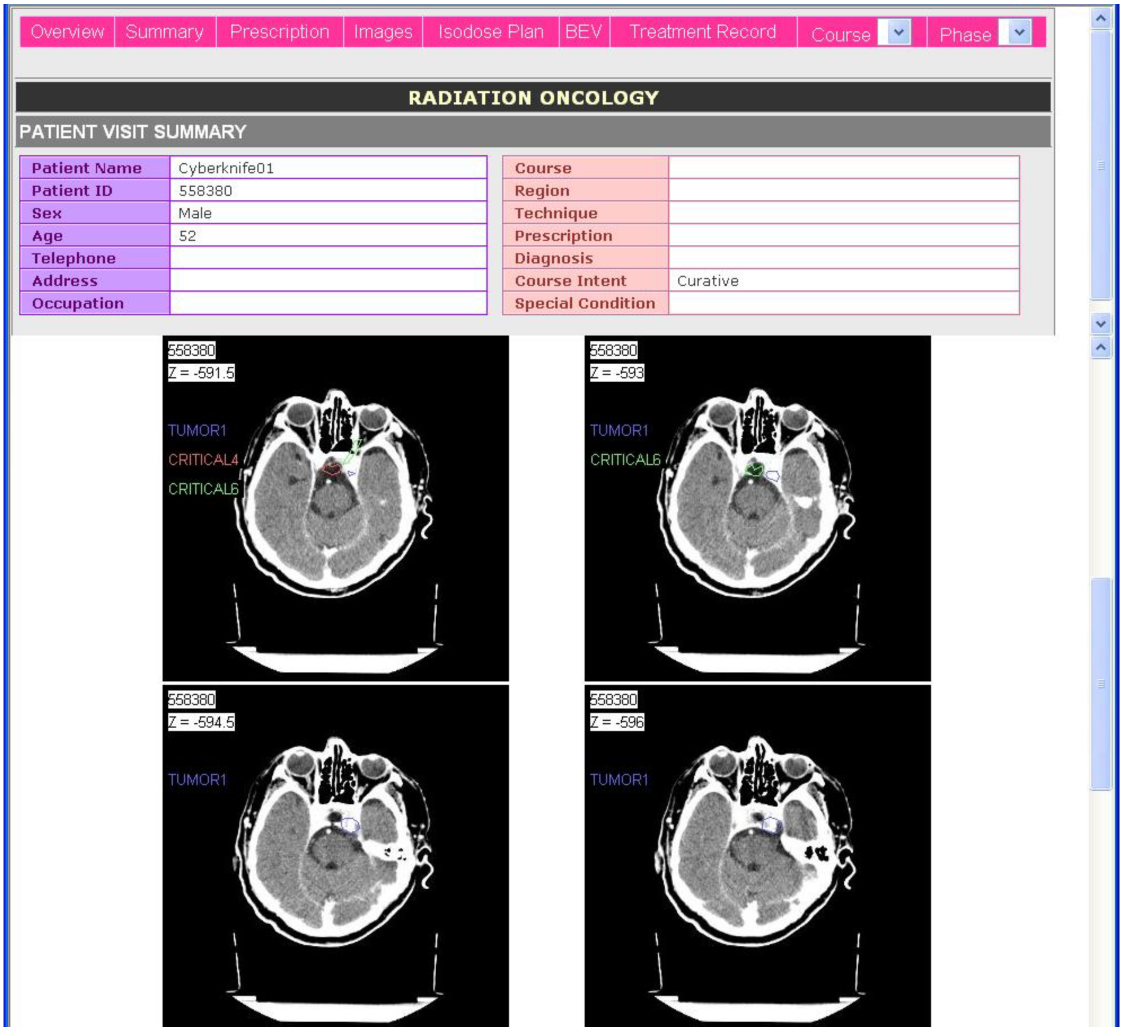


Figure 6: Display Window showing CT diagnostic images overlayed with critical structures obtained from the Cyberknife treatment system. The contours enclosing the tumor (TUMOR1, blue), the optic chiasm (CRITICAL4, red), and the left nerve (CRITICAL6, green) all happen to be displayed on this particular upper lefthand CT image slice.

Figure 7 is a timeline overview display showing that a CT and MR diagnostic exam was acquired. Since this particular case was an initial survey, some of the pertinent clinical data was missing, hence an incomplete patient overview. However, it is shown that clinical data from the Cyberknife RT system can be integrated within the DICOM-RT based ePR system. Referring to Figure 7 (bottom), a conventional RT management information system only has the DICOM RT records but no DICOM RT plan, RT images, and DICOM images. On the other hand, the DICOM-RT based ePR system is able to display information extracted from all of the DICOM-RT objects and can be expanded for more detailed views from the icons on the timeline in the User Interface (both bottom & upper sections).

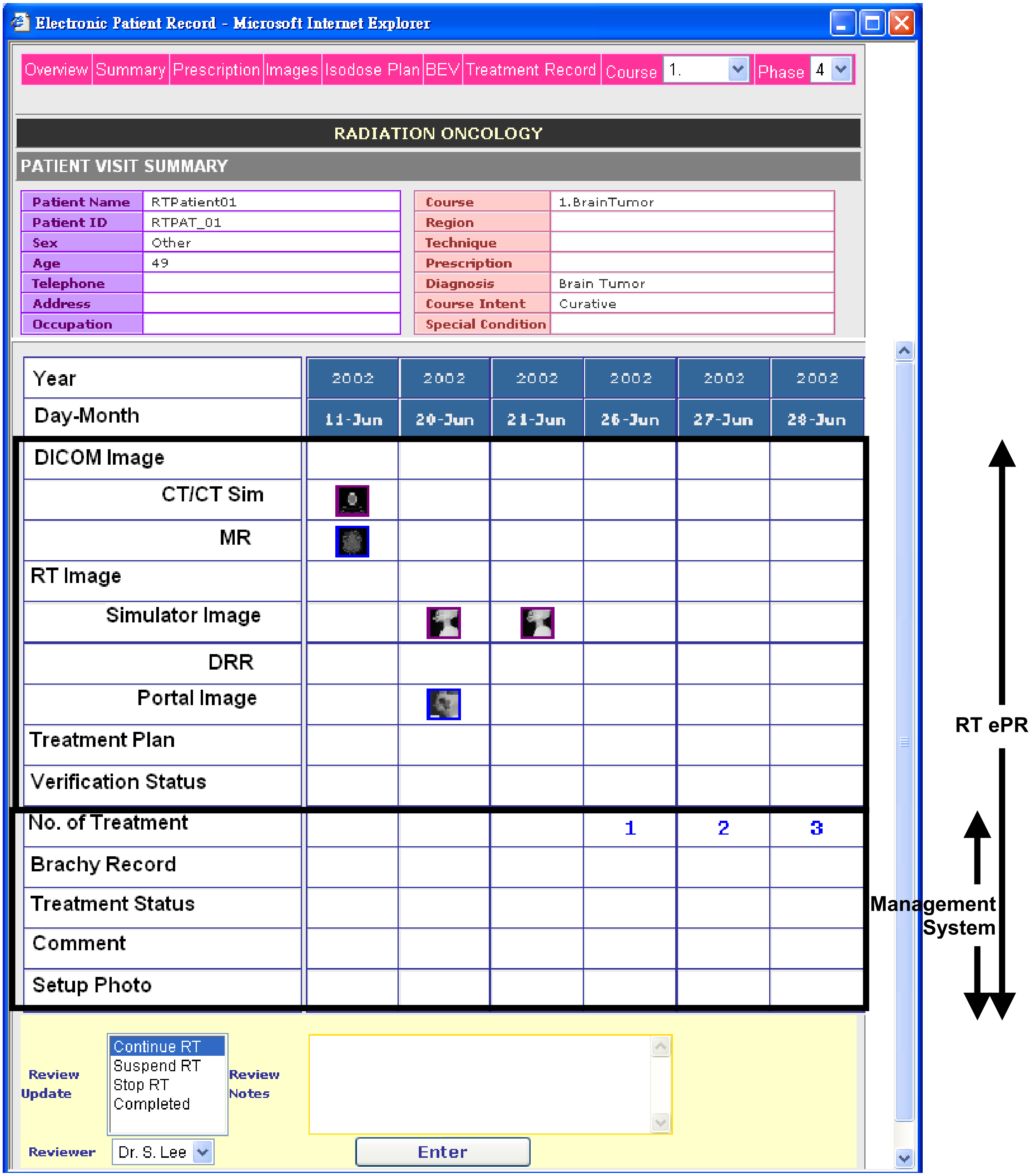


Figure 7: Timeline overview display of a patient in the DICOM-RT based ePR system. The RT ePR system has a richer database than the conventional RT information/management system. A given RT information management system has only the DICOM RT records (Bottom of figure), while the RT ePR is able to display the information extracted from all the DICOM objects including the DICOM RT plan, RT images, and DICOM images. They can be expanded from the icons on the timeline in the User Interface (Both bottom and upper sections of figure).

Finally, Figures 8 and 9 are a comparison of data from the Cyberknife Treatment Planning System (TPS) that was reconstructed and displayed to show how the results are identical to that of the screenshot results of the same TPS in Figure 9. The differences being that data in Figure 8 are converted to DICOM-RT objects that can be further distributed to other clinical areas and DICOM-compliant clinical systems while the data in the screenshot in Figure 9 are proprietary within the TPS and difficult to distribute throughout the healthcare enterprise.

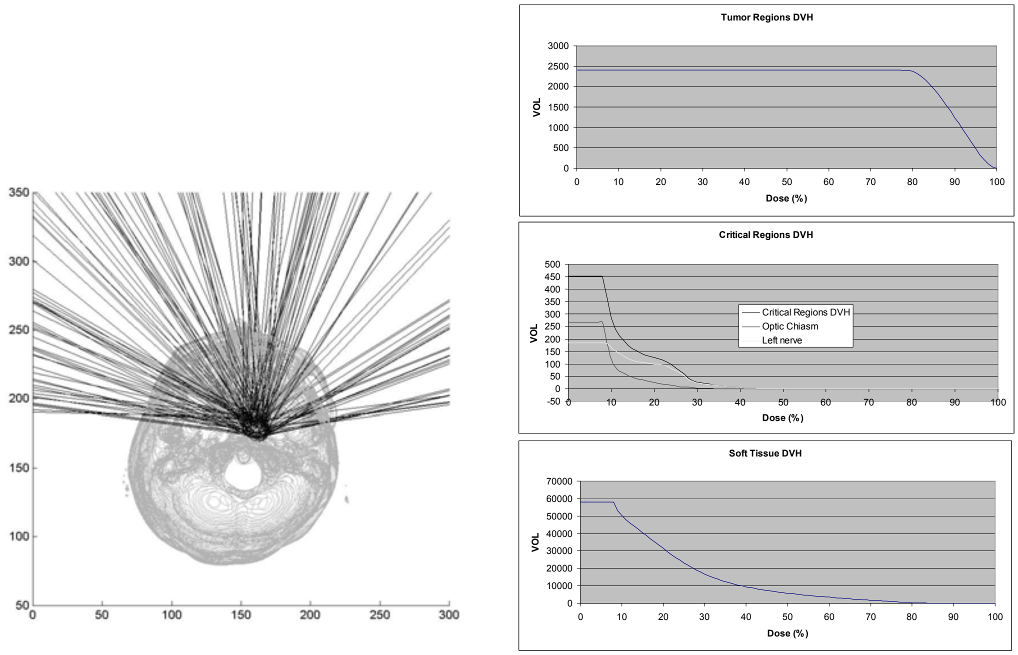


Figure 8: Reconstructed Data Obtained from the Cyberknife Treatment Planning System showing the DVH curves and the different radiation beams planned for dose delivery to the brain tumor.

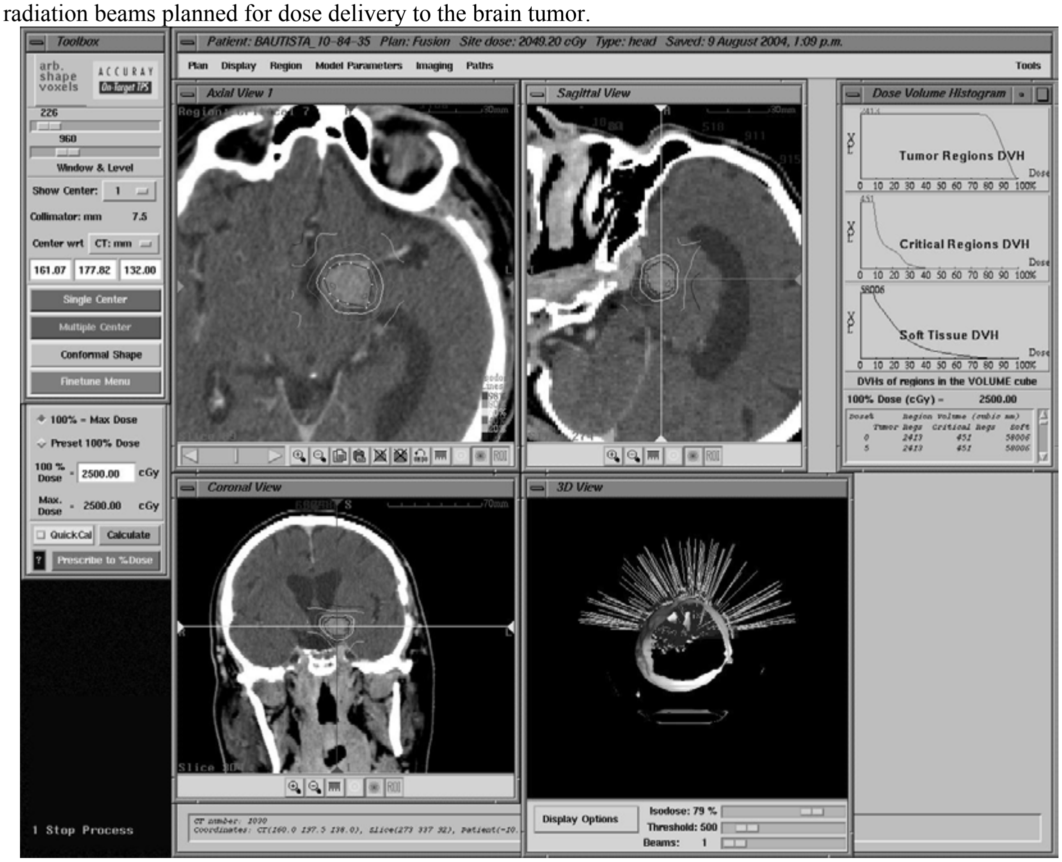


Figure 9: Screenshot from the Cyberknife Treatment Planning System. Note that the DVH curves and the radiation beams are confirmed to be identical to Figure 8.

4. CONCLUSION

A DICOM-RT based ePR system for managing patients with brain tumor cases was designed and developed within the Radiation Oncology Department at USC Health Sciences Campus. Data obtained from a brain tumor case treated by the Cyberknife system was collected and integrated within the ePR system as an initial first step. By implementing this DICOM-RT based ePR system, both clinical image and related informatics data are integrated into a one-stop source of pertinent clinical information necessary for making treatment decisions within the RT department and throughout the healthcare enterprise. The richness of the clinical data available was shown in comparison to standard RT information management systems. The initial results show the confirmation that the conversion of DICOM-RT objects are correct and can be displayed similar to what is displayed within the proprietary system. Preliminary surveys from users within the Radiation Oncology Department reveal that the ePR system is beneficial to their clinical workflow. Future work will include a more robust evaluation methodology to investigate the improvements of the clinical workflow especially at the four points-of-decisions described in the paper.

REFERENCES

1. Cyberknife, http://www.accuray.com/ck/local16.htm

2. Huang HK. 1999. PACS: Basic Principles and Applications. Wiley & Sons, NY, p. 521.

3. Palta J.R., Frouhar V.A., Dempsey J.F. Web-based Submission, Archive, and Review of Radiotherapy Data for Clinical Quality Assurance: A New Paradigm. In J Radiat Oncol Biol Phys, 2003; 57(5): 1427-1436.

4. Digital Imaging and Communications in Medicine (DICOM), Supplement 11: Radiotherapy Objects, 1997.

5. Digital Imaging and Communications in Medicine (DICOM), Supplement 29: Radiotherapy Treatment Records and Radiotherapy Media Extensions, 1999.

6. DICOM Standard 2003, http://medical.nema.org/dicom/2003.html.

7. USC/Norris:Cyberknife, http://ccnt.hsc.usc.edu/services/cyberknife/information.htm.

8. Law M.Y.Y., Huang H.K., Zhang X., Zhang J. The Data Model of a PACS-Based DICOM Radiation Therapy Server. Proceedings of the SPIE on CD-ROM, Medical Imaging 2003, 128-129, San Diego, CA, U.S.A.

9. Liu B.J., Huang H.K., Cao F., Zhou M.Z., Zhang J.Z., Mogel G., A Complete Continuous-Availability PACS Archive Server, Radiographics 2004; 24:1203-1209.

10. Liu B.J., Cao F., Zhou M.Z., Mogel G., Documet L. Trends in PACS Image Storage and Archive, Comput Med Imaging Graph 2003, 27:165-174.

The Data Storage Grid: The Next Generation of Fault-Tolerant Storage for Backup and Disaster Recovery of Clinical Images

Nelson E. King[a] , Brent Liu[a*], Zheng Zhou[a], Jorge Documet[a] , H.K. Huang [a]
Image Processing & Informatics Laboratory, Department of Radiology, University of Southern California, Marina del Rey, 90292

ABSTRACT

Grid Computing represents the latest and most exciting technology to evolve from the familiar realm of parallel, peer-to-peer and client-server models that can address the problem of fault-tolerant storage for backup and recovery of clinical images. We have researched and developed a novel Data Grid testbed involving several federated PAC systems based on grid architecture. By integrating a grid computing architecture to the DICOM environment, a failed PACS archive can recover its image data from others in the federation in a timely and seamless fashion. The design reflects the five-layer architecture of grid computing: Fabric, Resource, Connectivity, Collective, and Application Layers. The testbed Data Grid architecture representing three federated PAC systems, the Fault-Tolerant PACS archive server at the Image Processing and Informatics Laboratory, Marina del Rey, the clinical PACS at Saint John's Health Center, Santa Monica, and the clinical PACS at the Healthcare Consultation Center II, USC Health Science Campus, will be presented. The successful demonstration of the Data Grid in the testbed will provide an understanding of the Data Grid concept in clinical image data backup as well as establishment of benchmarks for performance from future grid technology improvements and serve as a road map for expanded research into large enterprise and federation level data grids to guarantee 99.999 % up time.

Keywords: Grid Architecture, Data Grid, Grid Computing, Fault-Tolerance, PACS

1. INTRODUCTION

Clinical image storage technology has matured significantly during the past ten years but many implementations still lack cost-effective backup and recovery solutions. For example, a large PACS vendor as recently as 2003 could not send a second image to a different storage device forcing film to be printed as the backup copy [1]. There are numerous backup strategies as summarized by Liu et al [2]. These include second and third backup copies often to tape, continuous availability servers, and the application service provider model. Some of these approaches are detailed in case studies in the book "PACS and Imaging Informatics [3]. Solutions to backup are evolving in step with the volatile archiving technology [4]. For example, tape as a medium for long-term storage is being challenged by low-cost disk arrays.

Nevertheless current solutions are expensive, difficult to administer, and often time consuming to effect recovery after a disaster. Many large-scale clinical image archive systems, such as PACS, still encounter downtime for hours or days, which have the critical effect of crippling daily clinical operations. According to the Disaster Recovery Guide, http://www.disaster-recovery-guide.com/, not all disasters are environmental such as earthquakes or floods. There are also organized and / or deliberate disruption, loss of utilities and services, equipment or system failure, serious information security incidents, and other emergency situations. Any of these disasters could make the images in a PACS unavailable. Radiologists unable to view comparison studies cannot make a diagnosis on a newly acquired exam which severely impacts the clinical workflow. PACS is taking an increasingly enterprise role such as image-assisted surgery and oncological treatments or non-invasive procedures. Therefore, backup and availability of historical image data is taking on increasing importance.

*brentliu@usce.edu; phone 1 310 448-9435; fax 1 310 444-9441; www.ipilab.org

Data grids evolving from Grid Computing represent the latest and most exciting technology to emerge from the familiar realm of parallel, peer-to-peer and client-server models, which is ideal to address this problem of downtime. We have researched and developed a novel Data Grid testbed that consists of several federated PACS integrated by adata grid architecture.

Grid computing is increasingly popular in medical research. Computing grids are being developed to assist research in biomedicine and bioinformatics. Several efforts underway in Europe include HealthGrid and Grid-Enabled Medical Simulation Services (GEMSS) Project. In the US, there is the Biomedical Information Research Network (BIRN) based at the San Diego Supercomputer Center whose data centered around brain imaging of human neurological disorders and associated animal models.

Grids currently under development for clinical applications are predominantly for digital mammography. The University of Pennsylvania Consortium developed the National Digital Mammography Archive Grid (NDMA). A small company i3Archive has commercialized the NDMA technology in the United States. eDiamond Grid will become a federated database of mammograms shared by numerous breast screen centres in the UK [5]. The Mammogrid aims to develop a pan-European mammography database.

Our focus is on data grids for clinical operations that depend on a 24/7 availability of PACS and the associated storage archives. Grids for bioinformatics access large amounts of data but medical research is not mission-critical. Similarly, grids for digital mammography focus on the archival and retrieval of a large dataset but rarely is image availability mission-critical. The emphasis on backup and recovery in clinical operations drove us to an architecture that could support heterogeneous platforms across geographic distances and DICOM compliance which means no added software to clinical equipment.

2. MATERIALS AND METHODS

Our grid concept involves the integration of a grid computing architecture to the DICOM environment so that a federation of clinical organizations can create a DICOM data storage grid. A failed PACS can then recover its image data from other archives in the federation in a timely and seamless fashion. The grid makes the specific hardware, protocols and physical location transparent to the end user. A DICOM client only needs to know the DICOM AE Title of the grid.

2.1 Data Grid Architecture

The grid means many things according to a survey by Gillett et al [6]. Our grid is anchored on an open set of standards and protocols called the Open Grid Services Architecture (OGSA) [7]. The design of our grid reflects the five-layer architecture of grid computing: Fabric, Resource, Connectivity, Collective, and Application Layers [8] that is used in the Open Grid Service Architecture (OGSA). These layers are described below in the context of our target clinical implementation built upon services in the open source Globus Toolkit 3.2 (GT3) created at the Argonne National Laboratory, and ISI (Information Science Institute) of the University of Southern California [9].

1. Fabric Layer: This is the lowest layer and includes the physical devices or resources (e.g., computers, storage systems, and networks). Our architecture consists of a DICOM compliant fault-tolerant (FT) backup Archive Server and a PACS simulator at the research laboratory (IPI: Image Processing and Informatics); two SANs (Storage Area Network) at two PACS clinical sites; and a communications network system including Gigabit LAN (local area network), Internet 2 and broadband WAN (wide area network - T1).
2. Connectivity Layer: This layer above the fabric layer includes the communication and authentication protocols required for Grid network transactions (e.g., exchange of data between resources and verification of the identity of users and resources). Some of the applicable GT3 services include Grid Security (GSI), high-performance secure data transfer (GridFTP).

3. Resource Layer: This layer contains connectivity protocols to enable the secure initiation, resource monitoring, and control of resource-sharing operations. Some of the GT3 services are remote job submission and control (GRAM) and Replica Location Service (RLS).
4. Collective Layer: This layer above the Resource Layer contains protocols, services, and APIs (Application Programming Interface) to implement transactions among resources (e.g., resource discovery, and job scheduling). This layer consists of services that interact between the User Applications and the services in the Core Middleware, such as database service (to find the best available database in the Data Grid), information service (to monitor the current active services in the Data Grid), and data service (to find the physical address of the logical data) as well as other services.
5. User Application layer: This is the highest layer and calls on all other layers for resources. This layer might obtain necessary authentication credentials (e.g., set up connection), query a replica catalog to determine the status of a fabric resource or the location of a file, submit requests (e.g., via DICOM), or monitor the progress of a transfer.

The connectivity and resource layer are often joined together and called "core middleware" as shown in Figure 1. The integration of DICOM with the data grid is shown below in a simplified view of the grid architecture showing a combined connectivity and resource layer. DICOM services are integrated at both the core middleware and collective layer or user level middleware. DICOM SCU (service class user), SCP (service class provider), and Query/Retrieve are key components integrated within the five layer architecture.

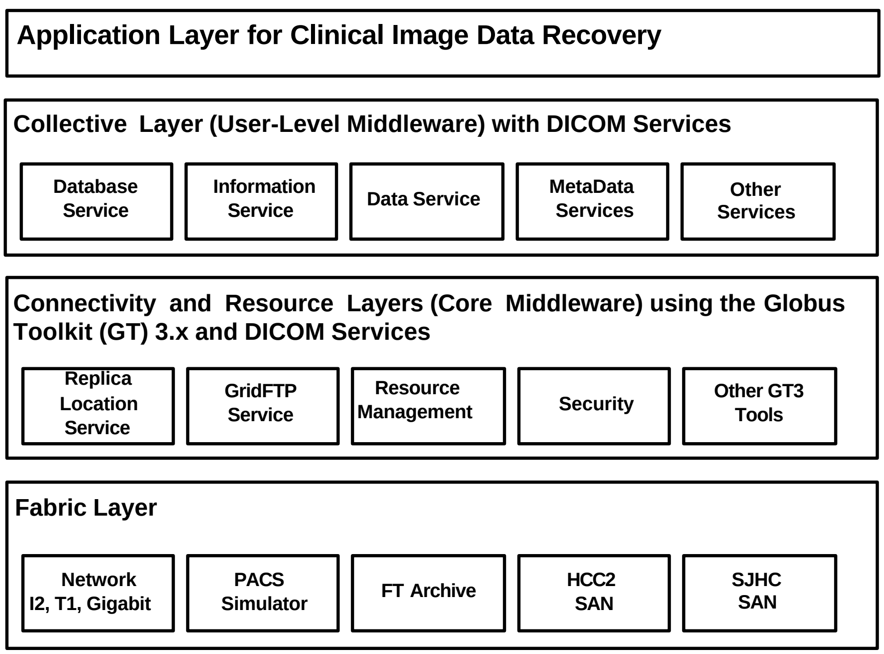


Figure 1 Layers of Grid Architecture with DICOM Services

2.2 Clinical Implementation of the Data Grid

Our data grid takes advantage of rapid decreases in storage hardware and the incremental cost of adding storage to a SAN. When a facility already operates a SAN, additional storage can currently be purchased for less than $10,000 a terabyte [10]. Storage devices can be physically isolated (e.g., 3U rack unit) or the SAN management tools can dedicate a partition to the Data Grid of the federation. This physical and logical isolation ensures that clinical data at the hosting site won't be mixed with data sent from federation members.

Geographical separation of alternate storage devices within a data grid is a benefit for several reasons. Data is protected from disasters that cause localized damage to a facility such as a plane crash. However, there are disasters that may not directly destroy the data center but nevertheless curtail its capability. For example, construction outside the hospital can cut a power cable. Generators can provide temporary power but high power consuming devices like disk-intensive storage are often the first systems to be powered down when power is limited. The PACS would then lose access to the images. However, a data grid would still allow complete access to images provided that the reading workstation and network switches still have power. Our implementation of a data grid would provide both backup of clinical images but also operational flexibility during recovery events.

The clinical implementation planned for this data grid consists of three sites as shown in Figure 2. The federation will consist of one research laboratory and two clinical PACS sites, Saint John's Health Center, Santa Monica, and the Healthcare Consultation Center II, USC Health Sciences Campus [2]. The first site is the IPI Lab where the major resources are the PACS Simulator and the DICOM Fault-Tolerant backup Archive. The latter serves as IPI's archive in the Data Grid. The second and the third sites are the Saint John's Health Center, Santa Monica, CA (SJHC) and the Healthcare Consultation Center II (HCCII) at the University of Southern California, Health Sciences Campus. These clinical sites already have previous research collaborations with IPI Laboratory facilitating their participation in the initial clinical evaluation. In addition, both sites have a clinical PACS with a SAN archive system although different storage hardware.

The three-site configuration of this Data Grid design example is shown in Figure 2. A partition of each SAN which is separate from the partition used to store clinical PACS image data will be used as backup archive resources in the Data Grid ("P2"). These storage resources are shown in gray on Figure 2 and found in the fabric layer of figure 1. This data grid architecture allows any DICOM resource such as a quality control (modality) workstation, reading workstation or clinical workstation outside of the Data Grid to access services within the Data Grid. The communication from the workstation to the Data Grid are shown as grey double-headed arrows.

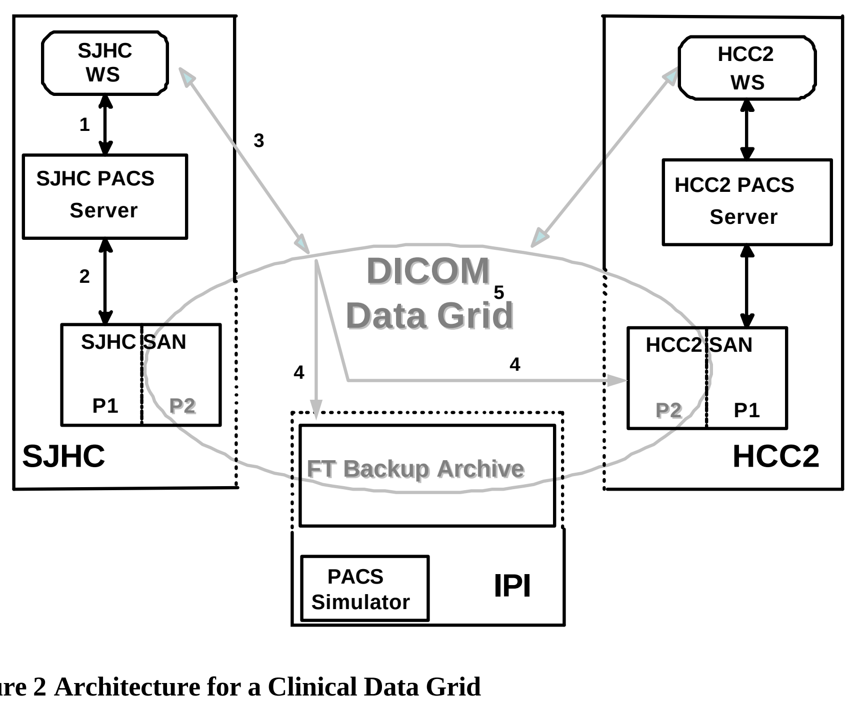


Figure 2 Architecture for a Clinical Data Grid

Our initial Data Grid implementation design uses three sites as a matter of convenience. Three sites are sufficient to demonstrate the principles of data grid technology. The grid architecture allows for unlimited expansion to additional sites. The inherent limitation of a data grid is the number of federation partners that can work with each other.

2.3 Backup of Clinical Images in the Data Grid

Clinical images generated in the three sites, IPI, SJHC, and HCCII are backed up by the Data Grid using the following protocol. The objective is to have two backup copies of image data acquired from any site in the Data Grid. In addition, there are several databases in the Data Grid to track the header information of the images and the physical storage location of both images. The numbers next to arrows in Figure 2 illustrate the backup policy from SJHC to the Data Grid in more detail as described by Liu et al [2].

1. After an examination is completed at the SJHC CT scanner, the modality workstation sends out two copies of the image data, one to the SJHC clinical PACS Server, and the second to the Data Grid.
2. The SJHC PACS Server receives the first copy of the image data and stores it in the SJHC SAN (P1) as its own clinical archive.
3. The Data Grid activates a resource to receive the second copy of the image. The Data Grid monitors the status of the resource and activates a second one if the first resource fails.
4. After receiving the image data, the resource automatically sends two copies, one to the IPI FT Backup Archive and the second to the HCCII SAN (P2). Since the Data Grid distributes redundant copies to two different storage sites, a single-point-of failure is avoided. After the image data has been successfully archived in the two storage sites, the physical location is logged within the Data Grid and physical storage information of each copy of image data is kept.
5. Patient information of the image data (header information but not the image itself) to a DICOM Data Model database in the Data Grid. The database is replicated and distributed to the FT backup Server at IPI, the SJHC SAN (P2), and the HCC SAN (P2). Three copies of the database are used to avoid a single-point-of-failure.

2.4 Image Recovery from the Data Grid

The scenario is that SJHC PACS archive goes down during a disaster and the physician from a SJHC workstation needs to query/retrieve the backup image data from the Data Grid immediately for review. The following is the design of a service-driven configuration and workflow for the image data recovery procedure as specified by Liu et al [2].

<u>Step 1: Find the Patient Information from the most optimal database copy in the Data Grid</u>

1) A DICOM Q/R SCU (Service Class User) in a SJHC workstation sends a Query/Retrieve request to the Data Grid.
2) The Data Grid receives this request and finds from the copies of the database in the grid, which holds the desired patient record.
3) The Data Grid surveys the status (e.g., availability and latency) of all the database resources (copy of the database at HCCII SAN (P2), copy of the database at IPI, and copy of the database in SJHC SAN (P2) within the Fabric Layer) and determines the most optimal selected database location. The selected database gets the patient information and returns to the DICOM Q/R SCU in the Application Layer. This step is necessary because the user will have to select studies and images from the patient information directory, which is a standard DICOM Q/R operation from the PACS workstation.

<u>Step 2: Retrieving Image Data from the Data Grid</u>

4) After the user selects a set of studies from the patient directory, the DICOM Q/R SCU in the Application Layer requests the Data Grid to retrieve the patient image data from the most optimal image data database where this image data resides in terms of the best latency and efficiency.

5) Because there are two copies in the two storage sites (HCCII SAN (P2) and IPI FT Backup Archive) in the Data Grid, it determines which replica is the best choice to perform the retrieval based on network connection. It is noted that this optimal location may be different from the previous optimal location for patient information where the patient information is located because a very large amount of image data will be transferred in this latter operation. Network availability, file sizes and types will be major determined factors.

6) The Data Grid selects the storage site with the best response (e.g., IPI Archive) and initiates a DICOM move client to request the DICOM server to transmit the file from the specified storage site to the SJHC workstation. Note that Step 2 is performed completely within the Data Grid and is seamless to the user.

3. RESULTS

A Data Grid Testbed was developed in the IPI Laboratory as an initial step. This Data Grid testbed mimics the implementation design model described in sections 2.3 and 2.4. The goal for the testbed was to implement a five layer grid architecture integrated with DICOM services to achieve the following:

- Show connectivity of the Data Grid by connecting resources (e.g., DICOM archives) across multiple networks to three sites
- Connect a clinical site to the Data Grid via multiple resources (e.g., Grid Access Points)
- Implement three copies of the database in the Fabric Layer to avoid a single-point-of-failure,
- Represent the IPI Archive, as well as the HCCII SAN and SJHC SAN partitions as storage resources for the Data Grid
- Develop the services in the Collective Layer, and modify the Applications such as DICOM query/retrieve, client/SCU and a DICOM storage client.

Our testbed currently simulates only one clinical site attached to the Data Grid due to hardware resource constraints. Figure 3 overlays the physical components of the testbed onto the clinical model of Figure 2 to graphically illustrate the parallelism of the testbed to the clinical setting.

The grid resources on Figure 3, shown in gray background with white text, are labeled "Grid Access Point", DICOM Store" and "DICOM DB". There is a modality simulator and reading workstation to represent one site. Three simulated DICOM storage archives were configured to represent the storage fabric resources located at "SJHC SAN P2", "FT Backup Archive" and "HCC2 P2". These are called DICOM storage resources which physically resided on Sun Ultra II servers. In addition to storage resources, the databases storing the image meta-data (e.g., DICOM header information) are shown as "DICOM DB". The transactions with the grid take place with a grid access point (GAP). The testbed used two GAP's for fault tolerance. The actual clinical implementation will include multiple GAP's at each of the sites to provide fault tolerance. Each GAP in the testbed was physically hosted on a PC running the Linux operating system. Actual clinical images with anonymized demographic patient data were used in testing the Data Grid. These images were from Computed Radiography, Computed Tomography, Magnetic Resonance, and Ultrasound modalities.

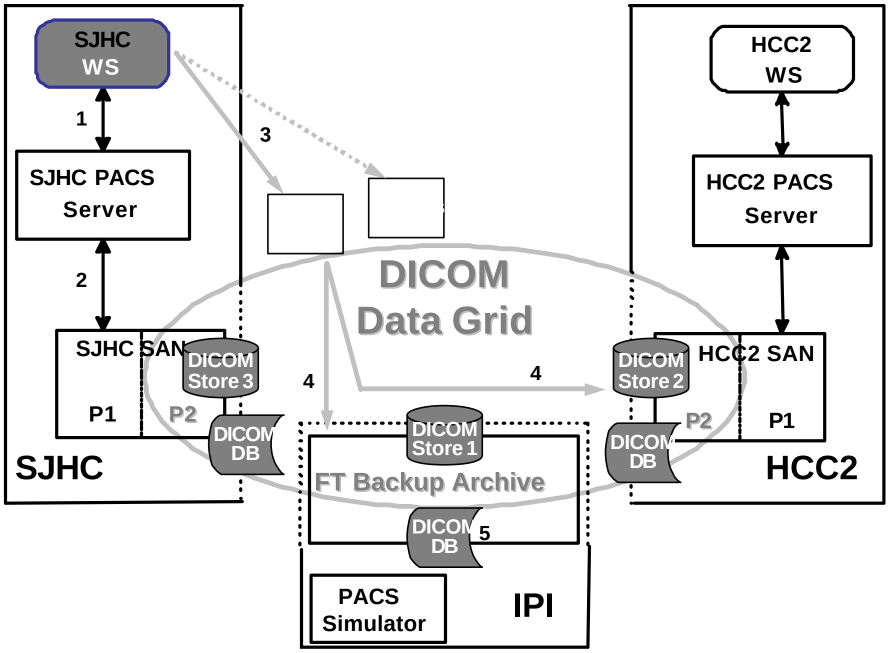


Figure 3 Data Grid Testbed Configuration

Section 3.1 Testbed Backup Demonstration

The testbed implements the clinical backup policy of Section 2.3 using the same steps previously described.

Step 1 - Modality simulator representing SJHC sends an image to the PACS Server.

Step 2 – PACS Server stores the primary image to the clinical partition (P1).

Step 3 – Modality simulator sends image to the Data Grid

The Data Grid receives images from the modality simulator via the grid access point. Grid Access Point 2 is the primary connection shown in Figure 3 that receives the images from the modality. If there were problems, failover could occur to Grid Access Point 1.

Step 4 – Data Grid sends two copies to Storage Resources

Grid Access Point 2 simultaneously sends a copy to Grid Storage Resource 1 and 2. These storage resources are intended to represent geographical separation from the storage resource at SJHC labeled “Grid Store 3”. Storage Resources 1 and 2 were placed on different sub-networks to represent delays from passing through network devices (e.g., firewall). If there is a problem with a storage resource, the Data Grid buffers the data until the connection is re-established.

Step 5 – Data Grid updates the DICOM databases

The Data Grid processes the image headers and updates the DICOM databases for the Data Grid.

Steps 1 to 5 demonstrate the basic functionality of the Data Grid. We added additional tests to demonstrate the fault-tolerance and recovery capabilities of the Data Grid.

"Pull the plug" test demonstrated fault-tolerance and failover. The RJ-45 connector to Grid Storage Resource 1 is removed while images are being sent. Transmission immediately stops. Once the connector is reinserted and a network connection established, the sending of images should resumes. The same test is done by "pulling the plug" on Grid Storage Resource 2 during image transmission. A DICOM application on a workstation is then used to retrieve the images to verify that all the images were received.

Alternate Grid Access Point test demonstrates the fault-tolerant backup to the Data Grid continues by pulling the plug on Grid Access Point 1. The modality then switches to Grid Access Point 2. Grid Access Point 2 simultaneously sends a copy to Grid Storage Resource 1 and 2. The fault-tolerance of the Data Grid using Grid Access Point 2 is then tested. The sequence of tests in Step 1are repeated.

There are combinations of test scenarios that can be exercised to demonstrate the operation of sending a study from the modality to two storage devices. The first factor is the condition of the grid access point. GAP1 may be primary so the failover is to GAP2. The second factor is the condition of the storage resource. The resource may be connected and operational during the entire time the study is being sent. The resource may not be available when first requested by the grid access point but a few seconds later may once again be connected. The third condition of the resource is that the connection is lost during the middle of a transmission. In this case, the study should resume sending. The single modality sending studies to two grid access points results in 36 potential tests. We tested about 1/3 of the conditions as shown in Table 1.

	Grid Access Point Condition			
Storage Status	GAP1 Primary	Failover to GAP2	GAP2 Primary	Failover to GAP1
Storage 1				
Available at start	Sent	Sent	Not tested	Not tested
Not available	Started on recovery	Not tested	Not tested	Started on recovery
Interrupted during send	Resumed after recovery	Resumed after recovery	Resumed after recovery	Not tested
Storage 2				
Available at start	Sent	Not tested	Not tested	Sent
Not available	Not tested	Not tested	Started on recovery	Not tested
Interrupted during send	Not tested	Resumed after recovery	Not tested	Not tested
Storage 3				
Available at start	Not tested	Not tested	Sent	Not tested
Not available	Not tested	Not tested	Not tested	Not tested
Interrupted during send	Not tested	Not tested	Not tested	Not tested

Table 1 Test Scenarios for DICOM Send to Data Grid

Section 3.2 Testbed Recovery Demonstration

The ability to retrieve images during an outage of the local storage archive is the next series of tests. The procedure was discussed in Section 2.4. The testbed hardware is identical to the one used for backup in Figure 3. This sequence of test illustrates the utility of the Data Grid when recovery of images is necessary from alternate storage resources.

Step 1 – Grid Storage Resource 1 is unavailable – Grid Access Point 1 is primary connection

The workstation with a DICOM client wishes to retrieve a study but the local PACS is unavailable. Using DICOM client software (e.g., Cedara Iview), a query is made to the Data Grid for a particular study. A DICOM connection is established with the grid access point so that the results of the query (e.g., search) can be returned to the client. Typically the local storage resource has the best response time. If the plug is pulled on Grid Storage Resource 1 then the Data Grid routes the query to see if the study is on Grid Storage Resource 2. The results are then returned to the client.

The next test is to retrieve the image. A retrieve request is made to Grid Access Point 1 which routes the request to the Storage Resource that provides the best response. A DICOM move is then executed by the storage resource sending the image directly to the DICOM client. The storage resource then returns a message to the client that the images are being sent. The images can then be checked once they are received.

Step 2 – Grid Storage Resource 1 and Grid Access Point 1 are not available

The same sequence as step 1 occurs. Using a DICOM client, a query is made to the Data Grid for a particular study. A DICOM connection is established so that the results of the query (e.g., search) can be returned to the client. However, failover to Grid Access Point 2 takes place which now processes the query and return of search results followed by the send.

Again, there are a large number of combinations to be tested. We examined a few conditions as shown in Table 2.

	Grid Access Point Condition			
Storage Status	GAP1 Primary	Failover to GAP2	GAP2 Primary	Failover to GAP1
Storage 1				
Available at start	Query returned: C-Move made	Query returned: C-Move made	Not tested	Query returned: C-Move made
Not available	Failover to Storage 2	Failover to Storage 2	Not tested	Failover to Storage 2
Storage 2				
Available at start	Not tested	Not tested	Query returned: C-Move made	Not tested
Not available	Not tested	Not tested	Failover to Storage 1	Not tested
Storage 3				
Available at start	No images stored	No images stored	No images stored	No images stored
Not available	No images stored	No images stored	No images stored	No images stored

Table 2 Test Scenarios for DICOM Query/Retrieve to Data Grid

4. DISCUSSION

Data collected from this initial Data Grid testbed demonstrates that a data grid architecture provides fault-tolerant backup and a capability for recovery during disasters. We chose to use DICOM services as much as possible so that there is no need for modification of the clinical devices which already use DICOM. These clinical devices are those that would like to store and access image data from the Data Grid. The insights gained from this version of the testbed are guiding the development of the next version. We discuss some of these insights below.

Our implementation of C-Store in the DICOM Data Grid provides a buffering mechanism so that any storage resource that is inactive or has the connection interrupted will receive all the images once the connection is restored. No image data is lost once sent into the Data Grid. Our next version will need to incorporate the IHE modality performed procedure protocol so that status information on the sending of the backup copies will be available as well.

The Data Grid also strengthens the DICOM Query command. Without the Data Grid a query attempt would fail if the requested DICOM database is unavailable or busy. The Data Grid now passes the query to the best performing DICOM database and returns the search results to the requesting workstation.

The Data Grid does not have as great of a benefit to the retrieval of an image to the workstation using a DICOM C-Move. The Data Grid gives a reliable routing of the request to the best performing storage resource. However, C-Move directly sends the images to the requesting workstation bypassing the Data Grid. This means the Data Grid cannot help in finding the best connection path back to the requesting workstation. This is an area for further research.

Recovery of a network connection after an interruption takes many seconds depending upon the operating system. The unpredictable reconnection time has implications for our future testbed. First, periodically monitoring the health of the grid components is insufficient for a clinical operation. While the health of the grid needs to be known, speedy transfer of images cannot rely upon a status measured every 5 to 10 seconds. Each DICOM service that is requested should initiate a status inquiry to the most likely resources that will be utilized. For example, the local storage device and the next most likely storage device should be checked at the time of a request. However, checking on the status of other resources can be deferred until a failed status condition occurs. Second, the Data Grid must be able to buffer the images in process. The delays in transmission or temporary failure of a device cannot be predicted. Buffering of the images sent by the modality allowed our testbed to recover from the times we pulled the plug and resume sending images.

5. CONCLUSIONS

The successful demonstration of the Data Grid in the testbed provided an understanding of the performance issues in the Data Grid concept in clinical image data backup as well as establishment of benchmarks for performance from future grid technology improvements. In addition, the testbed can serve as a road map for expanded research into large enterprise and federation level data grids to guarantee CA (Continuous Availability, 99.999 % up time) in a variety of medical data archiving, retrieval, and distribution scenarios.

REFERENCES

1. G. Wiley, Decisions in Imaging Economics **11**, 41 (2003).
2. Brent J. Liu, M.Z. Zhou, and J. Documet, Computerized Medical Imaging and Graphics **in press** (2005).
3. H.K. Huang, *PACS and Imaging Informatics: Basic Principles and Applications*, 2 ed. (Wiley-Liss, 2004).
4. Herman Oosterwijk, *PACS Fundamentals*. (OTech Inc., 2004).
5. M. Brady, D. Gavaghan, A. Simpson et al., in *Grid Computing*, edited by Fran Berman, Geoffrey Fox, and Anthony J.G. Hey (John Wiley & Sons, Hoboken, NJ, 2003), pp. 923.
6. Frank E. Gillett, Richard Fichera, Galen Schreck et al., 2004.
7. I. Foster, C. Kesselman, and S. Tuecke, International J. Supercomputer Applications, (15), 3 (2001).
8. Ian Foster, Physics Today **55** (2), 42 (2002).
9. Ian Foster, Carl Kesselman, Jeffrey M. Nick et al., presented at the Open Grid Service Infrastructure WG, Global Grid Forum, 2002 (unpublished).
10. Logan G. Harbaugh, in *InfoWorld* (2005), pp. 32.

Implementation of an ASP model offsite backup archive for clinical images utilizing Internet2

Brent J. Liu[a], Sander SD. Chao[b], Jorge Documet[a], Jasper Lee[b], Michael Lee[b], Ian Topic[b], Lanita Williams[b]

[a]Image Processing & Informatics Laboratory, Department of Radiology, University of Southern California, Marina del Rey, 90292

[b]Department of Biomedical Engineering, Viterbi School of Engineering, University of Southern California

ABSTRACT

With the development of PACS technology and an increasing demand by medical facilities to become filmless, there is a need for a fast and efficient method of providing data backup for disaster recovery and downtime scenarios. At the Image Processing Informatics Lab (IPI), an ASP Backup Archive was developed using a fault-tolerant server with a T1 connection to serve the PACS at the St. John's Health Center (SJHC) Santa Monica, California. The ASP archive server has been in clinical operation for more than 18 months, and its performance was presented at this SPIE Conference last year. This paper extends the ASP Backup Archive to serve the PACS at the USC Healthcare Consultation Center II (HCC2) utilizing an Internet2 connection. HCC2 is a new outpatient facility that recently opened in April 2004. The Internet2 connectivity between USC's HCC2 and IPI has been established for over one year. There are two novelties of the current ASP model: 1) Use of Internet2 for daily clinical operation, and 2) Modifying the existing backup archive to handle two sites in the ASP model.

This paper presents the evaluation of the ASP Backup Archive based on the following two criteria: 1) Reliability and performance of the Internet2 connection between HCC2 and IPI using DICOM image transfer in a clinical environment, and 2) Ability of the ASP Fault-Tolerant backup archive to support two separate clinical PACS sites simultaneously. The performances of using T1 and Internet2 at the two different sites are also compared.

Keywords: PACS, ASP, Archive Backup, Internet2, DICOM

1. INTRODUCTION

With the development of PACS technology and an increasing demand by medical facilities to become filmless, there is a need for a fast and efficient method for providing data backup for disaster recovery and downtime scenarios. Based upon previous work performed at the St. John's Health Center (SJHC) using a fault-tolerant backup archive at the Image Processing Informatics Lab (IPI) [1,2] with a T1 connection, an extended ASP Backup Archive was implemented for the USC Healthcare Consultation Center II (HCC2) utilizing an Internet2 (I2) connection. [3]

Once implemented, the backup archive system offers fast and ease-of-use DICOM query/retrieve from HCC2 when the primary archive onsite is unavailable. For some PACS sites, there is no off site backup copy and the images are vulnerable to a catastrophic loss. PACS sites that do not use an offsite facility may store a second copy of the images in close proximity to the original, leaving both copies potentially vulnerable to loss. The simplest solution is to store the second copy in an offsite storage vault. However, storage in a vault does not allow quick retrieval of the data during a disaster since the data media must be retrieved from the offsite vault and a replacement archive must be in place before the data can be loaded. For this reason, the PACS site would be unable to resume operation within a reasonable amount of time. In addition, historical images are necessary in order for a Radiologist to read new cases which would be unavailable if the primary archive site suffers a downtime.

The need for an automated backup system is to allow facilities, such as HCC2, to go filmless and yet maintain their records digitally. The loss of a digital archive can be catastrophic to a Radiology department similar to losing an entire film library. Having up-to-date clinical images stored off site is vital to a filmless environment. However, one of the greatest technical roadblocks in implementing an automated offsite backup is finding a fast broadband connection capable of transferring images in a reasonable amount of time. The IPI Laboratory has previously established Internet2 connectivity, which is a federal government initiative setup in 1996 to replace the current Internet with one using a higher speed backbone for many applications including transfer of medical imaging data. At this time, Internet2, with over 200 research universities and more than 100 non-academic institutions connected to the Internet2 backbone, has transmission speeds up to 10Gbits/s in its backbone and is emerging as a great candidate for high speed image data transmission solution [4].

From previous experience, IPI Laboratory has been a pioneer in testing and developing Internet2 for clinical applications [5], but this research project provides the additional challenge of using the Internet2 for daily operational tasks and being added into an existing ASP solution. Figure 1 depicts the Internet2 connectivity between the HCC2 facility and the IPI Laboratory.

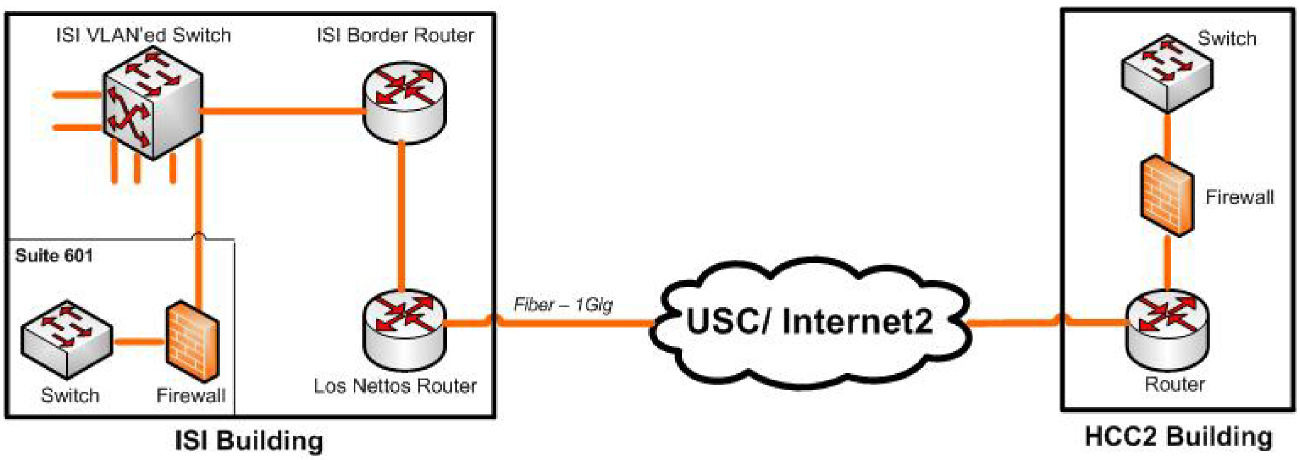


Figure 1: Connectivity between IPI Laboratory and HCC2

Because the offsite backup archive was already storing clinical images from SJHC, careful strategy and planning is needed to store a second site like HCC2's clinical images with minimal disruption to the normal clinical workflow of SJHC. Utilizing the backup server for two sites simultaneously presented a challenge in separating each clinical site's image data during storage and retrieval. It is a complex process that requires the repartitioning of both the long term storage and the RAID for short-term storage. This problem was difficult to address with the first site still running. This paper describes the implementation of adding a second clinical site with I2 Connectivity.

2. METHODOLOGY

There were 4 main steps that were implemented to ensure the success of the project

1. Clinical Workflow Development
2. Establishing & Evaluating Internet2 connectivity
3. Implementing the new site, HCC2
4. Monitoring and Performance Testing

2.1 Clinical Workflow Development

The initial step was to determine the workflow of the newly installed PACS implemented at HCC2. HCC2 installed a totally digital solution with integrated Hospital Information System (HIS)/Radiology Information System (RIS)/ PACS/ Voice Recognition (VR) in July 2004. The first step in the workflow is at the registration desk where patient information is obtained and an examination is ordered. Next, the technologist selects an ordered exam from a worklist and forwards the patient demographics to a modality workstation. After that, the patient is called in for the examination and immediately the study image quality is checked before being sent to the PACS server and diagnostic workstations. At the diagnostic workstations, a radiologist makes a diagnosis using voice recognition to generate a report which is then sent off as the finalized report.

To implement a fault-tolerant PACS system with an offsite ASP backup service a DICOM gateway must be added, the Internet2 connectivity must be established with proper security actions, such as configuring the firewall and restricting the access to the offsite PACS server, which also must be taken into consideration. Figure 2 describes in detail how this data workflow is presented at HCC2. It assumes a new patient with a single exam ordered.

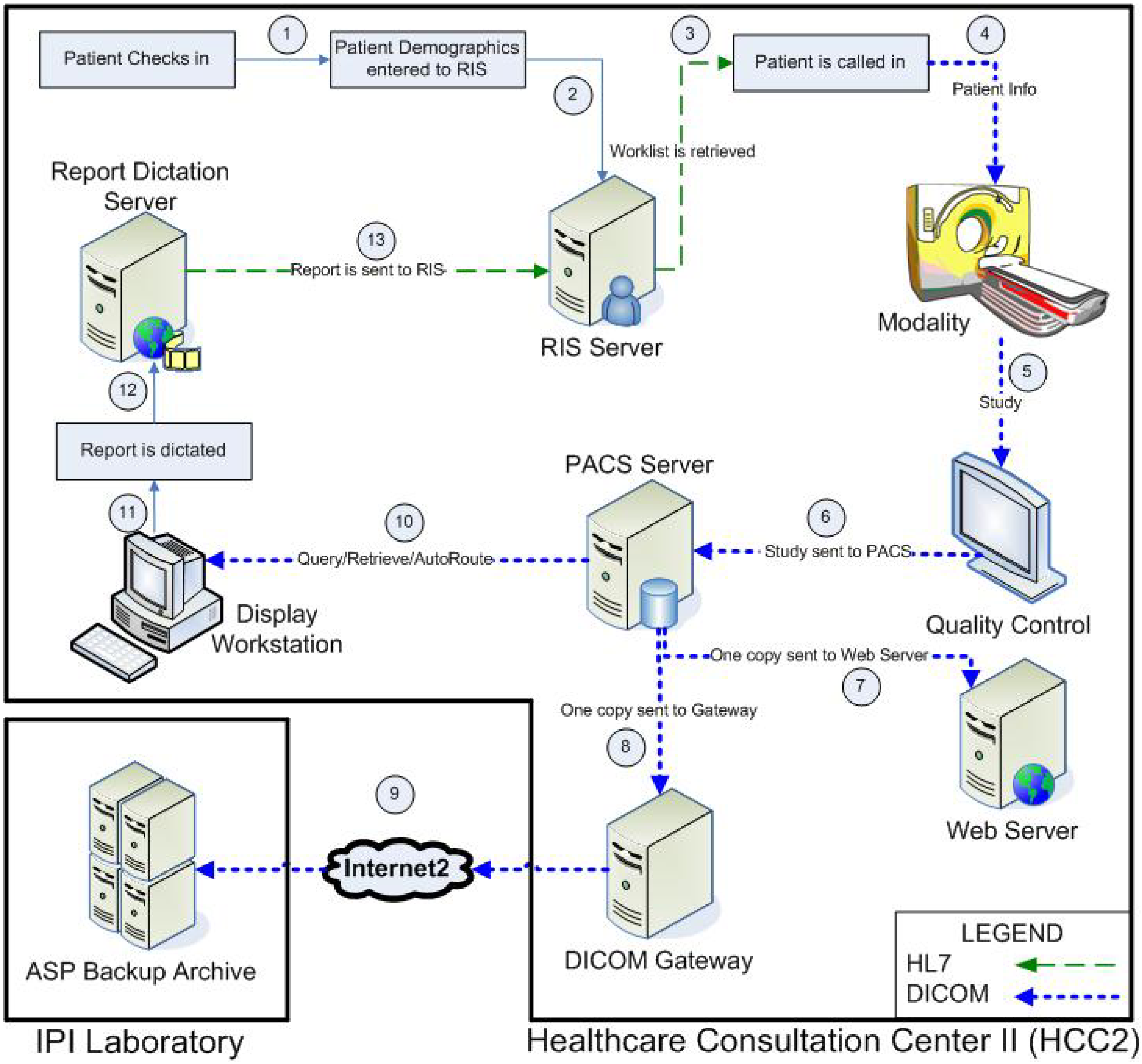


Figure 2: Detailed Data Workflow at HCC2

First, patient information is entered into the RIS. Then an exam is then ordered and sent via HL7 as part of a modality worklist. At this time the patient is called in to perform the examination at the corresponding modality. In step 5 the exam is sent to the Quality Control machine to ensure the study has been acquired correctly. After this, the study is sent to the onsite PACS server as shown is step 6. Once the PACS server receives a new image, it automatically sends a copy to a gateway (step 8) configured to forward images using Internet2 to the ASP backup server at IPI Laboratory (step 9). The image data is first saved in the RAID short term storage and eventually migrated to DLT long term storage. At the same time, another copy is sent to the web server to retrieval using web technology from outside the hospital. In step 10 the study is sent to the Display Workstation where the radiologist performs the diagnostics. Then the report is dictated and is digitally converted into a report by a server called Report Dictation Server, which then sends off the mentioned report back to the RIS Server as it is given in step 13.

2.2 Establishing & Evaluating Internet2 connectivity

An initial survey was performed between USC IT group to determine the connectivity of I2 from the Health Science Campus to the Main Campus. Since the IPI lab has already established an I2 connection, the survey was necessary to determine whether there was I2 Connectivity from the Health Science Campus to the Viterbi School of Engineering at the Main Campus. There was no need to test IPI's connection to the Viterbi School of Engineering because connection was already documented to be established previously. After communication between IPI and HCC2 was confirmed, simulated image data transfer was performed using both FTP and DICOM transfer protocol.

The PACS Simulator software [6] was used to gain familiarity with the operation of a PAC system during the early phase of the project. It was important to know how to use the system before beginning modification of the operational backup service to minimize interruption and gain a better knowledge of the data transfer workflow. The PACS Simulator modeled the different modules of a clinical PACS environment and it was installed on five networked desktop computers at the IPI Laboratory. The components listed in order of workflow are: the RIS (Radiological Information System), Modality Simulator, Image Acquisition Gateway, PACS Controller and Reading Workstation. Figure 3 below depicts this configuration.

Figure 3: Connection and workflow diagram of the PACS Simulator

The components of the PACS simulator system were used to evaluate DICOM file transfers between HCC2 and IPI. The Modality simulator software was setup on a workstation at HCC2 and an image acquisition gateway was setup on a workstation in the IPI laboratory. Images were then sent from the modality simulator to the image acquisition gateway to determine DICOM transfer capability between the two sites. The locations of the simulator components were reversed, and DICOM image files were transferred to determine if a 2-way communication was functional. Lastly, modality simulator and acquisition gateway software were installed on workstations at HCC2, and the gateway was configured to forward images to a test backup server at IPI. After the software installation and configuration was complete, the entire data flow (Figure 4 below) was tested to determine the functionality of the planned backup server configuration.

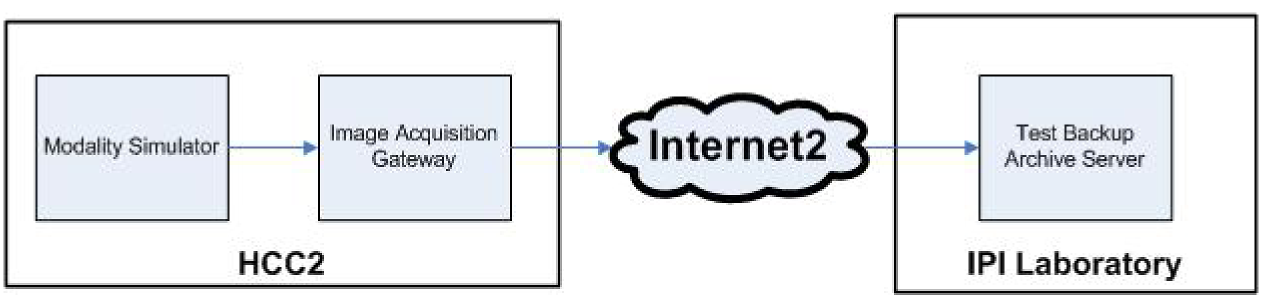


Figure 4: PACS Simulator component configuration testing planned setup

Data transfer speed was measured between the two sites using built in file transfer timers in the simulator software. Since previous T1 performance data between a clinical site (SJHC) and the IPI offsite back-up archive had already been collected, a performance comparison of HCC2 utilizing Internet2 and the clinical site utilizing a T1 connecting to the backup archive will be presented. The transfer data collected will be discussed later on in the results section of this paper.

2.3 Implementation

Reconfiguration of the ASP Backup Archive was necessary before image backup support could be extended to HCC2 because the archive was originally configured to receive data from only one clinical source. The first step was to estimate the amount of incoming data from both clinical sites and determine the storage requirements for amount of time previously accorded with all the parties involved: 90 days for HCC2 and 60 days for SJHC. Priority was given to data from HCC2 because the SJHC data being archived was a third copy of the hospital's image files, while HCC2 had no backup at all. After the storage requirements were determined, the backup server's short term RAID and long term DLT were reconfigured to reflect the new storage setup.

2.3.1 IPI Setup for Two Clinical Sites

The Backup Archive had been in clinical operation to serve SJHC for six months worth of clinical image data backup [7]. The storage volumes for St. John's during this period were partitioned as: 150 GB of RAID for short-term storage and 34 Digital Linear Tapes (DLT) at 35 GB each for long-term storage. SJHC stores close to 7.0 GB per day from their site to the ASP Backup Archive. With the addition of HCC2, it was necessary to reallocate the storage volume space to provide backup for both SJHC and HCC2. Prior to partitioning of the RAID for the two clinical sites, SJHC's data needed to be completely migrated from the RAID onto the long-term storage DLT. Careful planning was performed to determine the best time to repartition the RAID since the backup archive would not be available for SJHC in case a downtime occurred to the primary archive onsite. Based on their clinical workflow, it was determined that the best times for repartitioning should occur starting from late evening Friday and into the weekend. SJHC was notified of the scheduled downtime and SJHC stopped sending new PACS exams until the backup archive was reconfigured.

HCC2 was determined to have a greater need than SJHC for storage space because the ASP backup archive was its only source of backup. Taking into account this condition, the RAID was repartitioned with 110 GB for HCC2 and 40 GB for SJHC. However, because the 150 GB RAID storage was being used to 100% of capacity, space for the new HCC2 partition needed to be freed on the RAID disks before any repartitioning could be performed. The 110 GB of space was freed by changing the RAID migration settings such that all but 40 GB of the most recent data was migrated to the long term DLT tapes. A new RAID volume of 110 GB was then created from the space freed and set-up to receive data from HCC2.

The long term DLT storage at IPI was originally configured to store two copies of the SJHC data into separate tape pools. The second DLT copy of the SJHC data was determined to be redundant because it will be the fourth copy of the information (SJHC has a primary copy and a backup, plus the two at IPI gives four copies). Therefore, the DLT tape pool at IPI containing the second copy of the SJHC data was erased and reallocated to provide long term storage for the information from HCC2. The final amount of long term storage allocated to HCC2 was 10 tapes with a total capacity of 350 GB, which is approximately 3 months of data at a rate of 6 GB a day.

Once reconfiguration of the backup archive was complete, SJHC was notified to restart sending PACS exams to the ASP Backup Archive. After several weeks of monitoring, the data was determined to be storing and migrating as planned without any further modifications. When HCC2 opened a few weeks later, surveillance of the file transfers were performed and the files were observed to be sending correctly to each site's respective storage partition.

2.3.2 HCC2 Setup

A DICOM gateway was implemented based on the clinical workflow in Figure 2. The DICOM gateway acts as a buffer between the HCC2 PACS and the Offsite Backup Archive by reducing the bottleneck of transmission since it has higher bandwidth locally and can queue the PACS exams to be sent across the I2 connection into the Offsite backup archive. This reduces the burden of congestion and send queue overflow at the onsite PACS server and frees it to perform important clinical image distribution locally. Should there be a problem or bottleneck in the connectivity or a problem with the offsite ASP server, the gateway is able to buffer the transmission by managing its send queue to retry without affecting the performance of the onsite PACS archive server. Although the Internet2 connection could support a direct transmission of PACS data from the archive controller, the DICOM gateway is essential to the robustness and security of the clinical PACS system. Clinical workstations at HCC2 also need to Query/Retrieve PACS exam from the offsite backup archive in a disaster recovery or archive downtime situation. However, configuring the clinical workstations for

Query/Retrieve requires close coordination between the PACS vendor, hospital system administrator and ASP backup administrator.

2.4 Monitoring and Performance Testing

Finally, the HCC2 PACS was integrated with the ASP Backup Archive, where image data is temporarily stored for a period of 90 days at the IPI laboratory facility. System monitoring and maintenance support protocols were designed and developed. To maintain control and for maintenance purposes, it was deemed necessary to be able to remotely administer the system even if physical access to the sites were not available. The use of remote desktop technology aided the team greatly in our efforts to retrieve data and troubleshoot errors within the gateway computer during operations.

3. RESULTS

An ASP Backup Archive at IPI lab was successfully implemented for USC HCC2 PACS utilizing Internet2 connectivity. Performance metrics and other observations were gathered during a three month burn-in period. One of the key performance factors is the ability of the ASP Fault-Tolerant backup archive to support two separate clinical PACS sites simultaneously.

Transfer speeds were measured in megabits per second (Mbps) between the gateway computer sending from HCC2 and the ASP backup server receiving at IPI laboratory. The data was collected during normal clinical operation; therefore the speeds measured are actual indicators of how well an Internet2 ASP backup server performs in the real clinical environment. Table 1 and Table 2 below show a comparison of transfer statistics for the four different modalities implemented at HCC2.

	CT Exams	MR Exams	CR Exams	US Exams	Total
No. of Exams	41	55	89	15	200
Total Images	2214	1848	89	704	4855
Avg No. of Images Per Exam	54	34	1	47	136
Avg Size of Exam (MB)	0.51	0.27	6.67	0.35	7.81
Total Data Size (MB)	1145.71	391.116	593.78	258.9	2389.50

Table 1: Statistics of DICOM Images Transferred

	CT Exams	MR Exams	CR Exams	US Exams
No. of Exams	41	55	89	15
Total Transfer Time (Sec)	2450	2765	295	672
Avg Transfer Time (Sec)	59.76	50.27	3.31	42.90
Avg Transfer Speed (Mbits)	5.03	1.41	18.21	4.49

Table 2: Transfer Speed Statistics

These results show that the average CR image transfer speed was 18.21 Mbps with 95% of the transfers occurring in the range of 8.1 to 28.3 Mbps. In terms of time, it will take a standard 6.5 megabyte CR image between 1.8 to 7 seconds to transfer, with an average of 3.31 seconds. For CT it was found that the transfer speed averaged 5.03 Mbps with a range from 1.76 to 9.23 Mbps. The transfer time for an average transfer size of 28.8 megabyte took 59 seconds. Transfer sizes ranged from 0.5 to 81.6 megabytes which corresponds to a transfer time range of 1 to 144 seconds. In the case of MRI the transfer speeds averaged 1.40 Mbps with a range from 0.32 to 3.23 Mbps. The average transfer size was 7.11 megabytes which took 50 seconds. The transfer sizes ranged from 0.24 to 25.5 megabytes corresponding to transfer times ranging from 1 to 130 seconds. Finally, US transfer speeds averaged 4.49 Mbps with a range from 1.81 to 7.2 Mbps. Transfer sizes averaged 18.29 megabytes taking approximately 42.9 seconds to transfer. Overall transfer sizes ranged from 0.9 to 25.7 megabytes which took between 1 to 110 seconds to finish transferring.

The results obtained through DICOM testing consistently demonstrate that the Internet2 connection was vastly superior in terms of file transfer speed to a T1 internet connection. The Internet2 connection had an overall average DICOM transfer speed of 10 Mbps with a high of 28.72 and a low of 0.317 Mbps. FTP transfer speeds attained were even faster than that of the DICOM protocol, because of less processing overhead, yielding figures ranging from 12 to 40 Mbps. As a result, the real world Internet2 speeds recorded were generally much faster than the theoretical 1.5 Megabits per second maximum of a T1 connection. This can be seen from Table 4 where the FTP transfer speed utilizing the T1 Line between SJHC and IPI Laboratory is about 1.39 Mbps only and the DICOM average transfer speed yields to 1.31 Mbps. [1]

	IPI to HCC2	HCC2 to IPI
No. of Transfers	17	17
Total Size of Transfers (MB)	27.77	25.82
Total Time of Transfers (sec)	18.6	7.4
Avg. Transfer Speed (Mbps)	10.00	27.91

Table 3: FTP Transfer Statistics using Internet2 connection.

	Using FTP	Using DICOM
Avg. Transfer Speed (Mbps)	1.39	1.31

Table 4: T1 Line statistics

During DICOM testing, it was observed that the number and size of the files to be transferred significantly affected the data transfer rates attained. Table 5 shows the relationship between average file size and transfer speeds for the different modalities. Generally, the smaller the file size and the greater the number files to transfer will result in a slow transfer speed. MRI, US, and CT image transfers fall into the category of small and many files. As expected, MRI which has the greatest number of files to transfer as well as the smallest file size has the slowest overall transfer speed. In the same fashion, CR resulted to be the faster in transfer speed due to the bigger average size of the files transmitted.

Modality Type	Average Transfer Speed (Mbps)	Average File Size (Megabytes)
CR	18.21	6.67
CT	5.03	0.51
US	4.49	0.35
MR	1.40	0.27

Table 5: DICOM Transfer Statistics

Upon evaluation of our results, it seems that an Internet2 ASP Backup server should be capable of replacing a hospital archive when it comes to transferring CR images. In the event of a hospital image archive failure, transfer speeds of 1.8 to 7 seconds per image should be enough to allow diagnosis to continue at a slower pace. For MRI, CT, and US images, the transfer times (depending upon the total number of files to be transferred) can be much longer than that of CR images, up to 2 minutes. Nonetheless, it should be possible to continue with diagnosis of the images at a slower rate than normal.

4. CONCLUSIONS

While not being able to provide the ultimate goal of transparently replacing a hospital archive, the Internet2 ASP Backup server should be able to provide hospitals with the ability to continue radiological diagnosis in the event of a hospital archive failure. However, the maximum amount of traffic the Internet2 connection can support needs to be determined for a hospital's network infrastructure, in addition to the actual increase in demand of bandwidth in the event of an archive failure.

The implementation of a fault-tolerant PACS system is a significant addition to clinical digital storage as well as disaster recovery. With the emerging appearance of Internet2's high-speed and distant-independent cost of connectivity, clinicians have a heightened level of availability that allows them to gain consistency in their practice, and a much larger scale of resources always ready to be used to them.

REFERENCES

1. B.J. Liu, H.K. Huang, F. Cao, L. Documet, D.A Sarti, *A Fault-Tolerant Back-Up Archive Using an ASP Model for Disaster Recovery*, SPIE Medical Imaging, Vol. 4585-21, February 2002.
2. H.K. Huang, F. Cao, B.J. Liu, J. Zhang, Z. Zhou, A. Tsai, G. Mogel, *Fault-Tolerant PACS Server Design*, SPIE Medical Imaging, Vol 4323-14, February 2001.
3. http://www.internet2.edu
4. H.K. Huang, *PACS and Imaging Informatics, Basic Principles and Applications*, Wiley-LISS, 2004.
5. L.W.C. Chan, R. Li, K. Leung, S.K. Hau, E.C.H. Wong, F. Cao, M.Z. Zhou, *International Internet2 Connectivity Image Workflow for Tele-Imaging Consultation*, SPIE Medical Imaging, February 2003.
6. H.K. Huang, B.J. Liu, F. Cao, M.Z. Zhou, G.T. Mogel, J. Zhuang, X. Zhang, *PACS Simulator: A Standalone Educational Tool, SPIE Medical Imaging*, Vol. 4685-21, February 2002.
7. B.J. Liu, F. Cao, L. Documet, H.K. Huan, J. Muldoon, *Clinical Experiences with an ASP Model Backup Archive for PACS Images*, SPIE Medical Imaging, Vol 5033-10, February 2003.

Digital Signature Embedding (DSE) for Medical Image Integrity in a Data Grid Off-Site Backup Archive

Zheng Zhou, H.K. Huang, B.J. Liu
Image Processing & Informatics (IPI) Laboratory, University of Southern California,
4676 Admiralty Way, Suite1001, Marina del Rey, CA, 90292

ABSTRACT

One of the new trends to protect the PACS image data against disaster situations is to store clinical images at an off-site backup archive. In order to support the mission-critical clinical PACS, the backup archive must be 24/7 continuously available (CA). We have developed a novel Data Grid for this purpose using the grid computing technology. With the federation of several PAC systems in a grid, the Data Grid can provide the true CA (99.999%) backup for the PAC systems. However, image integrity becomes a new critical issue to the Data Grid where the image data are not under the protection of local PACS anymore. In this paper, we presented a digital signature embedding (DSE) method, which can assure image integrity in image transmission or archive. The DSE method permanently embeds the digital signature (DS) of the image in the image pixels using lossless data embedding approaches, which can completely recover the original image whenever desired. The permanently embedded DS in the image would provide the integrity assurance for medical image during its lifetime. The embedding process can be utilized by the local PACS archive server to embed the DS in every image before it is sent to the Data Grid. The embedded DS can then be extracted for verification to ensure image integrity when images arrived in the Data Grid or when images were retrieved back. Therefore, with the DSE method, we have extended our protection of image integrity from local PACS to the backup Data Grid.

Key words: Digital signature embedding, Lossless data embedding, Data grid, PACS, Backup archive

1. INTRODUCTION

Data integrity is an extremely important issue when medical images are archived or transmitted across public networks [1-3]. With current technology and know how, it is not difficult to get access to the network or computers, and to insert artifacts in medical images without detection. As a consequence, image integrity could be compromised. Figure 1(a) shows a magnetic resonance (MR) brain image after its acquisition, and Figure 1(b) shows the same image after it is retrieved from archive, or received by the receiving site (RS) from the sending site (SS). The question arises here: How do we know Figure 1(b) is the same image originally acquired as shown in Figure 1(a) especially considering the size of the artifact in relation to the overall image? With the artificial lesion camouflaged by the gray matter and the white matter in the brain, it requires quite an effort for its detection. Without the arrow pointing toward the artificial lesion in Figure 1(b), one would be hard pressed to identify the artificially inserted lesion even with both images displayed side by side. If the image in Figure 1(b) were utilized for primary diagnosis without the knowledge that the image integrity has been compromised, there could be certain misdiagnosis causing serious clinical consequences.

As a 24/7 mission critical system, picture archiving and communication system (PACS) must have a backup solution, thus the historical images can still be retrieved for clinical diagnosis if the main archive server went down during disasters. A recent new trend for PACS backup is to store images in an off-site backup archive using application service provider (ASP) modal [4]. In order to provide a continuous online support for the main archive, research efforts have been put in developing a fault-tolerant off-site backup archive using the emerging grid computing technology. This type of backup archive is usually called data grid [5,6]. With the federation of several PAC systems in a grid infrastructure, the data grid can provide the true CA (99.999%) image data backup for the PAC systems. However, the image integrity issue may happen in any situation when data is transferred between the clinical PACS and the off-site data grid, or when the data is stored in any computers inside the data grid. How to know the data retrieved back from the data grid is the original data sent to the data grid for backup still remains a critical issue before we can apply the data grid for clinical usages.

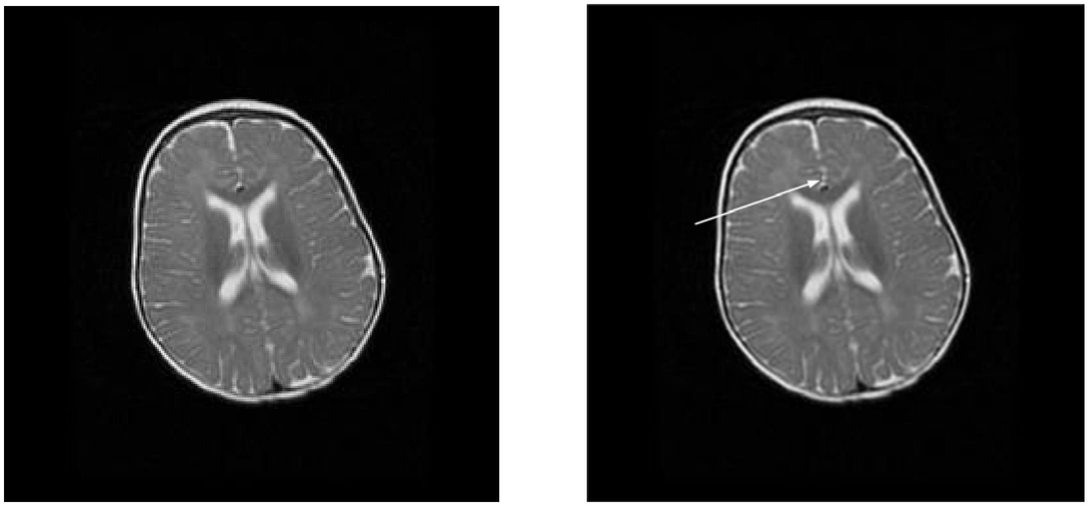


Figure 1(a) a brain MR image; (b) same MR image with an artificial lesion inserted (arrow)

Our previous work in image integrity is a digital envelope (DE) method, which embeds a DE in the least significant bit (LSB) of image pixels located in the dark background area of the image [7,8]. The DE consists of the encryption of a digital signature (DS) and pertinent patient information. The DE method works well for current mammography and sectional images in point-to-point tele-radiology applications. However, it has three drawbacks:

1) In order to embed data in the pixels located in the background, it requires image segmentation preprocess to outline the boundary of the body segment,
2) When there is no body boundary, (e.g., a chest image), the data has to be embedded in the image body which may alter some vital pixel values essential for diagnosis,
3) It needs RS's public key to generate the DE for every individual RS, which makes it very difficult to integrate the DE method with large-scale imaging systems such as the data grid where images may need to be verified in large amount of computers.

For these reasons, we have developed a novel digital signature embedding (DSE) method using lossless data embedding approaches.

2. METHODS

The goal of DSE is to provide a robust integrity assurance to medical images in various application environments and to reserve the original image acquired from imaging modalities. Hence, the DSE was designed to embed the digital signature (DS) permanently in the image pixels and be able to completely recover the original image whenever desired.

The DSE method consists of two opposite processes shown in Figure 2:

1) Sign & Embed operation

a. Generate the DS of the image pixels with the image owner's private key:

$$s = S_{k,priv}(I) \qquad (1)$$

where

s	DS
S	signature signing process
k,priv	owner's private key
I	medical image

b. Embed the bit stream of DS into the image pixels using lossless data embedding approach:

$$I^* = I \oplus s \qquad (2)$$

where
I* signature embedded image
$\oplus$ lossless data embedding process

2) Extract & Verify operation

a. Extract the DS from the signature embedded image and recover the image from the embedding process:

$$(s, I') = \Theta I^* \qquad (3)$$

where
I' the recovered image
Θ data extracting process.

b. Verify the extracted DS with the owner's public key:

$$v = V_{k,pub}(I', s) \qquad (4)$$

where
v verification result
V signature verification process
k,pub owner's public key.

If the verification result is true, the image integrity is assured. Otherwise, the image has been modified.

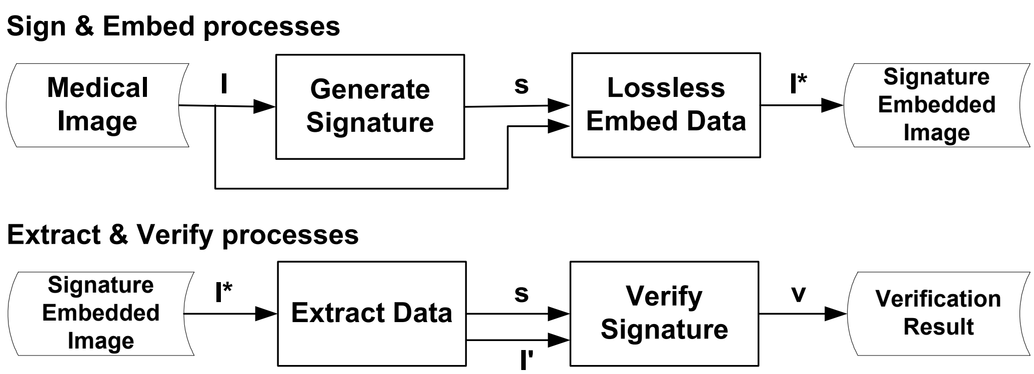


Figure 2 Data flow of Sign & Embed, and Extract & Verify processes in the DSE method. I: Original Image, I*: Signature Embedded Image, I': Recovered image, s: Signature of the Original Image, v: Verification Result.

Based on two different lossless data embedding approaches, we have developed two DSE methods including random pixel DSE method and RS (regular/singular) Groups DSE methods.

2.1 Random Pixel DSE

2.1.1 Algorithm

As shown in Figure 3, the methodology of random pixel DSE method starts with extracting a sequence of LSBs (least significant bits) from randomly selected image pixels. The extracted LSBs form a concatenated bit sequence, which is then lossless compressed. The binary length of the digital signature (DS) determines when the LSB extraction and compression processes ends. When the resulting space defined by subtracting the binary length of the compressed LSBs from the binary length of the original LSBs is no less than the binary length of the DS, the LSB extraction and compression processes are complete. Next, the bit stream of the DS is appended to the bit stream of the compressed LSBs to form a concatenated bit stream. This bit stream then replaces the extracted LSBs in the image pixels. When recovering, the same LSBs in these random pixels are determined and extracted. The DS is first recovered from these LSBs and then the compressed LSBs are decompressed. The decompressed LSBs replace the modified bits, thus restoring the original LSBs in the image pixels. The lossless compression is achieved using adaptive arithmetic coding.

2.1.2 Example

Figure 4 shows an example of embedding the DS in an ultrasound (US) image using random pixel DSE. Assuming that the binary length of the DS is 512 bits and the compression ratio is 10%, we would need at least 5120 random bits to embed the DS. In this case, we started with extracting the LSBs from 6000 randomly selected US image pixels including the dark background. A pseudo-random number generator was utilized to generate these 6000 unique random numbers for selecting the pixels. After lossless compression of these 6000 bits, a compressed bit stream of 5400 bits is obtained. The bit stream of the DS was appended to the compressed bit stream (5400 + 512 = 5912). This newly formed bit stream was then embedded into the original image replacing the first 5912 bits of the extracted 6000 bits to generate the signature-embedded image. Since most LSBs of US image contain only noise due to the characteristics of the modality, the embedded signature will not induce any visual degradation to the signature embedded image. The extraction is the reverse process.

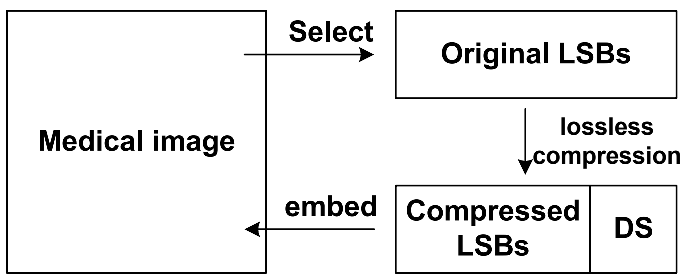


Figure 3 Data flow of the random pixel DSE. After the embedding process, the DS is embedded in the original image. Both the original image and the DS can be recovered.

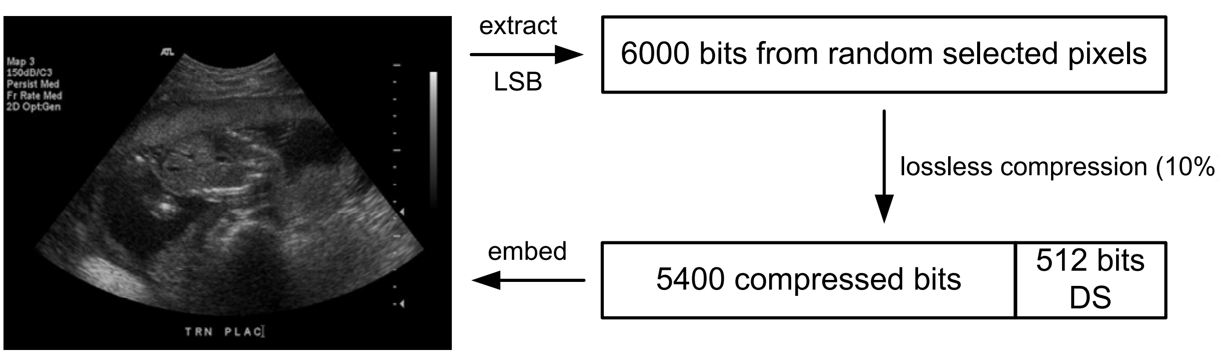


Figure 4 Example using the random pixel DSE to embed the DS in an US image. 6000 random pixels were needed for embedding.

Random pixel DSE adds additional protection to the embedded DS by using a secure random seed. However, the least significant bit of most current images is random noise whose compression ratio is very low, thus it is very difficult to acquire enough space for embedding the DS. In order to increase the compression ratio, one solution is to move up the bit plane from LSB to higher bit planes, which have less random noises. Another solution is to increase the compression ratio utilizing the correlation of adjacent pixels. Fridrich has proposed a novel RS (regular/singular) Groups lossless data embedding approach, which increase the compression ratio of the extracted bit stream by dividing the image pixels into three groups based on the correlation of adjacent pixels [9]. In this research, we developed the RS groups DSE method based on this data embedding approach.

2.2 RS Groups DSE

2.2.1 Algorithm

Consider the original N x M medical image with pixel values in the set P = {0, …, 4095, or higher}. The method starts with dividing the original image into disjoint groups of n adjacent pixels ($x_1, \ldots, x_n$). For example, we can choose n = 4. For this n, we define a discrimination function 'f' that computes the correlation coefficients of each pixel group G = ($x_1, \ldots, x_n$). For example, we can define 'f' as in Equation (5). We also define an invertible operation F on P called "flipping". F has the property that F(F(x)) = x for all x in P. For example, F_{LSB} is as $0 \leftrightarrow 1, 2 \leftrightarrow 3, \ldots, 254 \leftrightarrow 255$.

$$f(x_1, \ldots, x_n) = \sum |x_{i+1} - x_i|, \ 1 \leq i \leq n\text{-}1 \qquad (5)$$

One can see that there would be three situations when we apply the flipping operation F and the discrimination function 'f' to each pixel group G using Equation (6).

$$f(F(G)) \text{ compared to } f(G) \qquad (6)$$

We define these three situations as three groups: R, S and U.

Regular (R) group:	$f(F(G)) > f(G)$
Singular (S) group:	$f(F(G)) < f(G)$
Unusable (U) group:	$f(F(G)) = f(G)$

A new grouped image is formed with three possible states in the selected bit plane.

2.2.2 Embedding Process

At the selected bit plane, by assigning a 1 to R and a 0 to S we can embed one bit of the digital signature in each R or S group. The embedding starts with the scanning of the image pixels to find these R and S groups. Because the adjacent pixels in medical images are usually correlated, we can expect $f(G)$ to be smaller than $f(F(G))$ for most groups G, which would be less correlated after the flipping operation. Therefore, there would be more R groups than S groups, thus more 1s than 0s. With this uneven number of 1 and 0, the lossless compression can be performed more efficiently. The extraction and compression ends till there is enough space to embed the DS. After compression, the bit stream of DS can be appended to the compressed bit stream. The newly formed bit stream is then compared to the original 1/0 bit stream. If there is a difference, the corresponding group of pixels (R/S group) is flipped. After all the bits are compared, the embedding process is complete. In this process, the U groups are not used.

2.2.3 Extracting Process

Since the forming of R, S, and U groups, and the embedding are all reversible processes, we can extract the signature and recover the original image. The extracting process starts with the same scanning to find all the R and S groups from the embedded image. In consequence, we reconstruct the embedded 1/0 bit stream. The extracted bit stream is broken down to the compressed bit stream and the DS. The compressed bit stream can be decompressed to get the original R and S bit stream. To recover the original image, the decompressed R and S bit stream is compared with the extracted R and S bit stream. If any difference, the corresponding group of pixels is flipped. Since the flip operation is a permutation, the original image pixels can be completely recovered.

2.2.4 Examples.

Figure 5 shows an example of using RS Groups DSE method to embed the DS in a CT image.

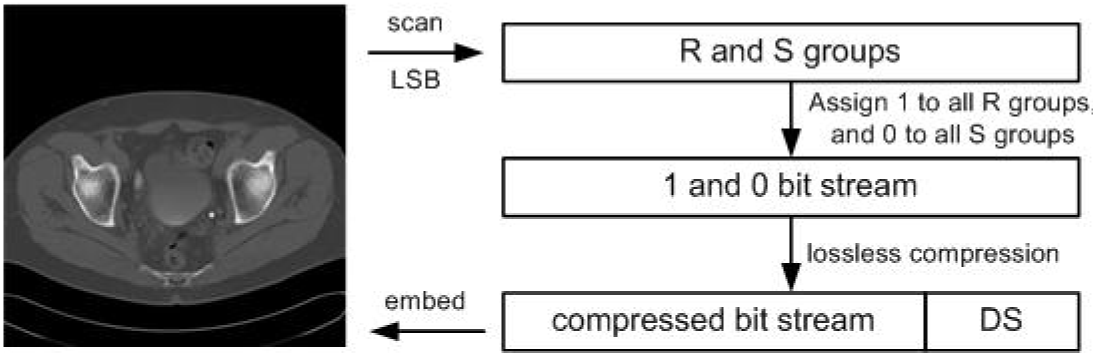


Figure 5 Example of embedding the DS in a CT image using RS Groups DSE method.

2.3 Three Dimensional (3D) DSE

The aforementioned DSE methods are performed on each individual two-dimensional (2D) image. With the advent of multi-detector and volume acquisition technologies, a CT, MR or US exam can generate thousands of images forming a 3D volume. To perform the DSE methods on each individual image in the volume is very time consuming and inefficient. For this reason, we have developed a novel 3D DSE method. As an example, we will use a CT volume with 500 images to illustrate the concept of the method.

2.3.1 Signing

As shown in Figure 6, the method starts with generating a pseudo random sequence with 500 numbers and arranging the pixel data of all the 500 images in a random order based on this random sequence. A hash value is computed for all the pixels in this newly formed volume. The hash value is then encrypted becoming a digital signature (DS) for the entire volume with 500 images. The bit stream is then lossless embedded in this volume using the following approach.

2.3.2 Embedding

3D data embedding is a complex problem. In this paper, we simply extended the 2D RS groups data embedding approach to 3D. However, the 3D method uses a Z shape walking to search the R and S groups in a selected bit plane instead of scanning the images row by row. Take the CT volume as an example, the Z shape walking through the 500 images would be like the Figure 7. It starts with the first four pixels in the first row of the first image, followed by the first four pixels in the first row of the second image until the 500^{th} image. Next, it continues at the second four pixels in the first row of the first image and so on. For all the found R and S groups, we convert them to 1/0 bit stream and lossless compress it. The walking and compressing processes end until there is enough space to embed the bit stream of the DS. Then, the compressed bit stream and the DS are embedded using the same flipping operation in 2D RS Groups DSE. After embedding, the volume data is re-arranged to the original order.

2.3.3 Extracting and Verifying

When extracting and verifying, the volume images are first arranged in the same random order in the signing and embedding processes. The same walking is performed to extract the embedded data. The extracted bit stream is used to recover the original volume data and the original DS. The DS is then verified with the recovered volume data. Finally, the successfully verified volume data is re-arranged to the original order for clinical usages.

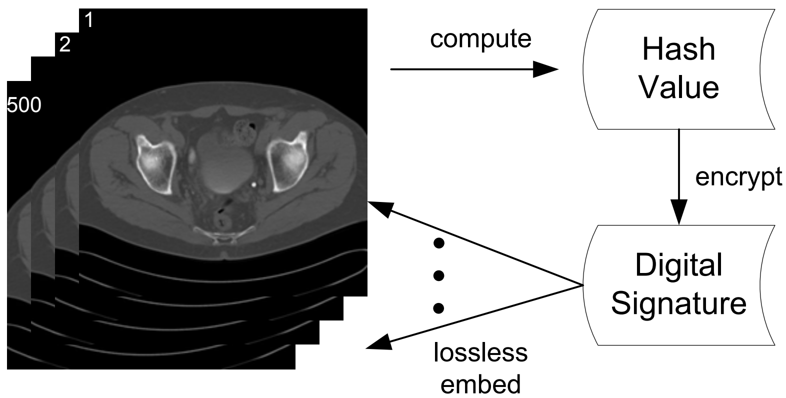


Figure 6 3D digital signature embedding (3D DSE). A single image digital signature is generated for the 3D volume to assure the integrity of the volume.

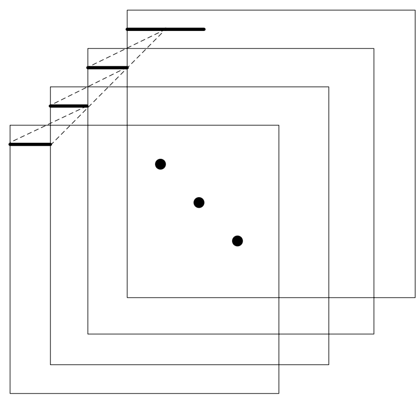


Fig. 7 Z-shape walking searches the R and S groups for data embedding.

3. INTEGRATION OF THE DSE WITH THE DATA GRID

3.1 Data Grid

Data grid utilizes the grid infrastructure to federate physically distributed storage devices into a fault-tolerant shared virtual storage device. Figure 8 shows an example of a data grid with federation of three sites. The first site is the PACS controller and FT backup archive located at the IPI Lab at ISI/USC. The second site is a partition of the SAN located at Saint John's Health Center. The third site is a partition of the SAN at Healthcare Consultation Center 2 at USC/HSC. A clinical workstation outside of the data grid is able to send images to the data grid for backup without the knowledge where images are actually stored.

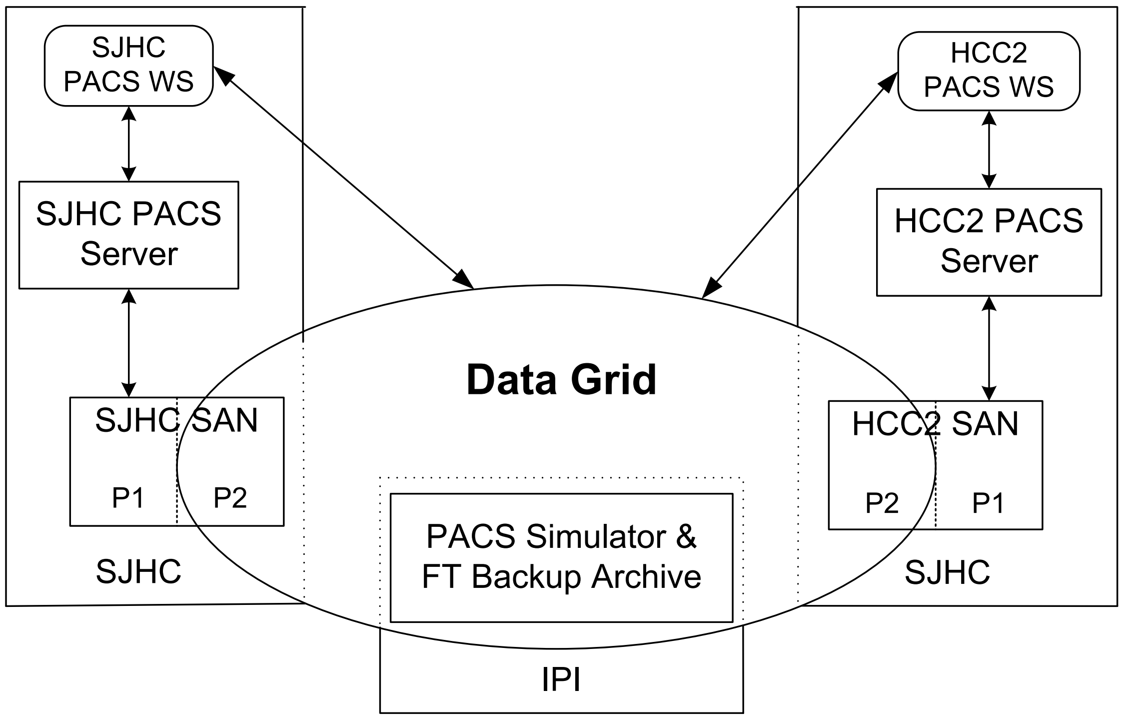


Figure 8 Example of the data grid infrastructure with the federation of three sites.

The image integrity problem can happen when images are transmitted between any two sites through public networks such as T1 or Internet2. It can also happen when images are archived in any site where images might be altered because of security holes. The DSE method can be utilized in the data grid to ensure the image retrieved back from the data grid is the original image that sends out to the grid. We have studied how to integrate the DSE with the data grid.

3.2 Integration of the DSE with the data grid

Figure 9 shows the integration of the DSE with the data grid. Take SJHC as an example, the DSE Sign & Embed processes is installed in the SJHC PACS WS (workstations). The Extract & Verify processes can be installed in both the data grid and the WS. Every image will be signed and embedded before it is sent to the Data Grid. After the image is archived in the data grid, the Extract & Verify processes can be called by any component in the grid to verify the signature of the image if desired. When the image is retrieved back in the WS, the image must be verified to ensure the image has not been altered. Therefore, the image integrity is assured when images are retrieved back from the data grid.

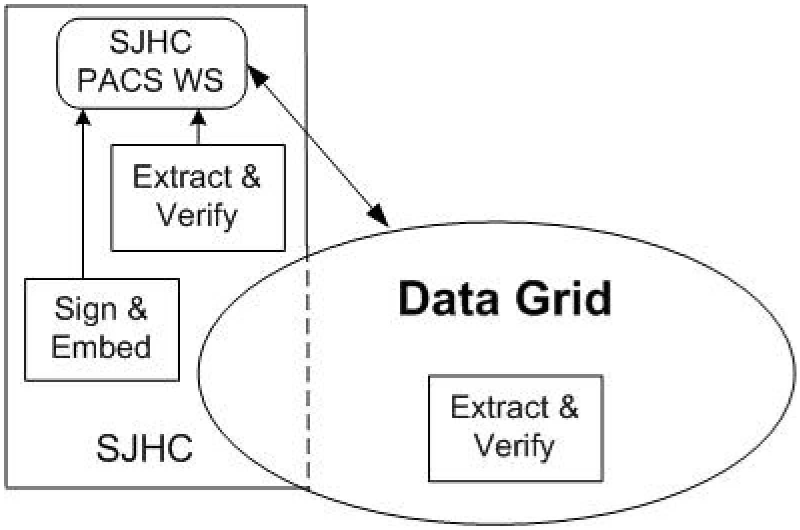


Figure 9 Integration of the DSE with the data grid.

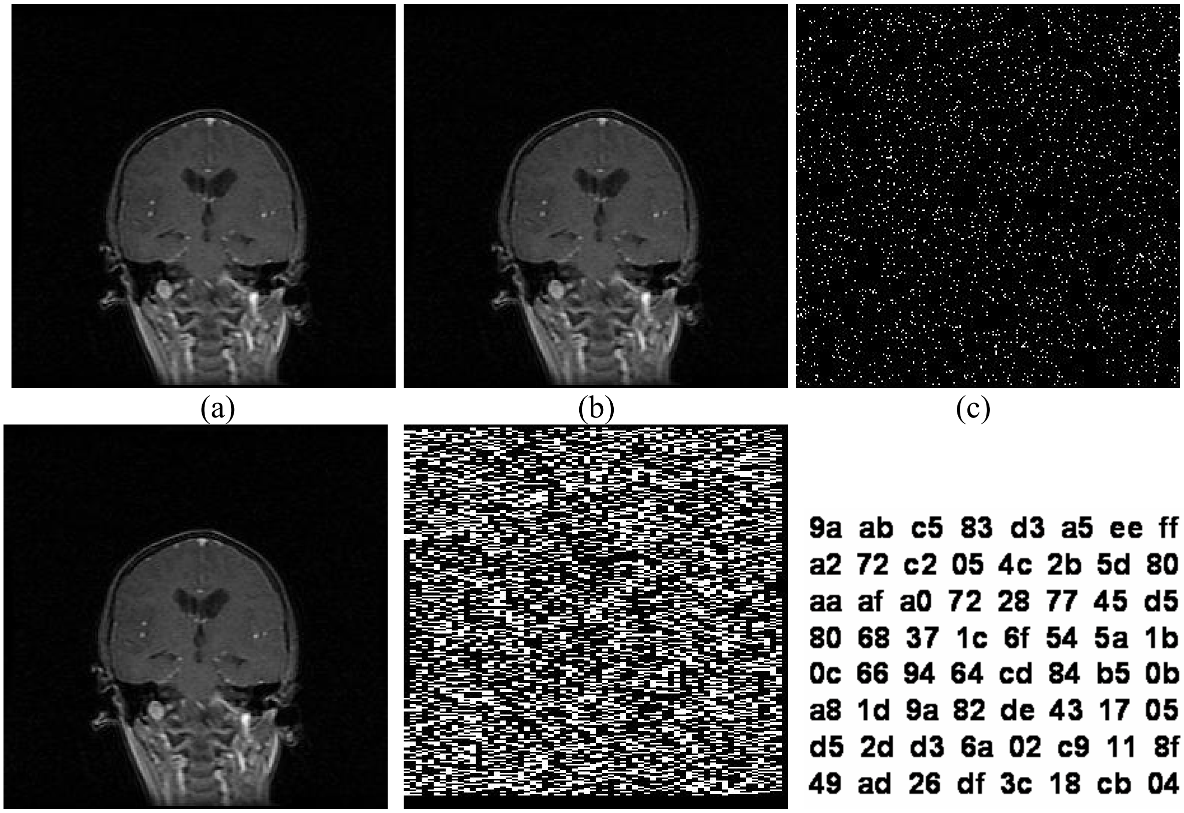
4. RESULTS AND DISCUSSION

We have tested the 2D DSE methods with all prevailing medical images including computed radiography (CR), CT, MR, and US. Following (Figure 10 and 11) are two examples obtained from the experiments.

Figure 10 shows the results of a MR brain image. Comparing the signature embedded images Figure 10(b) and 10(d) with the original image Figure 10(a), we found there is no visual quality degradation in the signature embedded images.

From the subtracted images Figure 10(c) and 10(e), we can see two types of patterns of the embedded data: random dots and block shape. The random dots pattern is from the random pixel DSE, which demonstrates that the embedded bits are randomly distributed in the entire image. The block shape pattern is from the RS Groups DSE. Because each 1 or 0 bit is corresponding to a group of four adjacent pixels, any one bit embedded would actually change four adjacent bits. Therefore, when we embed the bit stream row by row, the result of the embedded bits appears in a block shape. Obviously, more pixels will be changed during the data embedding process in the RS Groups DSE than in the random pixel DSE. For example, the total pixel changed in Figure 10(e) is 21840, whereas to 2550 in Figure 10(c). Divided by the total number of the original MR image pixels (65536), the % pixel changed is 33.3% and 3.9% respectively.

The digital signature (DS) embedded is a 64 bytes binary value. Figure 10(f) shows an example of the DS of the MR image 10(a).

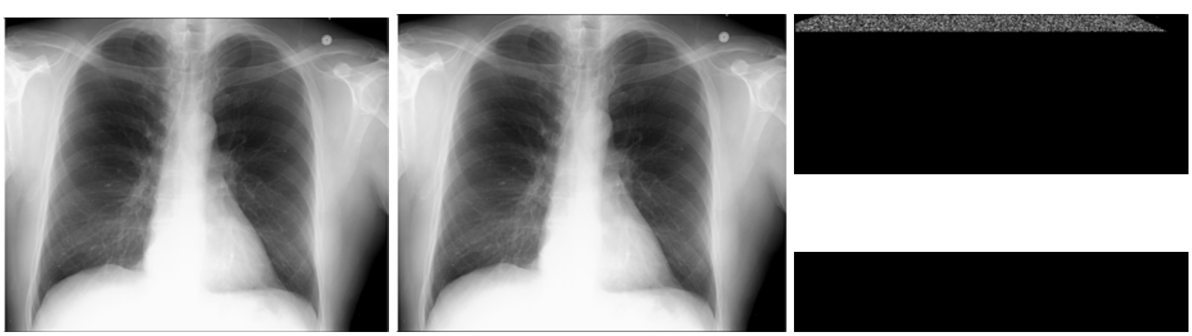


(a) (b) (c)

(d) (e) (f)

Figure 10 (a) Original MRI image. (b) Signature embedded image from the random pixel DSE. (c) Subtracted image between (a) and (b). (d) Signature embedded image from the RS Groups DSE. (e) Subtracted image between (a) and (d). (f) Binary value of the DS of the original MRI image.

Figure 11 shows the results of a CR chest image. Once again, we can see no visual degradation in the signature embedded image Figure 11(b) from the original image 11(a). As we have found, the random pixel DSE method is not very effective to CR image because the LSB contains most random noise. In contrast, the RS Groups DSE method is very effective to the CR image because it provides large amount of pixels to get R and S groups. One can see that the embedded bits only utilize very few pixels in the entire image shown in Figure 11(c). The actual % of pixel changed is 1.7%.



(a) (b) (c)

Figure 11(a) the original CR chest image, 11(b) the signature embedded image from RS Groups DSE, 11(c) the subtracted image between 11(a) and 11(b).

We also recorded the time performance of the DSE methods, since one of the most important factors to be able to adopting the DSE methods in the data grid is fast process time. Table 1 tabulates the time performance of two DSE methods in the experiments with four types of medical images including CR, CT, MR and US. The experiments were performed in a PC with one Pentium IV 2.8GHz processor and 512 MB memory.

As we can see, the time required for signing and verifying decreases from 0.22 seconds down to 0.008 seconds as image size decreases. Since the DS is generated using the secure hash algorithm SHA1withRSA approach [10], the time for signing is usually longer than the one for verifying demonstrated by other studies [10]. We can also see that the time performance for signing or verifying is almost same in both DSE methods, because the signing and verifying processes are independent of the embedding and extracting processes.

Imaging Modality	Random Pixel DSE (sec.)				RS Groups DSE (sec.)			
	Sign	Embed	Extract	Verify	Sign	Embed	Extract	Verify
MR	0.008	0.086	0.038	0.004	0.009	0.083	0.042	0.005
US	0.012	0.09	0.09	0.01	0.012	0.11	0.039	0.009
CT	0.027	0.12	0.085	0.032	0.027	0.08	0.026	0.042
CR	0.22	0.33	0.33	0.2	0.19	0.079	0.032	0.21

Table 1 Time performance of two DSE methods in the experiments.

The time performance of embedding using the random pixel DSE is much larger in CR image (0.33 seconds), than the time (about 0.1 seconds) for other types of images. The reason for this difference is that the tested CR image contains the most noise in the selected bit plane, which results in much more bits to embed the same size DS than the ones needed in other types of images. On the other hand, the time performance of embedding remains nearly the same for all the tested images using the RS Groups DSE. This indicates that the random noise in the selected bit plane has very few effect on the RS Groups DSE method.

Considering the performance of DICOM transmission of the images over T1 and Internet2 tabulated in Table 2, we can calculate the overhead of the transmission by

$$(T_{sign} + T_{embed})/T_{trans.} \quad (7)$$

or

$$(T_{extract} + T_{verify})/T_{trans.} \quad (8)$$

where T denotes the time performance of the process, e.g., T_{sign} the time performance of the sign process.

	MR (sec.)	US (sec.)	CT (sec.)	CR (sec.)
T1 (1.5 MB/s)	0.9	2	3.2	38
I2 ($\geq$ 350 MB/s)	0.8	0.95	1.1	3.3

Table 2 Time Performance of DICOM Transmission over Internet2 and T1.

Table 3 lists the overhead of the DICOM transmission using the DSE method through T1 or Intenet2. As we can see, the overhead is bigger for image transmission through Internet2 than through T1. For example, the overhead is 16% for CR image through Internet2 connection, whereas it is 1% for transmission through T1. Also, the overhead is bigger using the RS groups DSE method than the one using the random pixel DSE method, even the image is transmitted over the same network. For example, 8% for CR using the RS groups DSE method versus 16% using the random pixel DSE method. This difference between the two DSE methods decreases as the image size decreases. The overheads are almost same for MR image using both DSE methods in transmission through T1 or Internet 2. Considering the number of images in MR or CT exams, the current overhead indicates that the process time of the DSE methods still needs improvement in order to be able to apply in clinics.

		DICOM Trans. Overhead through T1 (%)				DICOM Trans. Overhead through Interne2 (%)			
		MR	US	CT	CR	MR	US	CT	CR
Random pixel	SE*	10	5	4	1	11	10	13	16
	EV*	4	5	3	1	5	10	10	16
RS Groups	SE*	10	6	3	0.7	11	12	9	8
	EV*	5	2	2	0.6	6	5	6	7

* SE: Sign & Embed processes, EV: Extract & Verify processes

Table 3 Time overhead of DICOM Transmission using the DSE method through T1 or Internet2.

5. CONCLUSION

In this paper, we presented the digital signature embedding (DSE) methods to ensure the integrity of medical images of the off-site backup data grid. A novel 3D DSE method was investigated for the recent 3D volumetric images. The 2D DSE methods have been tested in the laboratory experiments and some results from the experiments were presented. The system integration of the DSE with the data grid has also been studied. The laboratory testing of the system integration of the DSE with the data grid and clinical evaluations will be continued.

ACKNOWLEDGEMENT

This research is partially supported by a NIH Grant No. R01-LM06270 from the National Library of Medicine.

REFERENCES

1. S.B. Berger, B.B. Cepelewicz, *Medical-legal issues in tele-radiology*. Am. J. Roentgenoloy, vol. 166, pp. 505-510, 1996.
2. L. Berlin, *Malpractice issue in radiology-teleradiology*. Am. J. Roentgenoloy, vol. 170, pp. 1417-1422, 1998.
3. G.H. Kamp, *Medical-legal issues in teleradiology: a commentary*. Am. J. Roentgenoloy, vol. 166, pp. 511-512, 1996.
4. B.J. Liu, F. Cao, M.Z. Zhou, G. Mogel*, Trends in PACS image storage and archive*. J. Comp Med Imaging & Graphics, (27) 2-3, p. 165-74, 2003.
5. A. Chervenak, I. Foster, C. Kesselman, C. Salisbury, S. Tuecke. *The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets***.** *Journal* of Network and Computer Applications, 23:187-200, 2001.
6. L. Leoni, S. Manca, A. Giachetti, G. Zanetti. *A virtual data grid architecture for medical data using SRB*, EuroPACS-MIR, 2004.
7. X.Q. Zhou, H.K. Huang, S.L. Lou, *Authenticity and Integrity of Digital Mammography Images*, IEEE Trans. Medical Imaging, vol. 20, no. 8, pp. 784 – 791, 2001.
8. X.Q. Zhou, H.K. Huang, S.L. Lou, *A study of secure method for sectional image archiving and transmission*. SPIE Medical Imaging, vol. 3662, pp. 138-144, 1999.
9. J. Fridrich, M. Goljan, R. Du, *Lossless Data Embedding – New Paradigm in Digital Watermarking*, EURASIP Journ. Appl. Sig. Proc., vol. 2002, no.2, pp. 185–196, 2002.
10. A.J. Menezes, P.C. Oorschot, S.A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1997.

A HIPAA-COMPLIANT ARCHITECTURE FOR SECURING CLINICAL IMAGES

Brent J. Liu, Zheng Zhou, H.K. Huang
Image Processing & Informatics Laboratory, Dept. of Radiology,
Keck School of Medicine, USC, Marina del Rey, 90292

ABSTRACT

The HIPAA (Health Insurance Portability and Accountability Act, Instituted April 2003) Security Standards mandate health institutions to protect health information against unauthorized use or disclosure. One approach to addressing this mandate is by utilizing user access control and generating audit trails of the various authorized as well as unauthorized user access of health data. Although most current clinical image systems (eg, PACS) have components that generate log files as a solution to address the HIPAA mandate, there is a lack of methodology to obtain and synthesize the pertinent data from the large volumes of log file data generated by these multiple components within a PACS. We have designed and developed a HIPAA Compliant Architecture specifically for tracking and auditing the image workflow of clinical imaging systems such as PACS. As an initial first step, a software toolkit was implemented based on the HIPAA Compliant architecture. The toolkit was implemented within a testbed PACS Simulator located in the Image Processing and Informatics (IPI) lab at the University of Southern California. Evaluation scenarios were developed where different user types performed legal and illegal access of PACS image data within each of the different components in the PACS Simulator. Results were based on whether the scenarios of unauthorized access were correctly identified and documented as well as normal operational activity. Integration and implementation pitfalls were also noted and included.

Keywords: HIPAA, Security, Auditing, Monitoring

1. INTRODUCTION

Health Insurance Portability and Accountability Act (HIPAA) [1, 2] of 1996, Public Law 104-191, was officially instituted on April 14, 2003 to enforce healthcare providers to be compliant by April, 2005 deadline. The major goal and focus of HIPAA is to set and enforce broad standards in the attempt to protect the privacy and security of health data throughout the patient care environment. To date, there are four types of standards in HIPAA:

1) Transaction and Code Set Standards
2) Identifier Standards
3) Privacy Standards
4) Security Standards

In this paper, we focus on the fourth standard type - security. HIPAA Security Standards [3] are aimed at the protection of confidentiality, integrity, and public availability of electronic health information against unauthorized use or disclosure. This is accomplished by utilizing administrative, physical, and technical safeguards. In particular, the technical safeguards consist of technical methods to assure security of the health data. One such technical method proposed by HIPAA is the on-demand generation of an audit trail that can record and examine information system activities such as data access of a specific patient. Specifically, HIPAA compliant audit trails require following information for the health data access [4]:

- Identification of the person who accessed the data
- Identification of the accessed data
- Where the data was accessed
- Timestamp of when the data was accessed
- Types of access (eg, create, read, write, modify, delete)
- Status of access (eg, success or failure)

Because health data and information is such a broad area containing vast amounts of data types, the major focus of this research is on clinical imaging data that is generated and distributed through PACS.

Some efforts have been achieved by developing HIPAA compliant auditing tools for general health information systems [5-7]. These auditing tools generate audit trails by recording the health data transactions or changes in logs and extracting the pertinent auditing information from these logs on demand. This method is applicable for health information systems that have all the data transactions or data flow controlled by a centralized server, such as Radiology information system (RIS) [8]. However, the data flow is much different in integrated medical imaging systems, such as Picture Archiving and Communication System (PACS). There is no single component that controls and records the data flow of all the multiple components within PACS. This makes it very difficult for these auditing tools to record all the data transactions and changes in PACS. For example, the PACS archive server, even within client-server architecture, has no control of the workflow of the CT modality, and vice versa. Additionally, there are various other components within a PACS that require a system-wide architecture instead of a single component-based approach. Most current clinical imaging systems have no such ability to generate HIPAA compliant audit trails even though they generate activity logs. Furthermore, even though pertinent auditing information can be extracted from these logs to create audit trails, it requires tedious if not manual methods to produce the requested audit information and analysis. There is a lack of a formal methodology to interpret the potential large volumes of these log data and generate these HIPAA compliant audit trails. Therefore, a HIPAA compliant auditing architecture for integrated medical imaging systems needs to be tailored to the complex workflow.

In this research we present the design and development of a HIPAA Compliant Architecture specifically for tracking and auditing the image workflow of clinical imaging systems such as PACS. The architecture is designed to facilitate the generation of HIPAA compliant audit trails of image data access for a specific patient so that various types of queries can be performed on demand. It also provides the mechanisms to automatically monitor the data flow of PACS and facilitating the detection of unauthorized image access and other abnormal activities. As an initial first step, a software toolkit, called HIPAA Compliant Auditing System (HCAS), was implemented based on partial components from the HIPAA Compliant architecture. This initial HCAS toolkit was implemented and evaluated within a testbed PACS Simulator located in the Image Processing and Informatics (IPI) lab at the University of Southern California. Evaluation scenarios were developed and results were based on whether the scenarios of unauthorized access were correctly identified and documented as well as normal operational activity.

2. METHODS AND MATERIALS

2.1. Design Criteria

In order to apply the HIPAA Compliant Architecture for auditing and tracking clinical images to various PACS generating different format log files, it must be independent from any individual PACS architecture or manufacturer. For this reason, we define the necessary architecture criteria as follows:

1) **HIPAA Compliant**
   The ability to facilitate generation of the HIPAA compliant auditing trail report in terms of who access it, when, where, what are accessed, access status, and access types.
2) **Open and Extensible**
   Provide interfaces for integration of new auditing or monitoring techniques and the ability to support current HIPAA auditing requirements and accommodate new HIPAA additions in the future without affecting already existed components.
3) **Portable**
   Not tied-down to any individual PACS or PACS architecture.
4) **No interruption of clinical PACS workflow**
   Any interruption on the workflow of PACS is avoided.

2.2. HIPAA Compliant Architecture Development

Based on the criteria described above, the HIPAA Compliant Architecture was designed as a four-layer system shown in Fig. 1. The first layer (the lowest layer) is the Record layer, consisting of various logs within PACS components. By logically separating PACS logs from other components in HCAS, independence from PACS and portability can be achieved. The second layer is the Audit layer, which includes a centralized auditing database and other audit data analysis and interpretation tools. HIPAA compliant audit trails can be generated based on the auditing database. This layer also enables us to automatically monitor the data flow of PACS, which greatly assists PACS management. The third layer is Notification layer, which has a Notification component sending warning or alert messages of abnormal events to end users, such as PACS administrators. Finally, in the fourth layer, end users can decide to take certain actions against these abnormal events. These layers will be described in more detail in the following paragraphs.

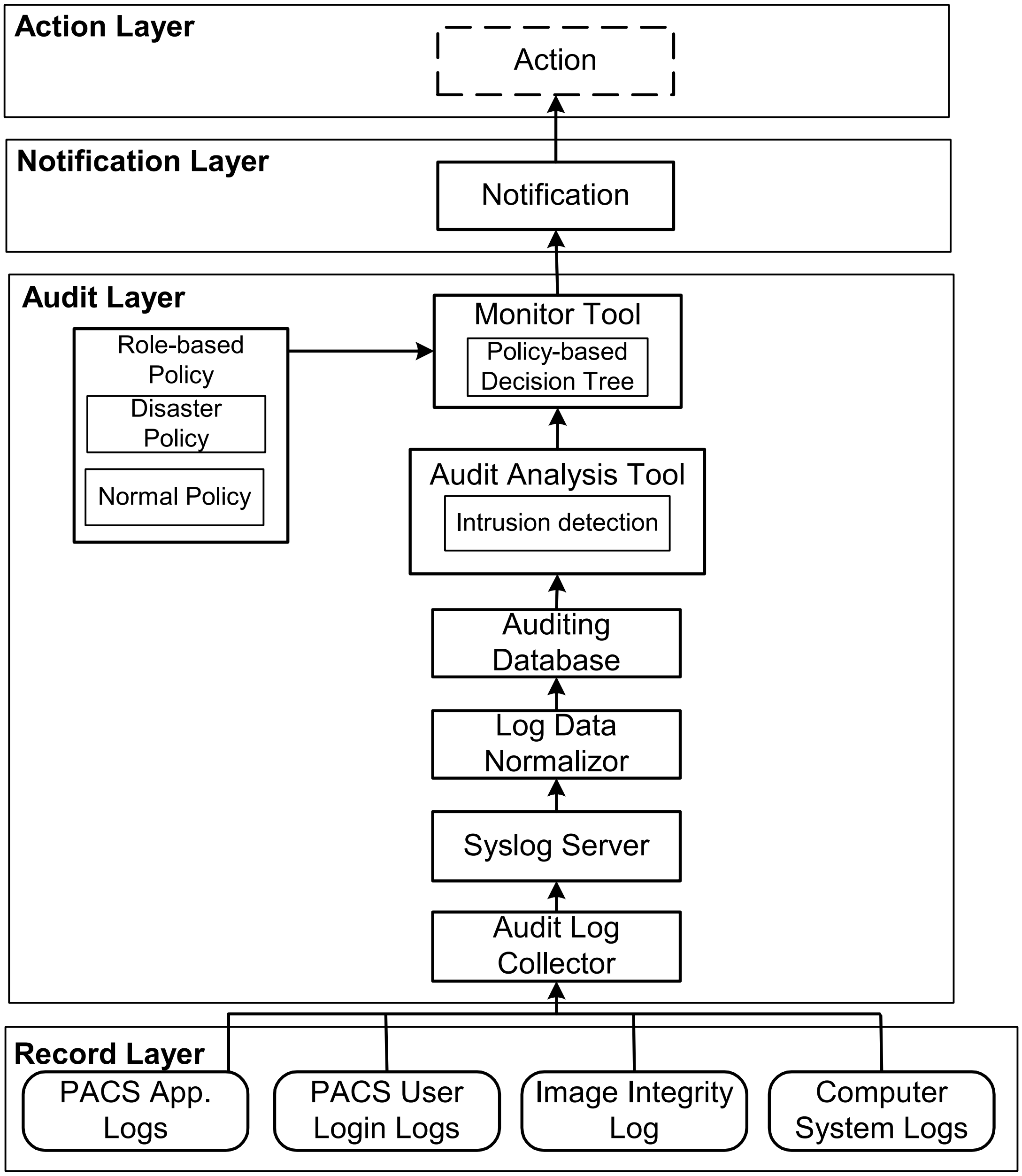


Figure 1: The four layer HIPAA compliant architecture for auditing of medical images in PACS showing the carious components for each layer: 1) Record layer, 2) Audit Layer, 3) Notification Layer, and 4) Action Layer.

<u>**Record Layer**</u>

This first layer is the data resource layer, including but not limited to the various types of log data shown in Figure 1. PACS application logs are event logs generated by the individual PACS applications. For example, an image query/retrieve event in PACS archive server may include such information as time, local host name, DICOM Application Entity Title (AET), patient information and query/retrieve status. PACS user login logs record user login events in each individual PACS component. Other computer system logs generated in PACS components, such as application access logs, can also provide supplement information.

Due to the flexibility of this architecture, new logs can also be added to this layer. For example, an image integrity log can be added to record image data integrity verification events. Data integrity, as one requirement of HIPAA Security Standards, refers to protecting image data from being altered or destroyed by unauthorized users. A Lossless digital signature embedding (LDSE) method has been developed to ensure the data integrity of medical images at IPI laboratory [9]. By recording signature verification time, local machine, and signature verification status in the integrity log, the LDSE method can provide logs to generate HIPAA compliant audit trails on the data integrity of image. To extract and interpret the pertinent information from thousands of log events requires proper methodology, which will be addressed in the second layer, the Audit layer.

Audit Layer

As shown in Fig. 1, the Audit Layer is the heart of the architecture. It collects the audit data from distributed PACS components and stores the data in a centralized auditing database. The database is then used for audit analysis and automatic monitoring. Currently, there are seven components in this layer: Audit Log Collector, Syslog Server, Log Data Normalizer, Auditing Database, Audit Analysis Tool, Role-based Policy and Monitor Tool.

a. Audit Log Collector

Because the audit data is scattered within large volume of logs, a collector was designed to extract the pertinent data from logs and send the data to the centralized Auditing Database. PACS logs may be stored with different formats, such as database tables or textual files. The collector must support all these types of logs.

b. Syslog (System log) server

The pertinent data extracted from PACS logs are distributed in different PACS components connected by digital networks. In order to store them in the centralized Auditing database, a transmission mechanism is needed.

Currently, syslog [10] is a de-facto standard for transport and storage of event notification messages in UNIX systems, network devices and network applications. Syslog is a client-server mechanism. The clients can be configured to locally store event messages or directly send event messages to the server without local storage. Syslog uses User Datagram Protocol (UDP) to transfer event messages. This feature can be utilized to reduce the overhead added to the image transmission in PACS caused by event message communication, since PACS uses DICOM (Digital Imaging and Communications in Medicine) Protocol and Transmission Control Protocol (TCP). For this advantage, syslog technology was adopted as the architecture standard to transfer pertinent log data. The data is converted to syslog format by the syslog client in each PACS component. The client then sends the data to Syslog server, which will forward the data to the Log Data Normalizer.

c. Log Data Normalizer

The pertinent data extracted from PACS components might have different terminologies for the same object. For example, the name “film librarian” in the CT modality might be named as “clerk” in an MR modality. For this reason, a Log Data Normalizer was designed to normalize the data into common terms and then add them to the Auditing Database.

d. Auditing Database

In order to generate HIPAA compliant audit trails in a short time, a centralized database was designed to preserve all the obtained auditing data. The structure of database was designed based on the requirement of HIPAA compliant audit trails, including: who, when, where, what, how, and status. Patient information, such as name and id, and other relevant information are also included in the database. The advantages to use database technology to preserve the log data are:

- **No loss of historical logs**: since all the logs generated in PACS components are obtained and stored in the database everyday, there is no loss of log data when these logs are overwritten and updated by PACS components.
- **Centralized management of data access information**: The image data access events for an individual patient usually happen in multiple PACS components. For example, an event that a CT image is generated in a CT modality and another event that the same CT image is retrieved to viewing workstation for clinical review are related to the same patient. But these two events were recorded in two different logs at two separate PACS components. Without centralized database, the pertinent information needs to be extracted from these two components every time HIPAA compliant audit trails of image access for this patient is desired. Therefore, a centralized database enables us to quickly generate audit trails.

e. Audit Analysis Tool

Most current PACS lack a mechanism to dynamically monitor the data flow, which results in PACS management mostly relying on the experience of PACS administrators. A monitoring tool that can automatically analyze the data to find abnormal patterns and make decisions on the patterns would make PACS management much easier. To develop such a tool, the information of data flow of PACS needs to be collected and analyzed in real time. With audit data collected in the auditing database, the HIPAA Compliant architecture can provide this ability using some data analysis techniques, such as Intrusion detection technology [12]. Audit Analysis Tool is the component to perform such data analysis functions.

f. Monitor Tool

After the Audit Analysis Tool finds abnormal patterns in the data flow of PACS, a Monitor Tool was designed to monitor the pattern, and make decisions whether it is an unauthorized data access for the abnormal pattern based on the Role-based Policy. Any pattern that violates the Policy would automatically cause a warning or alert result. For example, Audit analysis tool discovers an abnormal pattern of image query/retrieve by a PACS user "A", belonging to the role of "Clerk", which was defined to have no image query/retrieve right in the Policy. The Monitor tool automatically makes a decision that this is an unauthorized image query/retrieve and gives a warning message.

g. Role-based Policy

The Role-based Policy defines the roles for PACS users based on the roles they performed in the clinical environment, such as clerk, PACS manager, and Radiologists, and the image access rights for each role. Two types of policies, Normal policy and Disaster policy, are defined for two different conditions. Normal policy is for daily operation, whereas Disaster policy is defined for the emergency situations, such as earthquake, when normal policy can be bypassed.

Notification Layer

Notification layer consists of a notification component, which receives the warning or alert messages from the Audit layer and notifies PACS end users of the unauthorized image data access and other abnormal activities.

Action Layer

Action Layer is designed for PACS end users to take actions, such as access control, against the unauthorized image access and other abnormal activities. This four-layer architecture enables PACS to generate HIPAA compliant audit trails of image data access for a specific patient on demand. Meanwhile, it can automatically monitor the data flow of PACS facilitating PACS management. With an open and extensible design, the architecture can also easily incorporate new data analysis and monitoring techniques, and be extended to support future HIPAA requirements.

3. RESULTS AND DISCUSSION

A HIPAA Compliant Auditing System (HCAS) toolkit has been developed for automatic monitoring the data flow of PACS based on partial components of the Audit Layer in the architecture [12]. The HCAS toolkit and its GUI were installed in a UNIX machine. The HCAS toolkit currently includes such components as Audit Log Collector, Syslog Server, Auditing Database, Monitor Tool and Role-based Policy (Normal Policy). A simple role-based policy was designed according to the roles of healthcare providers. The policy consists of three tables: role table, resource table and policy table. The role table defines different roles, such as radiologists. The resource table records PACS applications, such as the viewing software. The policy table, shown in Table 1, assigns each user the role and the resources this user can access based on the right of his/her role. For example, one entry in the policy table is user "Michael", "System Administrator", and "All". "All" means all the resources can be accessed by this user.

User Name	Role Name	Resources
Michael	System Administrator	All
John	Clerk	Viewing software
Joe	Radiologist	All
Jessica	Technician	CT scanner

Table 1: Example of a role-based policy table which can be set up & configured based on each healthcare institution's own staff.

The toolkit can monitor the dynamic data flow of PACS. First, it collects pertinent auditing data from PACS application logs, PACS user login logs and other computer system logs. It then stores the log data in the Auditing Database. Next, the toolkit compares the user name in every record in the Auditing Database and the user name in the policy table. If a match occurs, the toolkit further compares the application name in the database record and the application name in the policy table. If any comparison failed, the HCAS toolkit gives out a warning message of unauthorized image data access in its graphic user interface (GUI). Otherwise, a normal message is given out. Currently, the toolkit lacks the ability to generate HIPAA compliant audit trails of image data access for a specific patient. This function is currently in development.

In order to evaluate the impact of this HCAS toolkit in PACS, a laboratory-based PACS Simulator [13, 14] was implemented with the toolkit to simulate the data flow of clinical PACS. The Simulator can simulate the complete data workflow of clinical PACS from patient registration to exam ordering, and to image generation, image archive and display. The clinical images used for simulation are replenished continuously through a clinical PACS connection but with the patient information in the DICOM header of the image removed. Figure 2 shows the implementation of the HCAS with the PACS Simulator for evaluation.

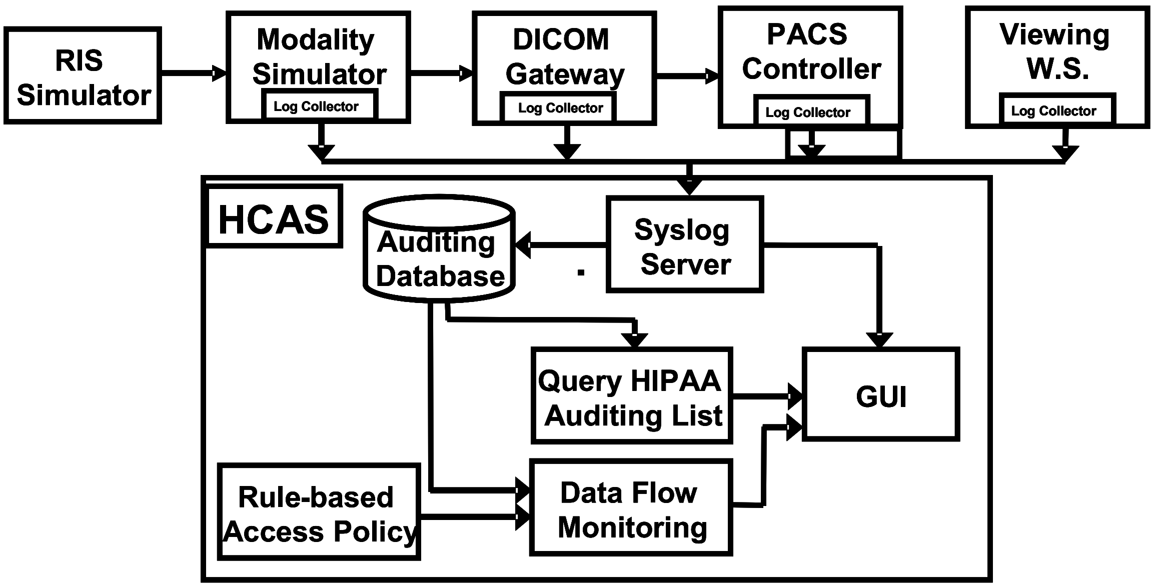


Figure 2: The HCAS toolkit implemented with the PACS Simulator in the laboratory environment showing Log Collector Clients at each of the PACS Simulator Components.

The log collector clients are installed in every component except the RIS Simulator to receive event messages of each image data access activity generated by these components. Log messages generated are automatically collected and inputted into the HCAS toolkist via the Syslog Server. The log data includes what, when and where images are accessed. In addition, user login logs and computer system logs are collected. These data, along with log data already collected by the PACS monitoring client in PACS controller are used as input data for the evaluation of the HCAS toolkit. Pertinent information in the log database of the PACS controller was collected and stored in the Auditing Database. The user login information and application access information from the computer system logs of the DICOM gateway and viewing workstations were also collected. For example, user "Michael" logged on a viewing workstation and made an image query/retrieve with the viewing software. A viewing software access event with the user name was recorded in computer system logs in the workstation. Meanwhile, the PACS controller also records this query/retrieve activity in its log database. Comparing the time of the software access event and the time of the image query/retrieve and assuming that only a single user was using the viewing software at that time, these two events can be related together and the name of the user who did this query/retrieve can be obtained.

The laboratory evaluation is currently ongoing.

4. CONCLUSION

The advent of HIPAA greatly impacts medical imaging systems, such as PACS, and even the entire health information systems. To be HIPAA compliant, every medical imaging system must satisfy the HIPAA requirement of audit trails.

In this research, we presented a HIPAA compliant architecture for auditing medical images in PACS. The architecture enables PACS to generate HIPAA compliant audit trails of image data access for a specific patient on demand. It also enables PACS to automatically monitor the image flow in the system, including detection of unauthorized usages of image data and other abnormal activities. As and initial first step a HIPAA Compliant Auditing System (HCAS) toolkit has been developed based on partial components of the Audit layer within the architecture. The toolkit can automatically monitor the image flow in PACS. Currently, the toolkit lacks the ability to generate HIPAA compliant audit trails, which is in development now. A PACS Simulator was integrated with the HCAS for laboratory evaluation of the toolkit. An evaluation methodology was developed and test scenarios designed accordingly to properly evaluate the HCAS toolkit. The evaluation is currently ongoing with promising initial data results.

ACKNOWLEDGEMENT

This research is partially supported by the NIH Grant No. R01-LM06270.

REFERENCES

1. HIPAA, http://www.cms.hhs.gov/hipaa/hipaa2/general/background/pl104191.asp

2. HIPAA, http://www.rx2000.org/KnowledgeCenter/hipaa/hipfaq.htm

3. HIPAA Security Standard, 2003, http://www.cms.hhs.gov/hipaa/hipaa2/regulations/security/03-3877.pdf

4. Cao F., Huang H.K., Zhou X.Q., Medical image security in a HIPAA mandated PACS environment, *Computerized Medical Imaging and Graphics*, Vol. 27, Num. 2-3, pp. 185-196, 2003.

5. HIPAA audit, http://www.datamirror.com/products/liveaudit/

6. HIPAA audit tool, http://www.axolotl.com/press/20021113/

7. HIPAA audit tool, http://www.peacefulpackers.com/it_solutions/hs2.htm

8. Huang H.K., PACS: *Basic Principles and Applications*, Wiley-Liss, P291.

9. Z. Zhou, H.K. Huang, B.J. Liu, *Digital Signature Embedding for Medical Image Integrity in a Data Grid Off-site Backup Archive*, SPIE Medical Imaging, in Process, 2005.

10. Syslog, http://www.loriotpro.com/Products/SyslogCollector/SyslogDataSheet_ENv3.php

11. White G.B., Fisch E.A., Pooch U.W., *Computer System and Network Security*, CRC Press, 1996.

12. Z. Zhou, B.J. Liu, H.K. Huang, J. Zhang, *A RIS/PACS Simulator Integrated with the HIPAA-Compliant Auditing (HCA) Toolkit,,* SPIE Medical Imaging, in Process, 2005.

13. Law M.Y.Y., Zhou Z., New direction in PACS education and training, Computerized Medical Imaging and Graphics, Volume 27, pp. 147-156, 2003.

14. Zhou Z., Huang H.K., Cao F., Liu B.J. Zhang J. Mogel G.T., Educational RIS/PACS Simulator, SPIE Medical Imaging, Vol. 4, pp. 139-147, 2003.

Educational RIS/PACS Simulator Integrated with the HIPAA Compliant Auditing (HCA) Toolkit

Zheng Zhou[a], B.J. Liu[a], H.K. Huang[a], J. Zhang[b]
[a]Image Processing & Informatics (IPI) Laboratory, University of Southern California,
4676 Admiralty Way, Suite1001, Marina del Rey, CA, 90292
[b]Shanghai Institute of Technical Physics, Shanghai, China

ABSTRACT

Health Insurance Portability and Accountability Act (HIPAA), a guideline for healthcare privacy and security, has been officially instituted recently. HIPAA mandates healthcare providers to follow its privacy and security rules, one of which is to have the ability to generate audit trails on the data access for any specific patient on demand. Although most current medical imaging systems such as PACS utilize logs to record their activities, there is a lack of formal methodology to interpret these large volumes of log data and generate HIPAA compliant auditing trails. In this paper, we present a HIPAA compliant auditing (HCA) toolkit for auditing the image data flow of PACS. The toolkit can extract pertinent auditing information from the logs of various PACS components and store the information in a centralized auditing database. The HIPAA compliant audit trails can be generated based on the database, which can also be utilized for data analysis to facilitate the dynamic monitoring of the data flow of PACS. In order to demonstrate the HCA toolkit in a PACS environment, it was integrated with the PACS Simulator, that was presented as an educational tool in 2003 and 2004 SPIE. With the integration of the HCA toolkit with the PACS simulator, users can learn HIPAA audit concepts and how to generate audit trails of image data access in PACS, as well as trace the image data flow of PACS Simulator through the toolkit.

Key words: HIPAA compliant, Audit trails, Auditing toolkit, Dynamic monitoring, PACS simulator

1. INTRODUCTION

Health Insurance Portability and Accountability Act (HIPAA) [1,2] of 1996, Public Law 104-191, became law on April 14, 2003. The purpose of HIPAA is to set and enforce the standards to protect the privacy and security of health data. Currently, there are four types of standards in HIPAA: Transaction and Code Set Standards, Identifier Standards, Privacy Standards and Security Standards. HIPAA mandates healthcare providers to fulfill these standards by certain deadlines. For example, the deadline for being compliant with the Privacy Standard is April 14, 2005. Among these four standards, the Security Standards [3] are the only one to protect confidentiality, integrity and availability of electronic health information. HIPAA specifies three types of protections including administrative, physical and technical safeguards. The technical safeguard defines the technical methods to assure the security of the electronic health data. One such method suggested by HIPAA is an audit trail, which can record and examine information system activities [3]. HIPAA requires health institutions to have the ability to generate audit trails on data access activities for a specific patient on demand. Specifically, HIPAA compliant audit trails require the following information for access to health data:

- Identification of the person who access the data
- Identification of the data that is accessed
- Location of where the data is accessed
- Date and time when the data is accessed
- Types of access (Create, read, write, modify, delete)
- Status of access (success or failure)

Most current health information systems have no such ability to generate HIPAA compliant audit trails, even though they do generate activity logs. These logs can be utilized to generate audit trails extracting the pertinent auditing information from the log data. However, there is a lack of a formal methodology to interpret the potential large volumes of the log data to generate HIPAA compliant audit trails. Some efforts have been achieved by developing HIPAA compliant auditing tools for text-based health information systems [4-6]. These auditing tools generate audit trails by recording the health data transactions or changes in logs and extracting the pertinent auditing information from these logs on demand. This approach is effective for health information systems that have all the data transactions or data flow controlled by a centralized server, such as Radiology information system (RIS) [7]. However, an integrated medical imaging system such as Picture Archiving and Communication System (PACS) [7] has a rather different data flow. There is no single component that can control and record the data flow of all components in PACS. For example, PACS archive server has no record of who logs in a CT modality and generates a CT exam, neither does the CT modality have any record of the PACS server receiving the CT exam. This makes it very difficult for these auditing tools to cope with the auditing of image data flow in PACS, because the data access events are distributed in separate components. Therefore, a HIPAA compliant auditing tool tailored to the complex data flow of PACS needs to be developed.

In this paper, we focus on developing a HIPAA compliant auditing (HCA) toolkit for auditing of image data flow in PACS. Section II will present the system design and development of the toolkit in details. In order to evaluate the performance of the toolkit, the PACS Simulator was be utilized as the test bed. The system integration will be described in Section III. Section IV will present some preliminary results from the laboratory experiments and discussion. The conclusion will be given in Section V.

2. SYSTEM DESIGN AND DEVELOPMENT OF HCA TOOLKIT

The goal is to design a HCA toolkit for auditing the image data flow in PACS and the imaging modalities connected to PACS with the log data from these applications as input sources. In order to be able to apply the toolkit in various systems instead of one specific system, the toolkit is designed to be independent from any individual PACS or imaging modality. Also, the toolkit has to be open architecture in order to be easily extended with future HIPAA requirements.

2.1 System design of the toolkit

The HCA toolkit was designed utilizing two layers from the HIPAA Compliant architecture for auditing and tracking clinical images [8] and is shown in Figure 1.

2.1.1 Record Layer

The first layer is Record layer, which is the data resource layer consisting of various types of logs from PACS components. Table 1 tabulates some examples of the logs. PACS application logs record data access events in the PACS components. For example, an image query/retrieve event in PACS archive server is one such event recorded with the information consisting of time, local host name, DICOM Application Entity Title (AET), patient information and query/retrieve status. PACS user login logs record user login events in each individual PACS component. Besides these conventional log data, there are also new logs such as image integrity log recording image data integrity verification events. Data integrity, as one requirement of HIPAA Security Standards, refers to the assurance that image data has not been altered or destroyed by unauthorized users. A digital signature embedding (DSE) method has been developed to ensure the data integrity of medical images [10]. By recording signature verification time, local machine, and signature verification status in the integrity log, the DSE application can provide logs for the HCA toolkit to generate HIPAA compliant audit trails on the data integrity of image.

All these logs provide the pertinent information needed to generate HIPAA compliant audit trails. However, to extract and interpret the pertinent information from thousands of log events requires proper methodology, which will be presented in the second layer, the Audit layer.

Table 1 Types of logs existing in the HCAS Record Layer.

Type of Logs	Time	Location	User name	Status	DICOM AET*
PACS Application Logs	Operation time	Local operation machine	N/A	Status of this operation	DICOM AET of this application
PACS User Login logs	User login time	Local machine user login	Login user name	Status of login	N/A
Image Integrity Logs	Integrity check time	Local machine does check	N/A	Status of integrity check	N/A

*AET: Application Entity Title

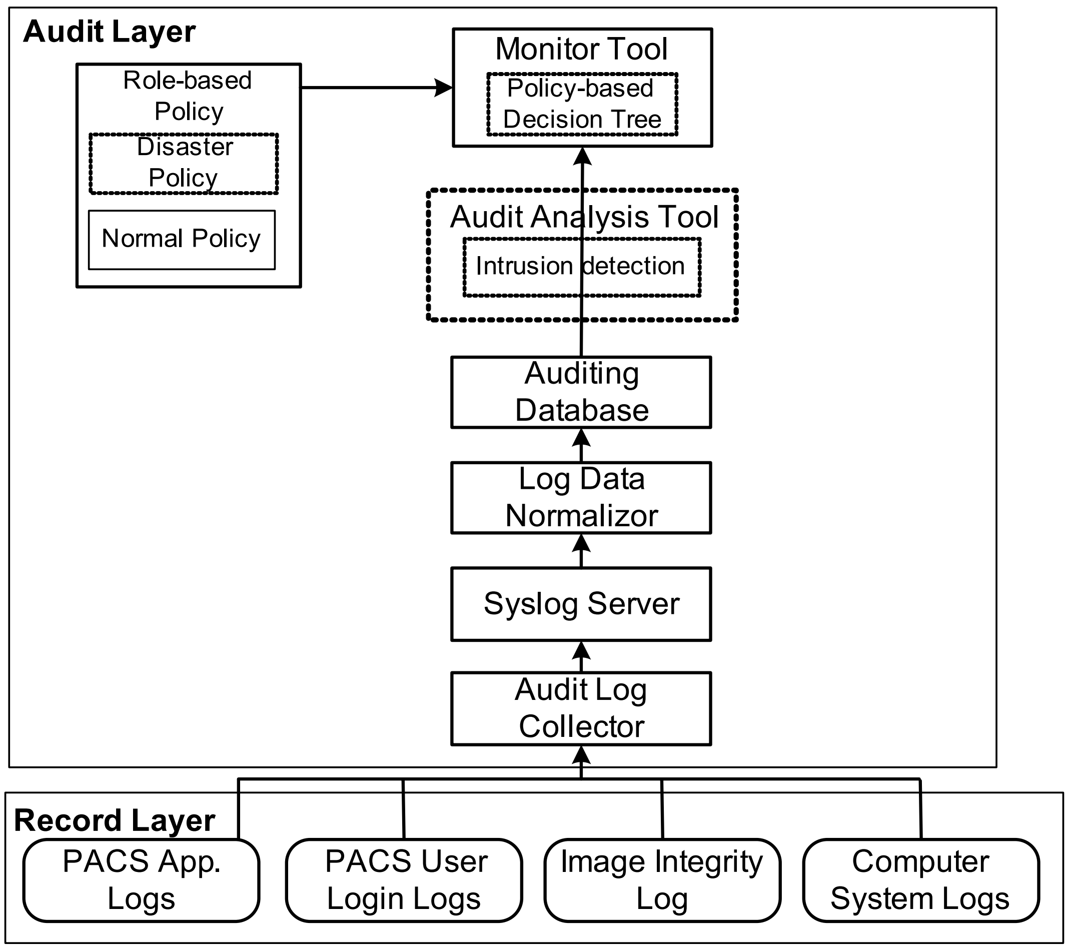


Figure1 Components and data flow of the HCA toolkit, where the components with dotted line have not been implemented.

2.1.2 Audit Layer

The Audit Layer collects the audit data from distributed PACS components, normalizes the data in a certain format, and stores the data in a centralized auditing database. The database is the heart of the Audit layer. It manages all the log data to facilitate the generation of the audit trails. It is also utilized in data analysis for dynamic monitoring the data flow of

PACS. The audit layer consists of seven components including Audit Log Collector, Syslog Server, Log Data Normalizer, Auditing Database, Audit Analysis Tool, Role-based Policy and Monitor Tool.

a. Audit Log Collector

Because the pertinent audit information is contained in a large volume of logs, a data collector was designed to extract the information from logs and send the pertinent information to the centralized Auditing Database. The logs could be stored in different formats such as database tables or textual files. The collector must be able to cope with all these types of logs.

b. Syslog (System log) server

The information collected by the Audit Log Collector is distributed in different PACS components. A transmission mechanism is needed to transfer them to the Auditing database. Currently, syslog [11] is a de-facto standard for transport of event notification messages in UNIX systems, network devices and network applications. Syslog is a client-server mechanism. The clients can be configured to store event messages locally or directly send event messages to the server. Syslog uses User Datagram Protocol (UDP) to transfer event messages. Thus, very little transmission overhead will be added to the image transmission in PACS because PACS uses DICOM (Digital Imaging and Communications in Medicine) Protocol and Transmission Control Protocol (TCP) to transfer data. For this reason, Syslog technology was utilized in the HCA toolkit to transfer the log data. Once the data is collected, the syslog client automatically transfers the data to the syslog server running in the computer where the central database resides.

c. Log Data Normalizer

Before the log data being added to the database, the format of the data needs to be normalized because different terminologies may be used for the same object in different modalities. For example, the name “film clerk” in the CT modality might be named as “clerk” in an MR modality. The Log Data Normalizer can convert the data to the correct format in the database, thus all the events related to the same object can be found when generating the audit trails.

d. Auditing Database

The database provides powerful management tool and search engine for the HCA toolkit. The advantages to using a database to preserve the log data are:

1) No loss of historical logs. Since all the logs generated in PACS components are immediately obtained and stored in the database, there is no loss of log data when these logs are overwritten or updated by PACS components.
2) Centralized management of data access information for every patient. The image data access events for an individual patient usually happen in multiple PACS components. For example, an event that an exam is generated in a CT modality and another event that the same CT exam is retrieved to the viewing workstation for clinical review are related to the same patient. But these two events were recorded in two different logs at two separate PACS components. Without the centralized database, the pertinent information needs to be extracted from these two components every time that an audit trail is requested. Therefore, a centralized database improves the time performance of the HCA toolkit to generate audit trails.

The structure of the database was designed based on the requirement of HIPAA compliant audit trails, including who, when, where, what, how, and event status. Patient information such as name and patient id as well as other relevant information are also included.

e. Audit Analysis Tool

Current PAC systems lack a mechanism to dynamically monitor the data flow. This results in PACS management being very difficult and relies upon on the experience of PACS administrators. A monitoring system that can automatically report the data flow of PACS and find abnormal activities would greatly ease the PACS management. To enable the dynamic monitoring, the information of the data flow of PACS needs to be collected and analyzed in real time. With audit data already collected in the auditing database, the HCA toolkit has the ability to provide such dynamic monitoring. An Audit Analysis Tool was designed to perform the data analysis to find abnormal patterns in the log data using technologies such as Intrusion detection approach.

f. Monitor Tool

After the Audit Analysis Tool finds the abnormal patterns in the data flow of PACS, a Monitor Tool makes decision on whether it is an unauthorized data access based on the Role-based Policy. Any pattern that violates the policy would automatically cause a warning or alert message. For example, Audit analysis tool discovers an abnormal pattern of image query/retrieve by a PACS user "A", belonging to the role of "Film clerk". This role was not assigned any image query/retrieve rights in the Policy. The Monitor tool automatically makes a decision that this is an unauthorized image query/retrieve and gives a warning message. The warning or alert message can be further utilized by PACS administrators for taking actions against the abnormal activities.

g. Role-based Policy

The Role-based Policy defines the roles for PACS users based on the roles they performed in the clinical environment, such as film clerk, PACS manager and Radiologists, as well as the image access rights for each role. Two types of policies, Normal policy and Disaster policy, can be defined for two different conditions. Normal policy is for daily operation, whereas Disaster policy is defined for the emergency situations, such as an earthquake, when normal policy needs to be bypassed. In a disaster situation, people are usually not in their normal positions anymore. Thus, most of the role assignments should be cancelled to allow everybody to be able to access all the available resources.

2.2 Technical development

According to the aforementioned system layout of the toolkit, a prototype of the toolkit has been implemented. The following programming languages, database software and platforms were chosen:

- JAVA programming language
- PostgreSQL database software
- LINUX platform (REDHAT 9)

The prototype consists of the Audit Log Collector, the Syslog Server, the Log Data Normalizer, the Auditing Database, a simple Monitoring Tool and a simple Role-based Policy. Other components with dotted line in Figure 1 are still in development.

3. INTEGRATION OF THE HCA TOOLKIT WITH THE PACS SIMULATOR

In order to evaluate the impact of the HCA toolkit in PACS, it has been integrated with the laboratory-based PACS Simulator.

3.1 Educational PACS simulator

The PACS simulator, developed as an educational tool and a test bed for PACS applications, can simulate the complete data workflow of clinical RIS/PACS from patient registration, exam ordering, image generation, image archive, and to image display [9]. As shown in Figure 2, the simulator consists of six key components: RIS simulator, Acquisition Modality Simulator (AMS), DICOM gateway, PACS Controller, Viewing workstations, and a PACS monitoring system. The AMS is connected to a clinical PACS and contains thousands of CT, MR, US, CR, and digitized film examinations. The images are continuously replenished through the clinical PACS connection for training and testing.

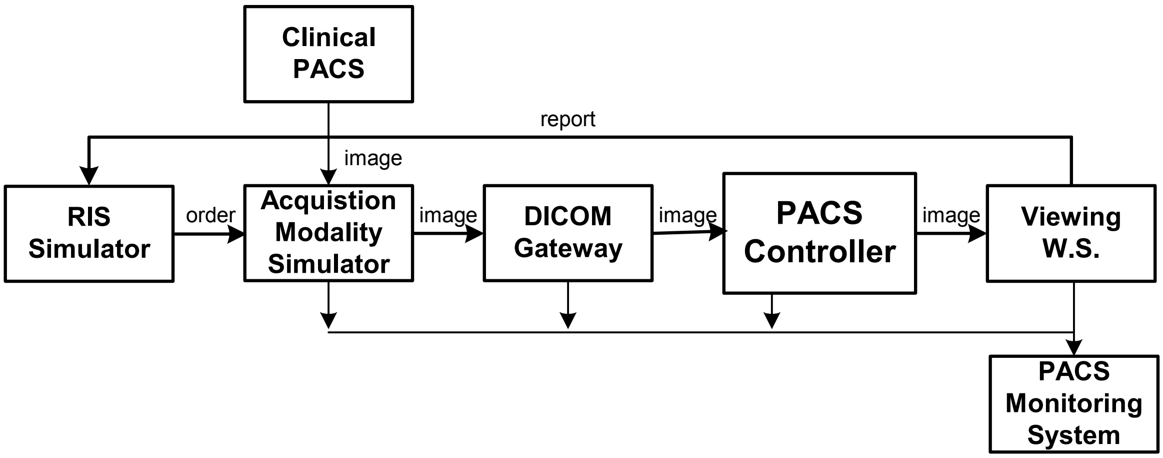


Figure 2 Data flow of the PACS Simulator and PACS monitoring system.

The PACS monitoring system was originally developed to monitor all the image data flow going through the components of the PACS Simulator in real-time. The PACS monitoring system consist of multiple clients and a centralized monitor server. The clients are installed in every component except the RIS Simulator to receive event message of each image data access activity generated by these components and generate event logs. The clients then send the log data to the monitor server, which can display the log data in a graphic user interface (GUI) for PACS users to monitor image data flow at each component. The PACS monitoring system has two limitations:

1) No preservation of historical log data
   Without a database to preserve all the log data, most of the historical log data are lost after being displayed in the GUI of the monitor server. Therefore, no pertinent information can be extracted to generate the HIPAA audit trails. Neither can any data analysis be performed to facilitate more sophisticated dynamic monitoring.
2) No PACS user information
   No user information is included in the log data generated by the monitoring system.

3.2 System integration

Because of no log data preserved in the monitor server, the log data for the HCA toolkit has to be collected from each component including AMS, DICOM gateway, PACS controller and the viewing workstation. Figure 3 shows the integration of the HCA toolkit with the PACS simulator. Because the implementation for different components is different, the log data of the PACS simulator components are stored in different formats. For example, the log data of the AMS is stored in Microsoft Access Database, whereas the log data of the PACS controller is stored in Oracle Database. Also, the user login information needs to be collected from different operating systems such as WINDOWS and UNIX. An interface to extract the log data is specially developed for each component based on their log format. The interface is then called by the Audit data collector, which is installed in each component. The data collector automatically forwards the collected log data to the centralized HCA toolkit. The HCA toolkit provides users a user friendly GUI to understand HIPAA audit trail concept by searching the data access events on specific patient or user. Meanwhile, it also provides a set of features allowing users to configure the role-based policy and dynamically monitor the image data flow of the PACS simulator based on the policy.

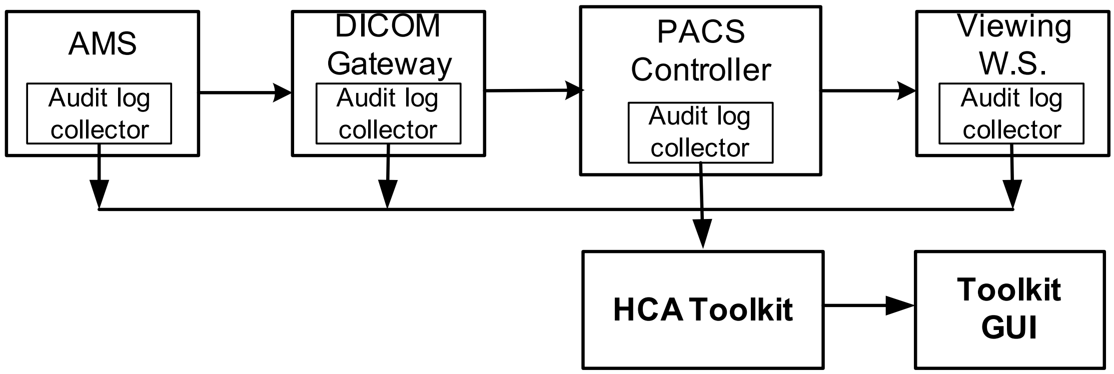


Figure 3 System integration of the HCA toolkit with the PACS simulator.

4. PRELIMINARY RESULTS AND DISCUSSION

The aim of the laboratory evaluation is to test two main functions of the HCA toolkits in the PACS environments, including:

1) Generation of the HIPAA compliant audit trail for any specific patient or user.
2) Dynamic monitoring of the image data flow of the PACS simulator to find the possible violation of the role-based policy.

4.1 Define and create testing scenarios

The testing scenarios for laboratory evaluation of the HCA toolkit are anchored around creating a clinical simulation of the PACS workflow from the exam generation to the retrieval of the exams for clinical review. From the end user point of the view, the entire PACS workflow can be broken down to two types of scenarios:

1) passive scenario, where no end user action is involved, e.g., exam transmission from DICOM gateway to PACS controller,
2) active scenario, where end user action is required, e.g., exam generation in AMS.

Three testing scenarios representing the most typical clinical PACS operations were created. These scenarios were 1. The AMS simulates the generation of an exam, 2. DICOM gateway automatically forwards the exams to the PACS controller, 3. The viewing workstations query/retrieve the exams for clinical review. Scenarios 1 and 3 are active examples, while scenario 2 is a passive example.

4.2 Experiments

During the laboratory experiments, the HCA toolkit was tested with five CT exams, five MR exams, and five CR exams. Table 2 tabulates the test data. Most of the experiments were performed with all three scenarios in the order of 1, 2 and 3. Some experiments only included one or two scenarios performed in random order. The experiments were conducted over several months.

Table 2 Exams tested in the laboratory experiments.

	CT Exams	MR Exams	CR Exams	Total
No. of Exams	5	5	5	15
Total Images	186	288	11	485
Avg No. of Images Per Exam	37.2	57.6	2.2	
Data Size (MB)	93.9	53.0	83.3	230.2

4.3 Examples

Two examples are shown in Figure 4 and Figure 5.

The first example shows the HCA toolkit searching the Auditing Database to generate a HIPAA compliant audit trail based on a given patient name “Jim Johnson”. Two operations need to be performed to generate the audit trail for a patient. First, search the patient in the database with “Patient Name” or “Patient ID”. In this case, “Patient Name” was chosen. Since the toolkit supports wildcard searching, a “J” was typed in the search field to search the patient with the first name starting with “J”. A list of patients matching the search criteria was returned in a new popup window. Second, double click the patient name “Jim Johnson”. A HIPAA compliant audit trail is then generated as seen in Figure 4. As we can see, the line 1 in the table in Figure 4 shows one example of the scenario 1, where the user “PACS” performed a MR exam generation at the modality simulator “ipi-pc2” on “2004-11-15 15:40”. The line 2 shows an example of the scenario 2, where the DICOM gateway stored the exam in the PACS controller. The line 4 shows an example of the scenario 3, where the user “Tech” performed a DICOM query/retrieve of the exam at the viewing workstation “IPI-VIEW”.

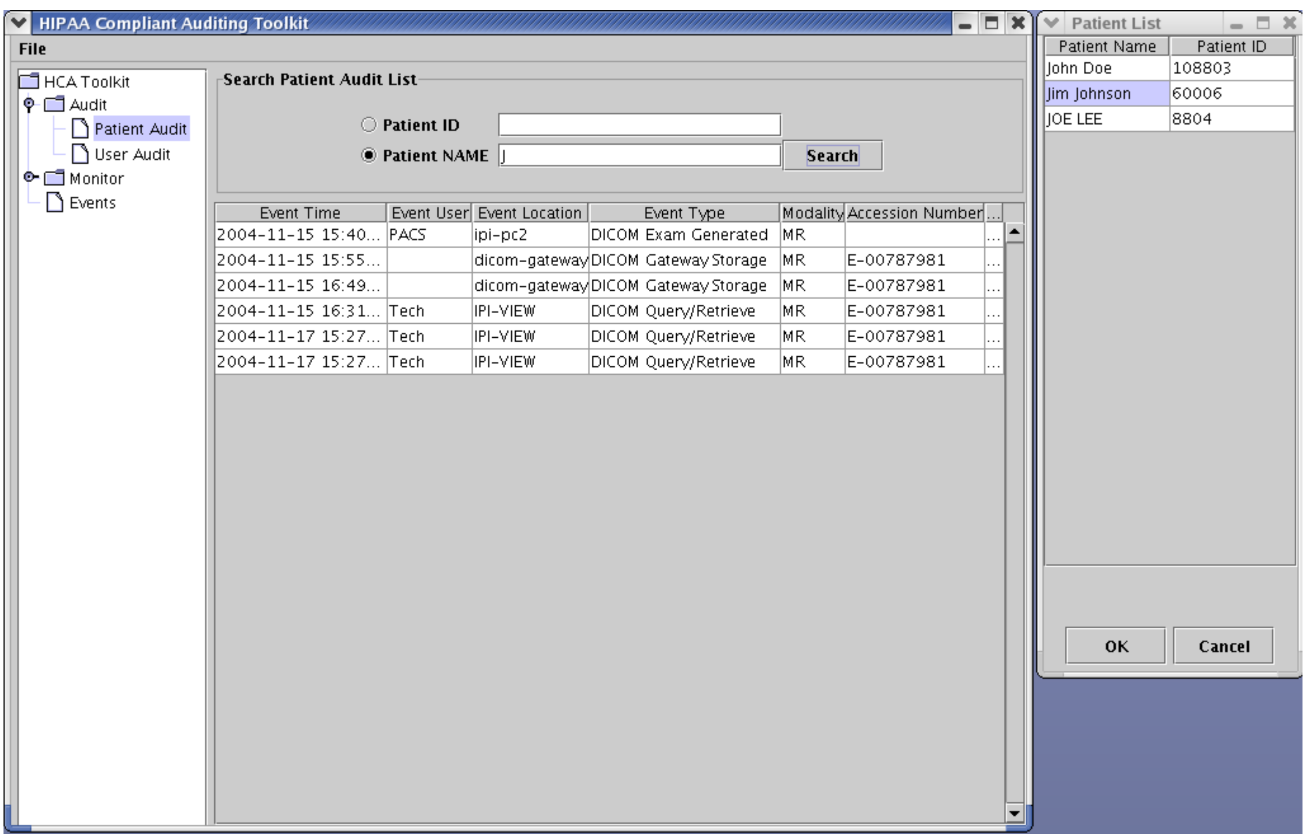


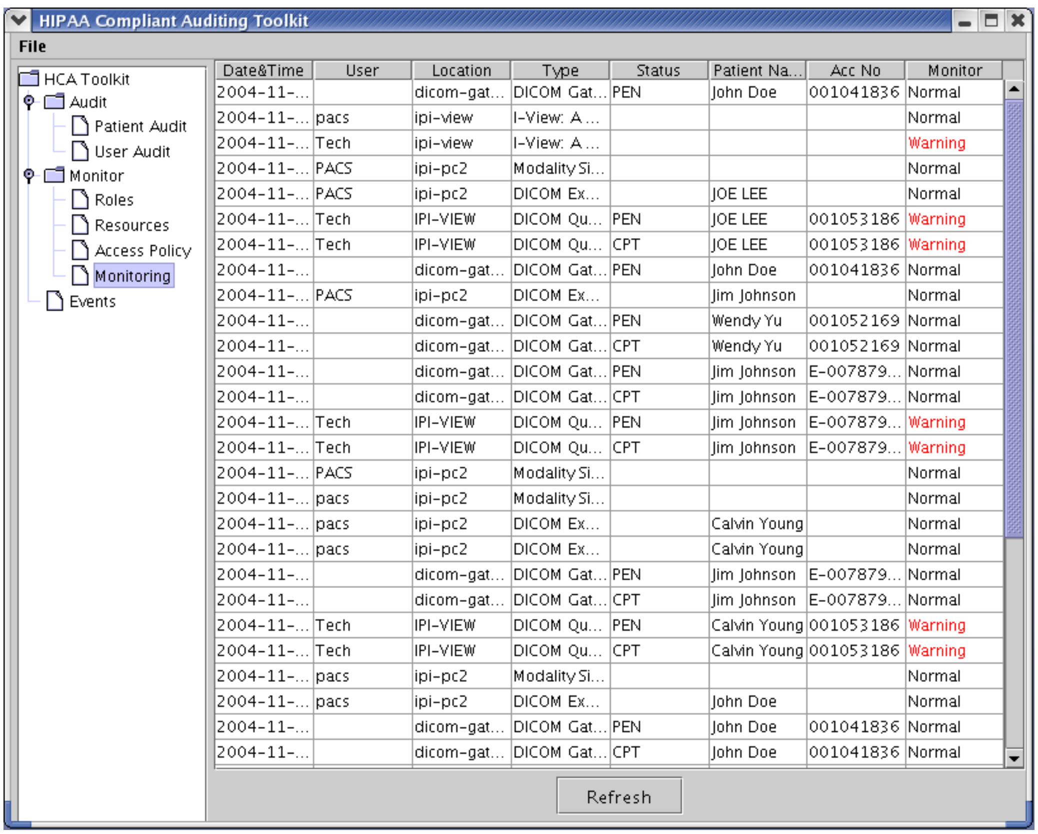
Figure 4 Example of generating the HIPAA compliant audit trails for a patient “Jim Johnson”.

The second example shows the dynamic monitoring of the image data flow of the PACS Simulator. Two types of results, "Normal" or "Warning", were given for each data access event based on the Role-based Policy.

The role-based policy needs to be configured before using the monitoring function. The GUI of the toolkit provides users an interface called "Access Policy" to add, modify, or delete the role-based policy. Table 3 lists the policy used in the experiments. In this case, user "pacs" was defined as "PACS administrator" with the right to perform all the applications in the PACS, whereas user "Tech" was defined as "Technician" with the right to perform the exam generation in the Modality Simulator. Thus, it would be a violation of the policy if user "Tech" performed the exam query/retrieve at the viewing workstation only for radiologists. For example, the "Warning" in the sixth row in Figure 5 indicated a violation happened when user "Tech" retrieved an exam with patient name "JOE LEE".

Table 3 Example of the role-based policy table.

User Name	**Role Name**	**Resources**
Pacs	PACS administrator	All
Clerk	Film Clerk	Viewing workstation
Rady	Radiologist	All
Tech	Technician	Modality Simulator



Date&Time	User	Location	Type	Status	Patient Na...	Acc No	Monitor
2004-11-...		dicom-gat...	DICOM Gat...	PEN	John Doe	001041836	Normal
2004-11-...	pacs	ipi-view	I-View: A ...				Normal
2004-11-...	Tech	ipi-view	I-View: A ...				Warning
2004-11-...	PACS	ipi-pc2	Modality Si...				Normal
2004-11-...	PACS	ipi-pc2	DICOM Ex...		JOE LEE		Normal
2004-11-...	Tech	IPI-VIEW	DICOM Qu...	PEN	JOE LEE	001053186	Warning
2004-11-...	Tech	IPI-VIEW	DICOM Qu...	CPT	JOE LEE	001053186	Warning
2004-11-...		dicom-gat...	DICOM Gat...	PEN	John Doe	001041836	Normal
2004-11-...	PACS	ipi-pc2	DICOM Ex...		Jim Johnson		Normal
2004-11-...		dicom-gat...	DICOM Gat...	PEN	Wendy Yu	001052169	Normal
2004-11-...		dicom-gat...	DICOM Gat...	CPT	Wendy Yu	001052169	Normal
2004-11-...		dicom-gat...	DICOM Gat...	PEN	Jim Johnson	E-007879...	Normal
2004-11-...		dicom-gat...	DICOM Gat...	CPT	Jim Johnson	E-007879...	Normal
2004-11-...	Tech	IPI-VIEW	DICOM Qu...	PEN	Jim Johnson	E-007879...	Warning
2004-11-...	Tech	IPI-VIEW	DICOM Qu...	CPT	Jim Johnson	E-007879...	Warning
2004-11-...	PACS	ipi-pc2	Modality Si...				Normal
2004-11-...	pacs	ipi-pc2	Modality Si...				Normal
2004-11-...	pacs	ipi-pc2	DICOM Ex...		Calvin Young		Normal
2004-11-...	pacs	ipi-pc2	DICOM Ex...		Calvin Young		Normal
2004-11-...		dicom-gat...	DICOM Gat...	PEN	Jim Johnson	E-007879...	Normal
2004-11-...		dicom-gat...	DICOM Gat...	CPT	Jim Johnson	E-007879...	Normal
2004-11-...	Tech	IPI-VIEW	DICOM Qu...	PEN	Calvin Young	001053186	Warning
2004-11-...	Tech	IPI-VIEW	DICOM Qu...	CPT	Calvin Young	001053186	Warning
2004-11-...	pacs	ipi-pc2	Modality Si...				Normal
2004-11-...	pacs	ipi-pc2	DICOM Ex...		John Doe		Normal
2004-11-...		dicom-gat...	DICOM Gat...	PEN	John Doe	001041836	Normal
2004-11-...		dicom-gat...	DICOM Gat...	CPT	John Doe	001041836	Normal

Figure 5 Example of the dynamic monitoring of the image data flow of the PACS simulator based on the role-based policy.

As we can see, the events shown in Figure 5 are all examples of the three scenarios developed. For example, the line 1 is the example of the scenario 2, the line 2 the scenario 3 and the line 5 the scenario 1.

4.4 Discussion

Three testing scenarios have been developed for the laboratory evaluation of the HCA toolkit integrated with the PACS Simulator. As one can see, for the evaluation of the function of generating HIPAA compliant audit trails, the results from the active scenarios 1 and 3 are more important than the ones from the passive scenario 2, because more chances of HIPAA security violation are likely when people are involved. On the other hand, all three scenarios are almost equally important for the dynamic monitoring because any unreported event may indicate a failure in the image data flow. The three scenarios developed are the most typical data flow in PACS and can represent most of the processes in PACS. However, one important process, image storage or archive, was not included in the testing. The image storage is very important to both evaluated functions of the toolkit because the stored image data could be compromised without detection. By integrating the aforementioned image integrity logs with the toolkit, the image storage issue can be solved.

The current functionality of dynamic monitoring of the toolkit is still limited. A more sophisticated monitoring system can be developed using intrusion detection technology and policy-base decision tree approaches. The role-based policy can also be improved to a multiple-factor-based policy instead of a single role-based one.

We have found that some image transmission events were not collocated in the database due to the PACS controller did not record the event in its logs. Therefore, robust and well-formatted logs generated by PACS and imaging modalities would greatly facilitate the functions of the HCA toolkit.

5. CONCLUSION

A HIPAA compliant auditing toolkit has been developed to audit the image data flow of PACS and the imaging modalities connected to PACS for the purpose of generating a HIPAA compliant audit trail when requested. The toolkit also dynamically monitors the data flow in real time based on a role-based policy. The toolkit has been integrated with the PACS simulator for evaluation of the performance of the toolkit in a PACS environment. Some preliminary results have been acquired during the laboratory experiments and presented in this paper.

Overall, the toolkit is just a prototype with limited functionality. However, with the enforcement date of the security portion of HIPAA rapidly approaching, this toolkit can still provide some useful hints to healthcare providers in understanding the HIPAA compliant audit trails of image data access and dynamic monitoring of the data flow in PACS environment. Further development and evaluation of the toolkit is ongoing.

ACKNOWLEDGEMENT

This research is partially supported by a NIH Grant No. R01-LM06270 from the National Library of Medicine.

REFERENCES

1. HIPAA, http://www.cms.hhs.gov/hipaa/hipaa2/general/background/pl104191.asp.
2. HIPAA, http://www.rx2000.org/KnowledgeCenter/hipaa/hipfaq.htm.
3. HIPAA Security Standard, http://www.cms.hhs.gov/hipaa/hipaa2/regulations/security/03-3877.pdf, 2003.
4. HIPAA audit, http://www.datamirror.com/products/liveaudit/.
5. HIPAA audit tool, http://www.axolotl.com/press/20021113/.
6. HIPAA audit tool, http://www.peacefulpackers.com/it_solutions/hs2.htm.
7. H.K Huang., *PACS and Imaging Informatics,* Wiley-Liss, 2004.
8. B.J. Liu, Z. Zhou, H.K. Huang, *A HIPAA-Compliant Architecture for Securing Clinical Images,* SPIE Medical Imaging, in Process, 2005.
9. Z. Zhou, H.K. Huang, F. Cao, B.J. Liu, J. Zhang, G.T. Mogel, *Educational RIS/PACS Simulator*, SPIE Medical Imaging, Vol. 4, pp. 139-147, 2003.
10. Z. Zhou, H.K. Huang, B.J. Liu, *Digital Signature Embedding for Medical Image Integrity in a Data Grid Off-site Backup Archive*, SPIE Medical Imaging, in Process, 2005.
11. Syslog, http://www.loriotpro.com/Products/SyslogCollector/SyslogDataSheet_ENv3.php.

Queue Monitor Toolkit for Tracking of PACS Clinical Workflow

Nelson E. King*[a], Jorge Documet[a], Zheng Zhou[a], Brent Liu[a], H.K. Huang[a]
Image Processing & Informatics Laboratory, Department of Radiology, University of Southern California, Marina del Rey, 90292

ABSTRACT

Expectation of rapid image retrieval and distribution from PACS contributes to increased information technology (IT) infrastructure investments and continuing demands upon PACS administrators to respond to "slow" system calls. Studies show that it is important for computer users to be able to check on the progress of their task via progress indicators (e.g., time left to download file) to know that the computer is still working. By analogy, the ability to provide predicted delivery times to a PACS user may curb user expectations for "fast" response especially during peak hours. Allowing for some periods of slow response means PACS infrastructure do not have to be overbuilt and also reduce time spent by PACS administrators fielding user inquiries on image status. For this condition, a queryable PACS queue monitor is the cornerstone for providing a progress indicator to the user. The typical PACS server holds image file information and destination workstation information in a queue until the RetrieveSend process can send the image. We developed an agent that queries the contents of the PACS RetrieveSend queue in real-time and coded an algorithm to predict delivery time. Delivery time can be predicted from the number and types of images in progress and the download time of prior images that accounts for network load and performance at that time of day. We have developed a PACS queue monitor prototype that is being tested on clinical data using the PACS Simulator at the Imaging Processing and Informatics (IPI) Laboratory of the University of Southern California (USC).

Keywords: PACS system performance, image retrieval delivery time, response time, benchmark

1. INTRODUCTION

PACS designers attempt to minimize the delivery time of images to users at their workstations by incorporating high-speed networks and redundant servers. Researchers have developed monitoring tools to identify bottlenecks[1] between a PACS and workstations on a network. Careful selection of hardware and use of monitoring tools prevent most problems with slow delivery. However, a cost-effective PACS implementation will unlikely be able to accommodate "spikes" of maximum potential (e.g., worst case) usage during peak hours. A cost-effective solution will result in limited periods of peak use where delivery times are slower than normal.

Computer usability guidelines recommend progress indicators to mitigate user expectations when tasks take more than a few seconds such as in downloading files or slow system response due to busy servers[2,3]. A queue monitor provides the PACS administrator with a tool to examine bottlenecks or respond to user queries about a "slow" system. In addition, a PACS Queue Monitor could inform a user that the requested retrieval is pending and the exam will take several minutes to download. This user could then choose to take a short break or temporarily switch to other tasks rather than waiting. While some current PACS can provide the administrator with the status of a series (e.g., pending, sending), none to our knowledge can predict delivery time.

The paper begins with a description of our laboratory testbed that allowed us to identify the PACS parameters to be used in a prediction algorithm. Section 3 summarizes the simulation data that illustrates the interaction among the parameters. The results and implications to PACS designers and system implementers are discussed in Section 4.

*linking@acm.org; phone 1 310 448-9440; fax 1 310 444-9441; www.ipilab.org

2. METHODOLOGY

Our goal was to identify the PACS parameters that play a part in predicting the delivery time of an image under conditions of queuing such as during peak usage in a clinical setting. These parameters would be used to develop the algorithm to predict delivery time. We simulated a clinical environment in which the requests to retrieve clinical images from our laboratory's PACS Simulator resulted in requests being queued. The behavior of this clinical testbed could then be examined under various conditions. The prediction algorithm would be derived from the performance data for this PACS. The study assumption was that a slow PACS Server can be saturated with fewer requests for studies from reading workstations compared to a faster PACS. Saturation of a PACS capability would result in slow delivery times. A faster PACS could eventually be saturated with more study requests and larger studies (e.g., hundreds of slices). So the methodological premise of a queue monitor should be applicable to a wide range of conditions although the actual parameters and performance would vary dramatically.

A laboratory testbed that modeled a clinical setting allowed us to analyze PACS parameters without disrupting an actual clinical operation. Actual clinical images were requested at 5 to 10 minute intervals to represent the typical workflow of a radiologist. Sometimes, multiple studies were requested at about the same time to reflect the condition when prior studies had to be retrieved for comparative purposes. Section 2.1 describes the laboratory testbed and the workstations used. Section 2.2 summarizes the clinical data used in the study. Section 2.3 discusses the prediction algorithm.

2.1 Workstation Testbed Description

A clinical setting was simulated in our IPI Laboratory. A PACS and several workstations running DICOM client software were placed on two different network segments as shown in Figure 1. Three of the workstations were placed on the same network segment as the PACS. Two other workstations were placed on a different network segment to represent the situation of a reading room in a different floor or building.

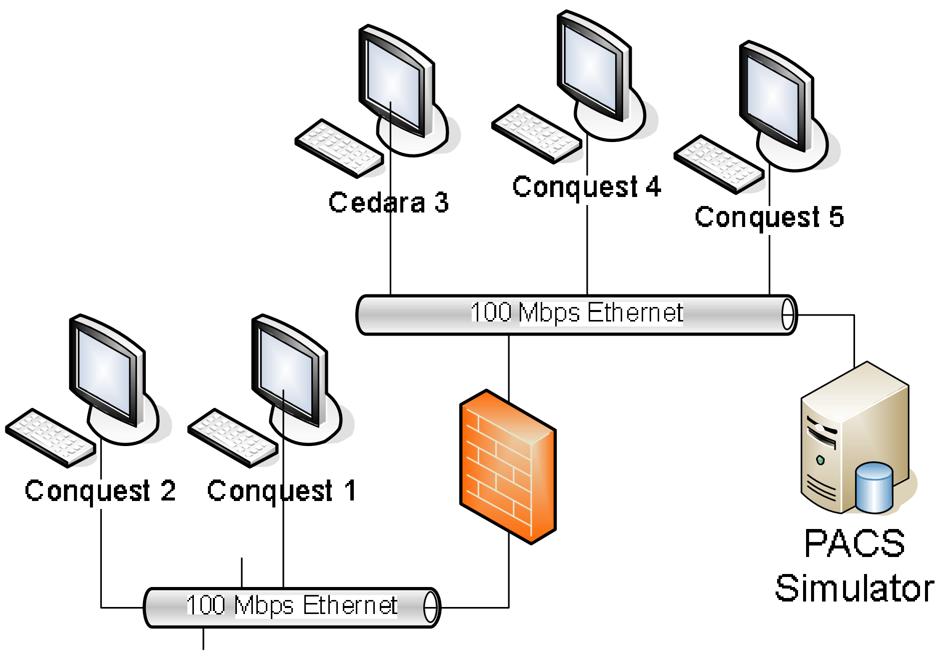


Figure 1 Physical Layout of Queue Monitor Testbed

The PACS Simulator in the IPI Laboratory was used for the PACS. The software runs on an Ultra 2 Sun machine using SunOS 2.X. The database was Oracle 9.x. The PACS Simulator is DICOM-compliant using the standard patient – study - series data model. There are four internal queues to which study requests are sent. A slower PACS was chosen so that fewer workstations could generate enough requests to saturate the PACS and cause some study requests to be delayed. Papers by Law et al [4] and Zhou et al [5] provide a more complete description of the PACS Simulator. There were five Windows-based computers acting as reading workstations configured as shown in Table 1.

Workstation Name	DICOM client	OS	CPU	Network	Database
Conquest1	Conquest 1.4.7	XPP	P4 2.8 Ghz	Same as PACS	Built-in dbIII
Conquest2	Conquest 1.4.7	XPP	P3 0.7 Ghz	Same as PACS	Built-in dbIII
Cedara	Cedera I-View 5	W2K	P4 1.3 Ghz	Same as PACS	Application-provided
Conquest4	Conquest 1.4.7	W2K	P4 2.8 Ghz	Subnet via firewall	Built-in dbIII
Conquest5	Conquest 1.4.7	W2K	P2 0.4 Ghz	Subnet via firewall	Built-in dbIII

Table 1 Workstation Specifications in Queue Monitor Testbed

2.2 Data Acquisition

The goal in generating data was to saturate the PACS server with send requests so that some studies would be queued. The sequence of requests and time between requests should be representative of a clinical setting. This precludes selecting a large number of exams via a wildcard search and requesting all the resultant studies to be sent all at once. Such an approach would greatly distort the data particularly with our PACS Simulator. Studies transferred under a single query are all sent to the same queue within the PACS Simulator. A backlog in the queue is created since processing is done sequentially. The throughput is even slower than having all four queues active since the other queues available on the PACS are not used. Thus it was necessary to manually query and retrieve every study by name. The clinical image files on each workstation were deleted and the DICOM server acting as a client was cleared and re-initialized so that all the requested images would have to be processed.

The data was gathered on a weekend so that there would be no other network traffic or use of the PACS that would confound the results. On the first workstation, a request was made by querying the PACS for a specific study. Once the results of the search are returned from PACS, a transfer was requested from the PACS to this workstation. The second workstation then queried a different exam based on study number. Once again the transfer was requested immediately after the query result was returned by the PACS. The same procedure was done for workstations 3 (e.g., Cedara), 4, and 5. The researcher would then return to workstation 1 and query another exam. The transfer was initiated and the process repeated on workstation 2. Sometimes one or two additional exams were requested on a client to simulate the retrieving of prior studies for comparison. Table 2 tabulates the breakdown of 131 studies that were requested over a 40 minute period. Nearly 2 gigabytes of data were transferred and almost 4500 individual DICOM images were sent. There were studies from modalities of computed radiography (CR), computed tomography (CT), magnetic resonance (MR), a single ultrasound (US) study and a highly compressed CR study containing two files.

Workstation	Total MB	Total Images	Studies	Modality				
				CR	CT	MR	US	CR (comp)
Conquest1	372	920	25	3	13	7	1	1
Conquest2	355	918	25	3	13	7	1	0
Cedara	455	892	29	7	14	6	1	1
Conquest4	390	931	25	5	11	8	1	1
Conquest5	444	803	26	7	11	6	1	2
Total	2012	4464	131	25	62	34	5	5

Table 2 Distribution of Studies by Workstation Destination

After all of the requested studies were processed on the five workstations, the queue transaction table of the PACS Simulator was copied using an SQL client and an ODBC connection. The data preserved in the queue table included DICOM requestor and recipient, queue used, time of retrieval request, start time of retrieval, finish time of retrieval, and study identifier.

2.3 Algorithm Description

The prediction algorithm must take into consideration location on the network (i.e., segment), computer configuration, DICOM application software, load on the PACS, and clinical study specifics. All of these factors contribute to the transfer rate for a workstation.

Network segment affects speed of transfer since the file may pass through any combination of firewall, switch or hub. In addition, each network segment may have a different speed (e.g., 10 Mbps versus 100 Mbps).

Computer configuration determines the capacity to receive images. Since this testbed used DICOM servers as the client, there is processing that takes place to process the DICOM header and store the file on the hard disk. A computer may be slower due to processor speed, data bus, memory type and size, and hard disk performance. A slower computer takes longer to receive the images since analyzing the DICOM header is a database activity and folders must be created on the hard disk.

Workstation software varies in choice of database and processing algorithms. Some open multiple threads while others rely upon buffers. The connection approaches to the DICOM archive will also vary.

Load on the PACS reflects the capability of the PACS hardware and software to send multiple studies at the same time. There may be limitations on throughput due to network card, buffer or thread implementation, memory and processing power.

Clinical study specifics include modality, number of images or slices and compression.

Transfer rate reflects all of these factors. This leads to the following equations with variables identified in standard Java notation in which the first letter of interior words are capitalized:

1) Predicted Finish Time of Study (predictFinishTime)

predictFinishTime = predicted start time of sending a study (predictSendStart) + sending time based on transfer rate of the client on its network segment given the particular study specifics (transferRateClient) given the load on a PACS.

2) Predicted Start of Sending Time of Study (predictSendStart)

predictSendStart = cumulative time to complete sending studies 1 to N that are already in the same queue based on transfer rate of the client on its network segment considering study specifics and overall load on PACS (e.g., transferRateClient).

3) Transfer Rate of Specific Client (transferRateClient)

transferRateClient = Transfer rate in Mbytes per second for a particular DICOM client application running on a particular computer on a particular network segment receiving study specific images (e.g., modality) with a particular load on the PACS. This is a measured average value in Mbytes per second.

3. DATA

A log from the PACS Server was generated from the simulation conducted with five workstations over 40 minutes (46,000 to 48,500 second timeline) requesting 131 studies. The analysis of this log is presented in this section. Performance characteristics between workstations are contained in Section 3.1. Section 3.2 describes the impact of modality upon workstation performance. The performance impact of a backlog in one of the PACS queues is described in Section 3.3. The impact of a PACS sending a number of studies to different clients is shown in section 3.4.

3.1 Workstation Performance

Differences among workstations depend on configuration parameters of the prediction algorithm. These include the computer components of the workstation, location of the network, and application software. The table below provides an overall comparative performance showing relative differences between the workstations. The average time for processing a study in the first column ranges from 10.2 seconds on Conquest4 to 53.6 seconds on Conquest5. This is the time from when the PACS received a request to the time that PACS records in the log that the study has been sent. The average transfer rate is measured in megabytes per second as shown in the second column, where the total size in megabytes of the study is divided by the time from when the PACS begins transmitting the images to the workstation until the study is completed. The third set of measures is the average start delay in seconds as shown in column 3. This is the time from the PACS receiving a request to the time the transmission of images has begun. The PACS Simulator used in this simulation uses four queues. There were 22 studies that became backlogged in a queue. Most of these 22 studies had lengthy delay times simply because of waiting for the previous study to finish. The average delay in seconds excluding these 22 studies is found in column 4.

Workstation	Average Time for Study (sec)	Average Transfer (MB/s)	Average Start Delay (sec)	Average Start Delay w/o Backlog (sec)
Conquest1	53.6 (22.5)	0.34 (0.15)	7.0 (3.5)	7.0 (3.4)
Conquest2	48.1 (12.6)	0.38 (0.17)	7.1 (5.9)	5.2 (2.9)
Cedara	23.5 (12.3)	2.01 (2.26)	7.8 (6.7)	5.4 (3.4)
Conquest4	10.2 (5.7)	4.21 (1.94)	6.5 (5.5)	5.3 (3.4)
Conquest5	17.0 (8.2)	1.72 (1.16)	6.2 (3.8)	5.9 (3.3)
All	29.8 (21.5)	1.77 (2.01)	6.9 (5.2)	5.7 (3.3)

Table 3 Transfer rates by workstation destination (standard deviation in parentheses)

The numbers in this table are comparable at a high level. Table 1 showed there was a similar number of studies and distribution of modalities sent to each workstation. The sequencing of requests during the simulation was also similar. So the large standard deviations shown in parentheses for average study time and average transfer rate indicate that further explanation is necessary as provided in Sections 3.2 to 3.4.

The average study time and average transfer rate show several differences at this level of granularity. First, network configuration makes a big difference. The two workstations that sat outside an internal firewall were quite slow. Conquest1 had the same hardware configuration as Conquest 4 but four times slower in terms of average study time. Conquest2 had slightly older hardware than Conquest5 but nearly three times slower in terms of average study time. Second, the Cedara workstation was surprisingly slow suggesting that application software makes a difference.

The actual sequence of studies that were transferred is shown in Figure 2. The actual study times for each workstation are plotted beginning with Conquest1 at the bottom. The delay time and length of transmission by modality is also shown. The differences in workstation performance are clearly seen by the relative length of study time (delay and sending) between the five workstations. When several studies are requested within a short time of each other from the same client there is an unexplained degradation of performance for that client. For example, Conquest1 at approximately 47250 seconds is receiving four studies at the same time.

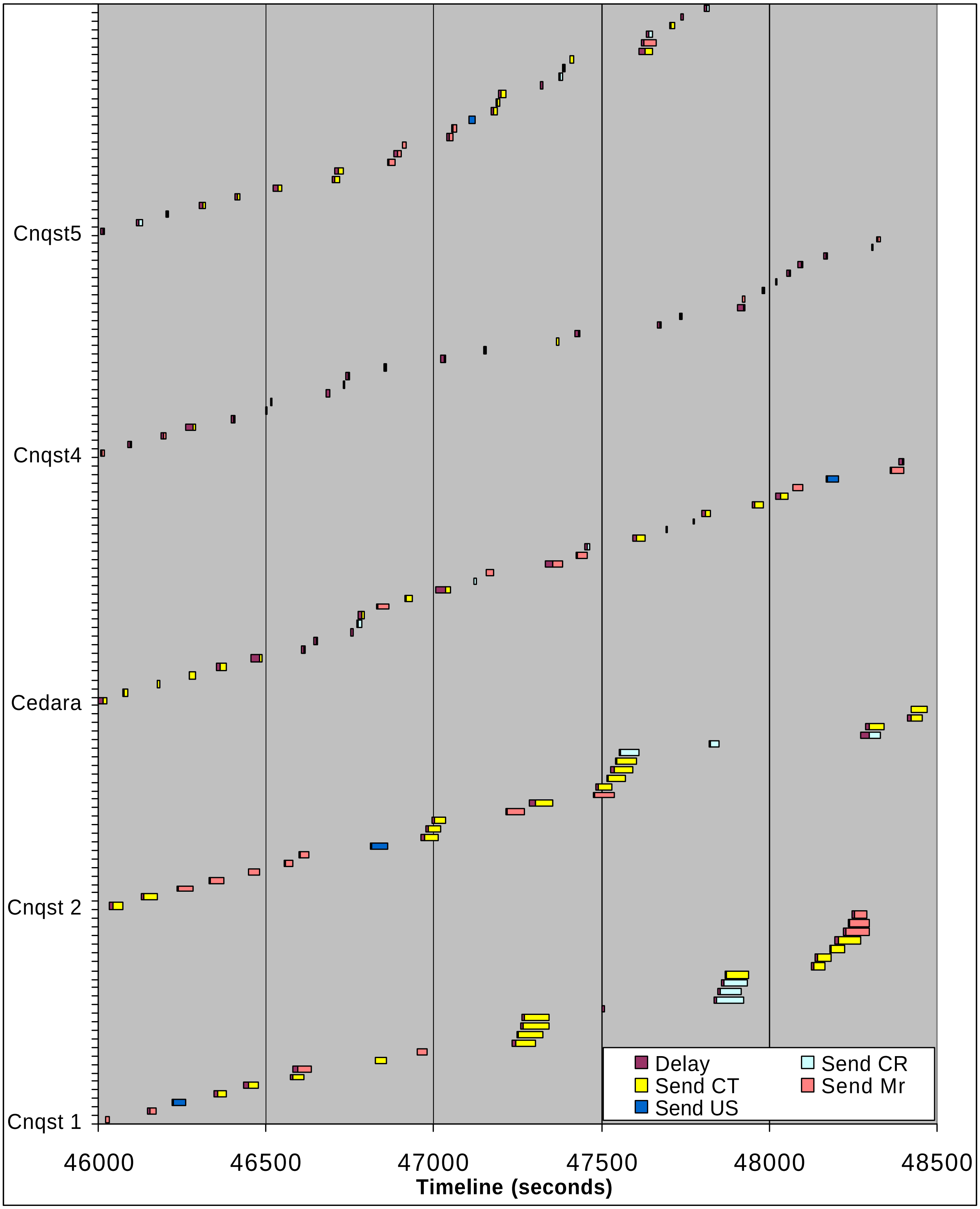


Figure 2 Delay and Transfer Time to Five Workstations (modality indicated)

3.2 Workstation and the Impact of Modality

Modality determines the study specifics especially image size (Mbytes) and number of images. A CR study has large images (e.g., 8 MBytes) but only a few images. A MR study has small images but many slices (e.g., 50 to 60 in this study). Since a DICOM server/client must process the header if stored locally, more workstation resources are spent processing an MR study than transferring the image. This is why the transfer rates for CR are typically higher than CT and MR. This is shown in the average transfer rate column of Table 4. The anomalous results of Conquest1 for CR reflect a relatively rare situation in this simulation where three studies were retrieved simultaneously. Section 3.4 explains how simultaneous transfers decreased throughput by nearly one-half.

	Transfer Rate (Mbytes/sec)		
Workstation	CR	CT	MR
Conquest1	0.3 (.03)	0.4 (0.1)	0.3 (0.2)
Conquest2	0.6 (0.2)	0.4 (0.1)	0.2 (0.1)
Cedara	5.1 (2.5)	1.4 (0.8)	0.3 (0.1)
Conquest4	6.4 (1.2)	5.0 (0.6)	2.2 (1.0)
Conquest5	3.3 (0.6)	1.7 (0.4)	1.7 (0.4)
Average	3.8 (2.6)	1.7 (1.7)	0.8 (1.0)

Table 4 Transfer rates by modality by destination

3.3 Backlog in Queues

A queue backlog occurs when there are more requested studies than queues to process the requests. Figure 3 shows the concept of a queue backlog taken from the study data from time 47250 to 47400. Conquest 1 workstation requested a study at 47247. This request is assigned to Queue 2 finishing at 47326. The next study request assigned to Queue 2 comes from Conquest 5 workstation. The request occurs at 47318 but cannot start sending until 47327 because each queue processes requests sequentially. Similarly, Conquest 2 workstation requests a study at 47283, which finishes at 47354. The study requested by Cedara workstation at 47334 must wait until 47354 to begin sending. This figure also shows that this PACS distributed three requests from Conquest 1 workstation made within 20 seconds to queues 1, 2, and 4.

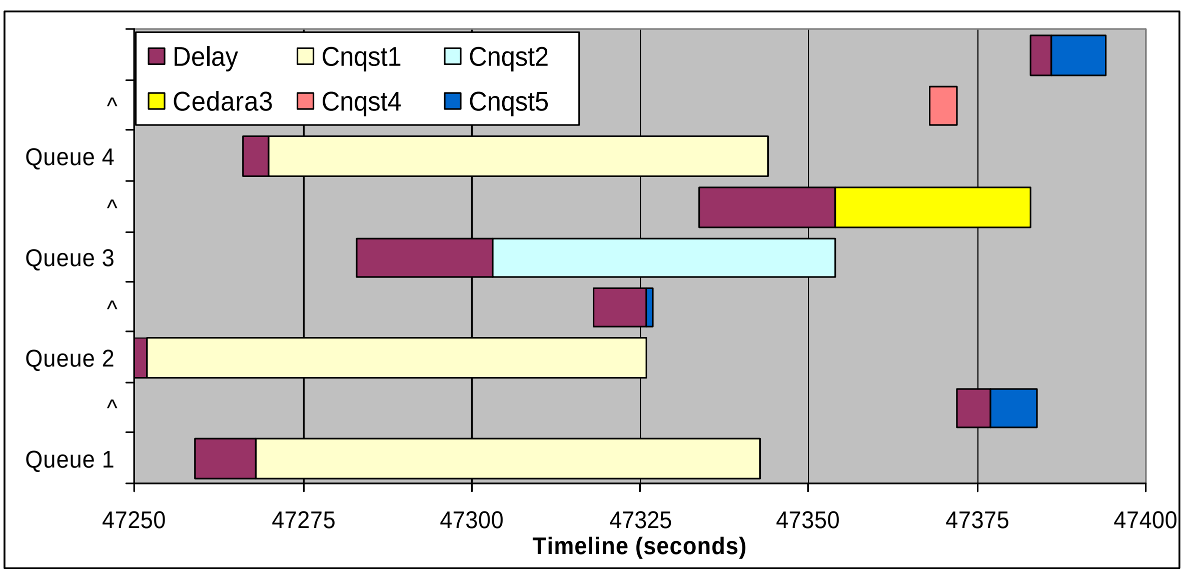


Figure 3 Delay Time in Sending Study by Queue

The figure below with the logarithmic delay scale shows that the studies with delay times over 15 seconds consisted of the studies that had a queue overlap as indicated by the darkened circles. That is why the average delay time drops from 6.9 to 5.7 seconds when these 22 studies in which the queue overlap are excluded. The standard deviation drops from 5.7 to 3.3 seconds. The queue overlap cannot be explained by workstation (shown below) or choice of queue. There were other implementation factors in the simulation that could not be isolated with the data available in the PACS log. Thus the delay time will be assigned a value of 5.7 ± 3.3 seconds.

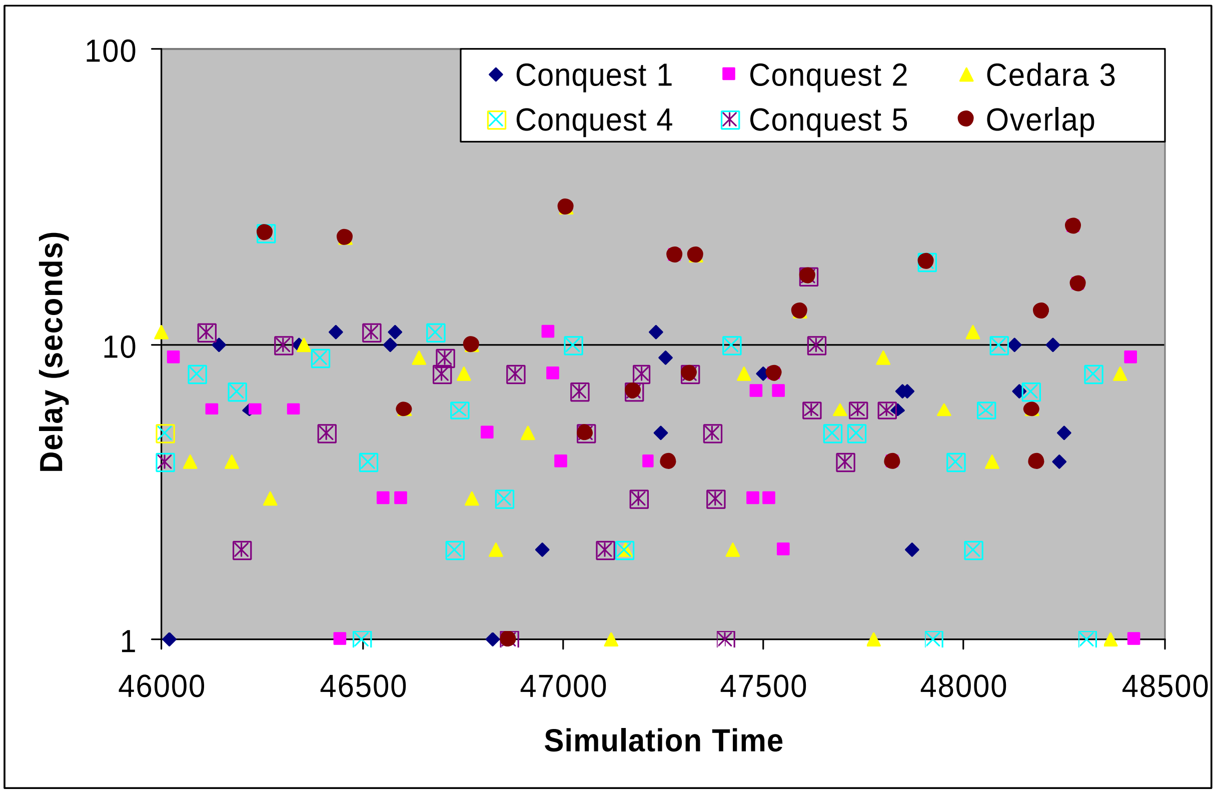


Figure 4 Delay versus Simulation Time

3.4 Simultaneous Studies in Transmission

What happens to overall throughput when multiple studies have been requested in the same time interval? Table 5 illustrates the queue backlog effect. There are five studies in process at time 47325. Four are transmitting and one is waiting in queue 2. The effect on throughput for this PACS Simulator and the configuration of workstations is observable.

The table below tabulates the number of studies that are being processed by the PACS while other studies are also in process. The arbitrary sequence among workstations to request a study resulted in Conquest1 and Conquest2 having the majority of studies in parallel with other requests. This behavior explains the slow throughput beyond the explanation of files passing through the firewall. In contrast, the workstations on the same network segment as the PACS had less than one-third of the studies competing with others for bandwidth.

Workstation	Studies	# studies simultaneous with others	% of Studies	Less than 3 in process	3 or more in process	Effect on Throughput
Conquest1	25	22	88%	11	11	Substantial
Conquest2	25	19	76%	16	3	Moderate
Cedara	29	8	28%	7	1	Not observable
Conquest4	25	5	20%	2	3	Not observable
Conquest5	26	5	19%	4	1	Not observable
Total	131	59	--	40	19	

Table 5 Simultaneous Transfer by Destination Workstation

The impact on transfer performance was substantial for Conquest1. The figure below shows the number of other studies in addition to the one being transferred on the x-axis. The transfer rate of the study is shown on a logarithmic scale. Partial overlaps in transmission were computed. For example, two studies being processed in parallel for only half the time is considered 0.5 additional studies in process. The scale from 1 to 10 is preserved since the peak transfer rates are 8 to 9 Mbytes per second for CR. Table 4 in Section 3.2 tells us that average transfer rates of 5 or 6 Mbytes per second are possible on the same network (e.g., Conquest 4). The figure shows that there is a log-linear relationship on throughput as the number of additional studies increase.

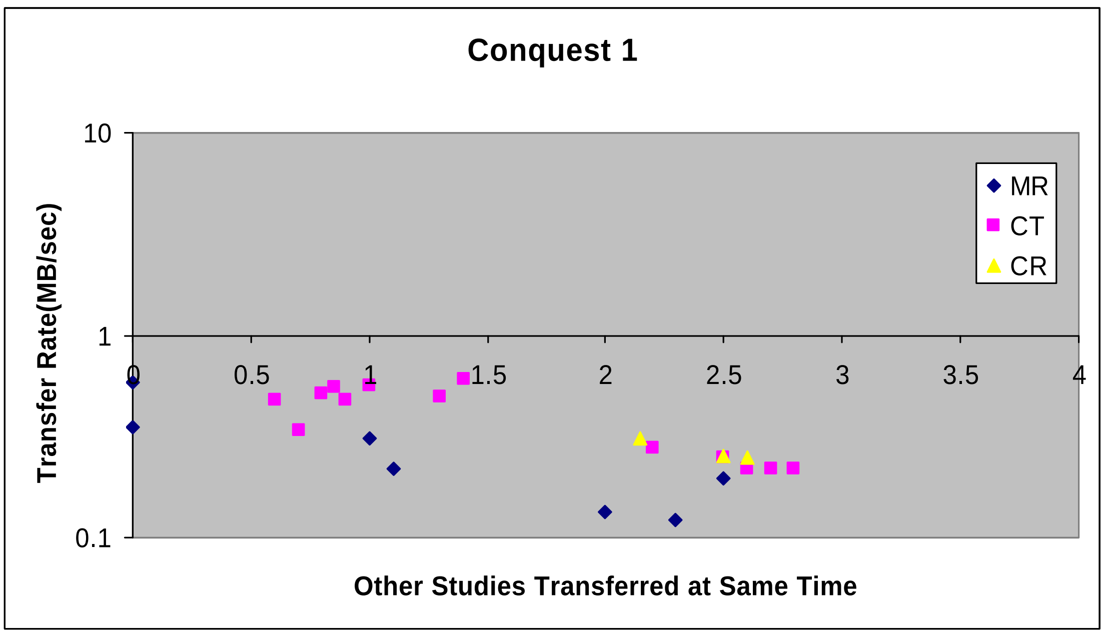


Figure 5 Transfer rate of Studies versus Number of Simultaneous Studies

4. RESULTS

The simulation of workload on a PACS in a clinical setting resulted in some important findings. Several parameters are of importance for at least our laboratory configuration as discussed in Section 4.1. Values for these parameters are used in the prediction algorithm as listed in Section 4.2. Section 4.3 discusses the implications for clinical systems. Our research plans are summarized in Section 4.4.

4.1 Important Parameters

The key parameters identified in this study include network configuration, client software, modality, simultaneity of studies, and a queue backlog.

A workstation placed on a network segment that requires passing through a network device such as a firewall appears to have a large impact on performance. Transfer times seem to be at least double. Oddly, network location did not seem to have an effect on delay time. Performance impact of a network segment needs to be included in a prediction algorithm.

The performance of workstations Conquest4 and Conquest5 compared to the Cedara workstation were very different. All three workstations were on the same network segment. The means in which DICOM transfers were processed for the two different software packages is the likely explanation.

Modality type must also be included in a prediction algorithm due to the differences in transfer rate that result from varying image size and number of images.

The dramatic impact of processing simultaneous studies shown in Section 3.3 is justification for a queue monitor. While reaching the point of simultaneous studies in a fast PACS may not occur often, there may be peculiarities in the way that buffers or threads are handled which can limit throughput under peak load conditions as was done in this research.

Delay time appears to be a function of the PACS Controller software. The dispersion in delay time reflects that some studies are not delayed for very long before sending begins while others take a long time. This dispersion occurs despite controlling for the long delays associated with a backlog in a queue as discussed in Section 3.3. The only factor not controlled was the processing load of queries being sent to the PACS. There was no evident pattern in the data controlling for workstation, network segment, and study parameters including modality, image size, and number of images. The only characteristic of this PACS Simulator is that a study backlogged in the queue will have no delay once the sending of the prior study is finished.

4.2 Prediction Algorithm

The typical prediction algorithm would be the normal study request in which the PACS is not overloaded. The algorithm is:

predictFinishTime = predictSendStart (= delay time) + sending time based on transfer rate of the workstation on its network segment given the particular study specifics (transferRateClient) given the load on a PACS.

For example, Conquest5 requests a CR study consisting of two 8 MByte images. A value of 5.7 ± 3.3 seconds will be used for the delay. Table 4 in Section 3.2 indicates a transfer rate of 3.3 Mbytes/sec under typical conditions. In the case of Conquest5 during this simulation, about 20% of the studies were done simultaneously with others. For the condition of a lightly loaded PACS Simulator, the predicted finish time is 5.7 + (16/3.3) ~ 10 seconds. However, what happens if the PACS is busy and the request from Conquest1 arrived before the request from Conquest5. The request is for a 200 slice MR (about 27 Mbytes) which means a download at 0.3 Mbytes/sec. So after 90 seconds (27/0.3), the backlogged request can begin. The predicted delivery time would have been 95 seconds.

4.3 Implications for a Clinical System

Testing a PACS with peak load conditions as done is this research will be difficult in a clinical environment. The speed of current PACS hardware makes it difficult to flood the PACS with requests so that a backlog in the queues is created.

One approach during acceptance testing would be to transfer very large studies with 1000's of images from as many computers as practical. This research showed that there is some interaction between DICOM client software and the PACS. Slow performance may be attributable to the interaction between the client software and the PACS. Our PACS placed requests for multiple studies (e.g., using wildcard search) in the same queue presumably since only a single connection was made. We have concerns that increasing use of web access means most of the requests are coming from clinician web browsers rather than reading workstations. Questions should be asked about the connection established between the web application server and the PACS. Persistent connections might be assigned to the same buffer or thread.

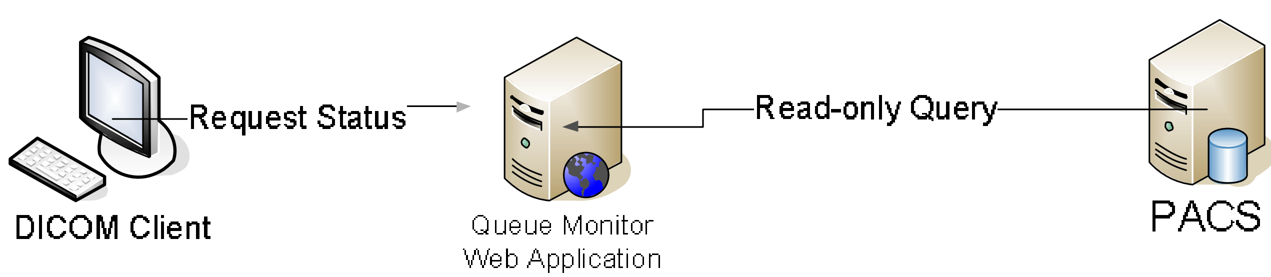
A properly designed clinical system means a queue monitor will not be used frequently. However, unplanned usage is possible such as storing digital mammography in the PACS. So a contingency plan is necessary – buy more hardware or fine-tune the existing system. A queue monitor helps to understand the behavior of the system such as our experience of workstation performance degraded not only by the network configuration but the load on the PACS.

A conceptual model for a queue monitor is shown in the figure below. In a clinical setting, the only time that a radiologist or clinician would want to know about the status of a study is when the system is responding slowly. Ideally, the user checks for themselves rather than calling the PACS administrator. Figure 6 shows that a user would request the status of a study through a queue monitor web application. A web application would provide the most versatile solution to adapt to the specific PACS architecture in different clinical settings.

The implementation of a queue monitor requires that the PACS software store sufficient information to assess the state of the queues. This information does not need to be in one table of the database. However, the information on the queues must be queryable in real-time (albeit read-only) from a queue monitor application. Unfortunately, a prediction algorithm in the form of an SQL query would have to be developed for every PACS since there are vast differences in implementation and database structure. A read-only query minimizes the load on the PACS server. Ideally PACS developers would place all of the critical queue information in a single table. By doing so, a read-only table copy would be possible rather than the computation-intensive joins needed in an SQL query.

4.4 Future Research

Our future research will be directed towards building a web application so that the prediction algorithms can be processed in real time. Other PACS queue implementations will also be studied so that a generalizable prediction algorithm can be written. Since this study was conducted on a PACS that utilizes older hardware, future testing on faster PACS hardware will require a reading workstation simulator. The simulator must provide for multiple DICOM clients residing on different network segments that request studies individually over a time interval consistent with radiologist reading habits. While some load on the PACS can be created with larger studies (100's of slices instead of 10's), simulating peak load conditions will require the use of multiple DICOM clients. The workstation simulator will also be useful in establishing benchmarks for a particular PACS. We expect each PACS to have vastly different performance behavior under peak load conditions because of the means in which simultaneous studies are processed.
Our prediction algorithm considered average transfer rates. In reality, there will be a large number of workstations whose configuration will be known including hardware components, software and network location.



A self-learning algorithm seems ideally suited for a queue monitor application. There are many know variables about each workstation and the type of study being sent. Each of these sets could be associated with measurable PACS parameters such as processor and memory utilization as well as queue status (e.g., simultaneous studies in process). The transfer rates could be periodically updated using all of the previous studies as a training set.

5. CONCLUSIONS

We demonstrated the feasibility of a PACS Queue Monitor algorithm so that performance of our PACS could be monitored under peak loads. Fast performance from a PACS requires more than just fast networks and fast workstation hardware. The queuing implementation and the connection established between a PACS and a workstation will also be factors. Our prediction algorithm can predict approximate delivery times but further understanding of PACS behavior under peak load conditions is necessary to refine the algorithm. Implementation in a clinical setting requires further investigation into integration of a Queue Monitor with user workstation and a platform for handling multiple queries (e.g., web application) in real-time.

REFERENCES

1 Paul G. Nagy, Mark Daly, Max Warnock et al., Radiographics 23 (3), 795 (2003).
2 J. Nielsen, *Usability Engineering*. (Morgan Kaufmann, San Francisco, 1994).
3 Dennis F. Galletta, Raymond Henry, Scott McCoy et al., Journal of the Association for Information Systems 5 (1), 1 (2004).
4 M.Y.Y. Law and Z Zhou, Computerized Medical Imaging and Graphics 27, 147 (2003).
5 Z. Zhou, H.K. Huang, B.J. Liu et al., SPIE Medical Imaging 5, 372 (2004).

Performance Evaluation of Secured DICOM Image Communication with Next Generation Internet Protocol IPv6

Fenghai Yu*, Jianguo Zhang*+, Xiaomeng Chen*, H. K. Huang+
*Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai
+Image Processing and Informatics Laboratory, Departments of Radiology and Biomedical Engineering, University of Southern California, Los Angeles

Abstract

1. INTRODUCTION

PACS (picture archiving and communication system) requires high-speed networks to transmit large image files between components. In case of intranet, that is, PACS within medical center, Gbits/s switches and Mbits/s connections to workstations are almost standard in most hospital and university network infrastructures. Their transmission rates, even for large image files, are acceptable for clinical operation. However, in case of the Internet, image data must be transmitted between hospitals and campuses. There are two important issues usually to be addressed when talking about medical image transmission over public Internet: first issue is cost-effective, the second is security. Current low-cost commercial WAN (wide area network) is too slow for medical imaging application, whereas high-speed WAN is too expensive for cost-effective use. To solve the first problem, the Next Generation Internet (NGI) technology with new communication protocol IPv6 emerges as a potential solution for low-cost and high-speed networks for image data transmission [1]. For second one, there are three organizations issued guidelines, mandates, and standards for image/data security: The ACR (American College of Radiology) Standard for Teleradiology, adopted in 1994, defines guidelines for qualifications of both physician and non-physician personal, equipment specifications, quality improvement, licensure, staff credentialing, and liability [2-5]. HIPPA of 1996, Public Law 104-191, which amends the Internal Revenue Service Code of 1986, requires certain patient privacy and data security. Part 15 of the DICOM Standard specifies security profiles and technical means for application entities involved in exchanging information to implement security policies (PS 3.15-2001) [6].

Despite these initiatives, to our knowledge, there have not been active systematic research and development efforts in the medical imaging community to seriously tackle the secured DICOM image communication over the NGI and evaluate their performance compared to the current DICOM communication with different secured methods and algorithms used to encrypt the image data. In this presentation, we first give the implementation of DICOM image communication library with IPv6 by using open source code; secondly, we discuss the two major security methods, IP Security (IPSec) and Secured Socket Layer (SSL), used in medical data communication; Third, we design the testing schema of IPv6/IPv4 enabled DICOM image communication with different security methods, different algorithms and operating systems, and evaluate the testing results; Finally, we discuss the impacts of our research results to clinical application.

2. IPV6/IPV4 SECURE DICOM COMMUNICATION SOFTWARE

Most network applications and protocols (Clinet/Server, Web, http, DICOM, …) used in the Internet or intranet are developed based on TCP/IP. They are partitioned into three layers, i.e. application layer, transport layer, IP layer, link layer and physical layer. TCP/IPv4 were developed in last centenary 80 years, it has a lot of problems such as: (a) address shortage; (b) IPSec is an add-on; (c)problems of multicasting; (d) Complicated header; (e) Fragmentation/retransmission problems; (f) poor QoS (Quality of Service); (h) Inability to handle large frames; (i) Limited auto-configuration support (needs DHCP).

TCP/IPv6 were developed for Next Generation Internet, and designed to solve many of the problems of the current version of IP (known as IPv4) with regard to address depletion, security, autoconfiguration, extensibility, and more. IPv6 include many associated protocols, such as IPSec, ICMPv6, etc. IPv6 has some features as follows: (a) Larger Address Space; (b) Aggregation-based address hierarchy and efficient backbone routing; (c) Efficient and Extensible IP datagram, such as no fragmentation by routers, 64 bits field alignment, and simpler basic header; (d) Auto configuration; (e) Security; (f) IP Renumbering as part of the protocol. Figure 1 shows the architecture of TCP/IPv4 and TCP//IPv6. From Figure 1, we see that the major difference of TCP/IPv4 and TCP/IPv6 is in IP layer.

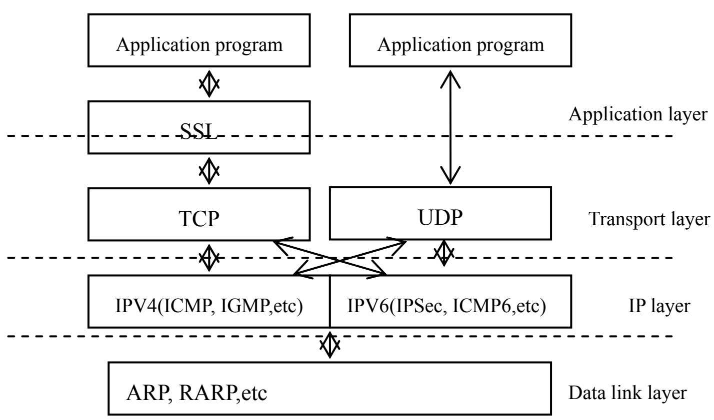


Figure 1 TCP/IP protocol family architecture

Most medical image communication uses DICOM Standard to transfer the image data or objects between the imaging modalities, PACS archiving server, workstations, and other components. In order to enable medical image transmitting through NGI with IPv6, it needs to develop DICOM communication library on top of TCP/IPv6, and also make it compatible with IPv4. We choose open source DICOM software CTN (Central Test Node) library, developed by the Mallinckrodt Institute of Radiology (MIR), to implement IPv6/IPv4 enabled DICOM communication software on different operating systems (Windows/Linux) by modifying its low level communication modules. To achieve this purpose, Two works have to be done: (1) installing the IPV6 protocol stack in computer and attaching it to operating system; (2) Modifying DICOM communication library to adapt to new TCP/IPv6 protocols, and update the DICOM service and application software. For first issue, we only need to install the IPv6 stack software and do some configurations, such as assigning IP address, configure the tunnel and configure router, since most operation systems, such as Windows XP, Linux and Solars have supported the IPv6. For the second issue, we need to replace the original TCP/IPv4 socket functions with RFC standard TCP/IPv6/v4 compatible socket functions, provided by operating systems, in CTN library, re-compile the library, and link it to DICOM services and applications. So, we got four basic IPv6/IPv4 enable DICOM communication services and applications:

(1) DICOM C-Store SCU/SCP
(2) DICOM C-Find SCU/SCP
(3) DICOM C-Move SCU/SCP
(4) DICOM QUERY/RETRIEVE SCU/SCP

3. DICOM IMAGE SECURITY COMMUNICATION AND IMPLEMENTATION CONSIDERATION

DICOM Standard Part 15 (PS 3.15-2001) provides a standardized method for secure communication and digital signature. It specifies technical means (selection of security standards, algorithms and parameters) for application entities involved in exchanging information to implement security policies. The implementation of the DICOM Part 15 Secure Transport Connection Profiles for DICOM image security transmission utilizes the framework and negotiation mechanism specified by the TLS (Transport Layer Security) version 1.0 protocol, which is derived from SSL (Secure Socket Layer) version 3.0. The security communication of IPv6/IPv4 enable DICOM image transmission utilizes IPsec protocol, which is now mostly used in VPN (virtual private network) applications, and will be widely used in NGI. So, in this section, we talked about the software implementations of IPv6/IPv4 DICOM communication with IPsec supported, and the SSL/TLS-based DICOM secured communication. In next section, we tested and compared the efficiencies of both secured DICOM communications with different modality DICOM images, different security algorithms, and different platforms.

From figure1, we see that there are two methods to ensure the security of transmitting medical images through public network. Generally speaking, we can implement it in different network layer, for example in IP layer, in TCP layer or in application layer. IPSec protocol is a method to ensure information security in IP layer, while SSL protocol can achieve this between TCP and application layer. In application layer, you can encrypt your medical images and then transmit them. Following we will introduce DICOM security implementation based on IPSec and SSL using CTN library.

3.1. DIOCM communication on IPSec

From figure 1, we can see that IPSec is member of IPV6 protocol family. It provides security to the IP and the upper-layer protocols.. IPSec is composed of two protocols: AH protocol and ESP protocol. AH is used to ensure the authentication and integrity, while ESP is used to ensure confidentiality. AH protocol uses hash message authentication codes (HMAC) to protect integrity. There a lot of algorithms can be used in AH, such as SHA, MD5, etc. ESP protocol uses the standard symmetric encryption algorithms to protect confidentiality, such as 3DES, AES and Blowfish, etc.

Figure 2 shows the data flow of the IPv6/IPv4 DICOM Storage communication SCU/SCP with IPsec supported. The DICOM SCU and SCP have its own certificate which is created by the same CA (Certificate Authority). There are three steps in this communication. The first is IKE (internet key exchange) protocol association. In this step, the ISAKMP daemons running in both SCU/SCP sites negotiate the IKE parameters and exchange certificate, which is used for IPSec association. In second step, SCU/SCP entities establish IPSec association. In this step, the both negotiate IPSec parameters and create session key, which is used for security communication of DICOM data. The third is transferring the DICOM data on the secure channel.

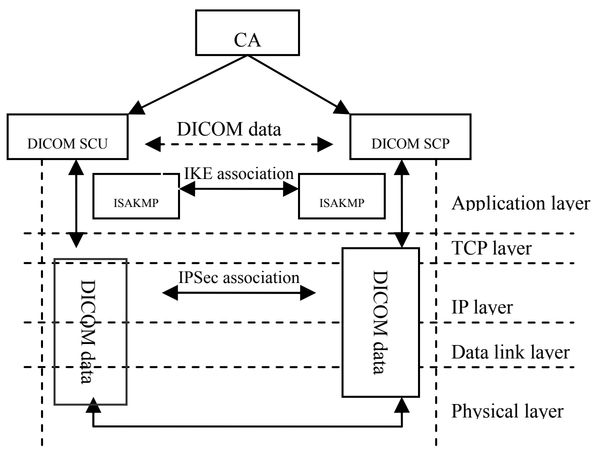


Figure 2. The data flow of the IPv6/IPv4 DICOM Storage communication SCU/SCP with IPsec supported

IPSec working in two modes: tunnel mode and transport mode. In tunnel mode, the entire IP datagram is encapsulated by a new datagram; while in transport mode only the payload (upper layer data) is handled by inserting AH header, ESP header, or both. Because of working in tunnel mode need routing, we did not test it.

We can set security association manually or uses internet key exchange protocol (IKE). IKE is more convenience and more security.

Because of Working in IP layer, IPSec has no effect on our DICOM communication program, which works in application layer. To test the performance of DICOM communication on IPSec, what we need to do is setup SA for peers to establish the secure channel. During the setup, we need create certificates for peers and set SA associated parameters.

3.2 DICOM communication based on SSL

SSL was originally developed by Netscape Communications to allow secured access of a browser to a Web server, SSL has become the accepted standard for Web security. It provides secure communication channel between client and server by allowing mutual authentication, which uses digital signatures for integrity, and encryption for privacy. The protocol is designed to support multiple choices of specific algorithms used for cryptography, digests, and signatures. SSL 3.0 is the basis for the TLS protocol, which is still been developed by the Internet Engineering Task Force (IETF)[8]. The SSL protocol uses both of public-key and symmetric key encryption. Symmetric key encryption is much faster than public-key encryption, but public-key encryption provides better authentication techniques.

SSL protocol consists of two protocols: the handshake protocol and the SSL record protocol. The handshake protocol defines how the peers exchange associated information, such as, SSL version, ciphers and authenticates certificate. The SSL record protocol defines the format of SSL record or message, in which all of the SSL associated message or application data should be transferred. SSL connect is executed in two phrases. The first is handshake. The second is transferring data.

SSL/TLS works between TCP layer and application layer and the most implementation of SSL(v.3.0)/TLS communication library simulate the style of the Berkeley Socket APIs. So, in the implementation of the SSL/TLS-based DICOM secured image communication, we replace TCP APIs with that by using OpenSSL toolkit [5] in the DICOM CTN library software, and recompile the CTN library, link the application with SSL/TLS enabled DICOM library. The data flow of the SCU/SCP applications of SSL/TLS enabled DICOM Storage services is shown in Figure 3.

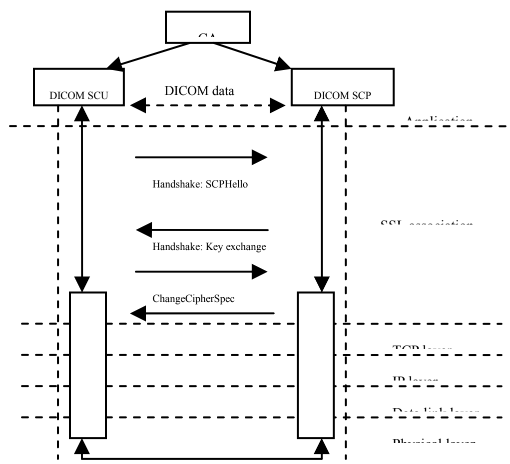


Figure 3. The data flow of DICOM Storage SCU/SCP communication with SSL/TLS supported.

4. EVALUATION OF DICOM IMAGE SECURITY COMMUNICATION

Since different transmission protocols, different PUD (Protocol Data Unit) sizes, different security algorithms, and different computer OSs may have different effects on different medical images in network transmission. So, we design some transmission experiments to test and evaluate images transferring efficiencies of IPv6/IPv4 enabled DICOM communication with different security configuration, algorithms, modalities, and operating systems, and compare them with SSL/TLS supported DICOM image communication.

We took the DICOM C-STORE as example to test the performance of images transferring efficiencies. One computer works as DICOM C-STORE SCU sending images to the other, and the other works as DICOM C-STORE SCP receiving images. To avoid the interfering of the network equipments such as network switches, routers, we connected two computers directly with the reticle connection. The computer configurations are:

(a) C-Store SCU: Dell Dimension PC with Intel P4 CPU:2.8G, RAM:512M, and NCI:100Mb/s;
(b) C-Store SCP: Dell Dimension PC with Intel P4 CPU:2.8G, RAM:512M, and NCI:100Mb/s;
(c) OSs: Red Hat Enterprise Linux V3.0., and Microsoft Windows XP.

The DICOM image data rested are as followings:

(a) CT : one series with 100 DICOM images (512x512x2byte), and total size 53,016,540bytes;
(b) MRI: one series with 200 DICOM images (256x256x2byte), and total size 35,219,944 bytes;
(c) CR: 10 DICOM images (2048x2495x2byte), and total size 89,150,896 bytes;
(d) US: one single multiframe DICOM images with total size 135,322,856 bytes.

For IPSec and SSL/TLS enabled security communication, we create X509 certificates for both sites of DICOM C-Store SCU and DICOM C-Store SCU from the same CA attached in OpenSSL tool kit.

5. EXPERIMENTAL RESULTS AND DISCUSSION

We tested the DICOM image communication with different protocols (IPv6/IPv4), different security configuration (IPsec and SSL/TLS) and algorithms, and different parameters. The figure 4 to 6 show the experimental results. In these figures, the Y axis represents the measured speed of DICOM communication (kb/s); The X axis represents the different kind of DICOM image communication with different protocols, for example, CT(IPV4) represent the test is on CT image and using IPV4 protocol. The Z axis represents the different PDU size used in TCP/IP communication, and different operation systems, as well as the different algorithms, for example, 4096(LINUX 3DES) means that the maximum PDU size is 4096, the OSs of DICOM C-Store SCU and DICOM C-Store SCP is LINUX, and the security algorithm is 3DES.

5.1 The performance evaluation of IPV6 and IPV4 DICOM image communications without IPsec security

Figure 4 shows that measurement results of IPv6 and IPv4 DICOM image communications under the experimental conditions mentioned in above section. From figure 4, we got following results:

(1) The DICOM image communication on LINUX is faster than that of on windows for same image data set;

(2) The DICOM communication based on IPv4 is mostly faster than that of on IPv6. This is because the size of IPv4 packet's header is smaller than that of IPv6. But sometimes it is fast on IPv6. We think this is because IPv6 has other features, for example better route performance in despite of

bigger IP packet,.

(3) The PDU size has strong decreasing effect on DICOM communication speed with Windows platform, and the size 4096 bytes is the best compared to other two of sizes (16384 and 65536 bytes). Meanwhile, it don't have obviously decreasing effects with Linux.

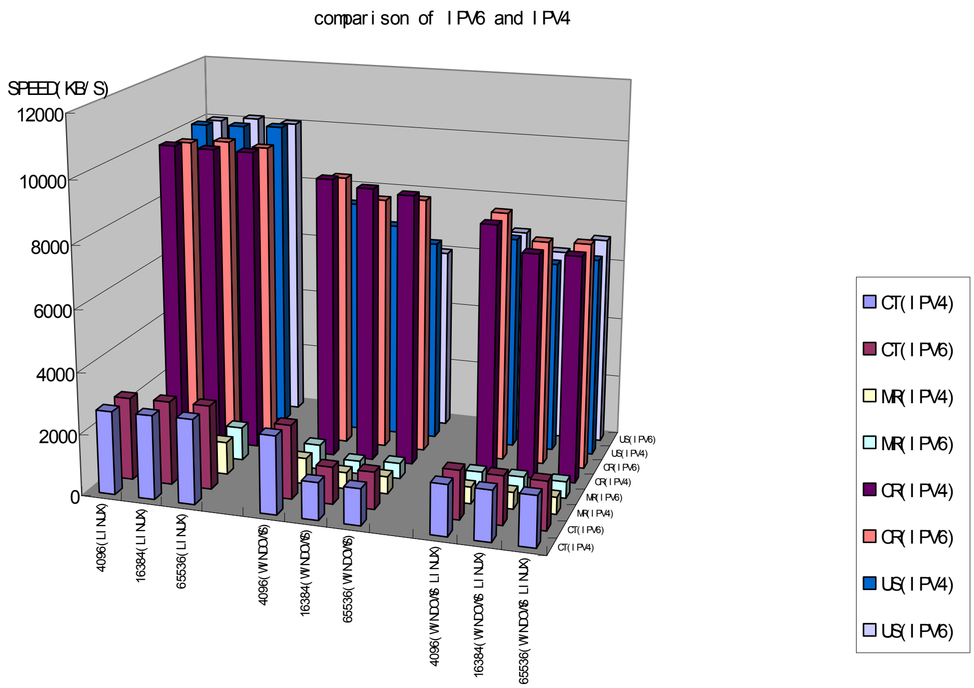


Figure 4. The performance comparison of IPV6 and IPV4 DICOM image communication without security

5.2 The performance evaluation of IPV6 and IPV4 DICOM image communications with IPsec security

Figure 5 shows the testing results of IPv6/IPv4 DICOM image communication with IPSec security. From figure 5, we see that:

(1) All the encryption algorithms have stronger decreasing effect on CR and US(DSA) images, while they are weaker on CT and MR images. The explanation is that it would take more times of SCU/SCP establishing association for CT/MR images than that for CR/US images with same quantity of image sizes, and these establishing times concealed that required for security association and encryption, so the encryption has less decreasing effect on small size images.

(2) For Linux platform, the algorithms of 3DES and Null_enc have stronger decreasing effect on the efficiency of DICOM image transmission than that of DES, AES and Blowfish448. For Window platform, the 3DES have stronger decreasing effect on the efficiency of DICOM image transmission than that of DES. The implementation of IPSec is not so good on Windows as it on Linux, since it only

supports DES and 3DES algorithms on IPv4 but does not support on IPv6.

(3) General speaking, the performance of DICOM image transmission is faster over TCP/IPV4 than that over TCP/IPv6 if IPsec was not enabled. But it is not always true when the IPsec security working. We found that the performance of DICOM image transmission is faster over TCP/IPv6 than that over TCP/IPv4 in sometimes if IPsec was enabled. The reason for this is that the integration of IPv6 and IPSec may be better than that of IPv4 and IPSec.

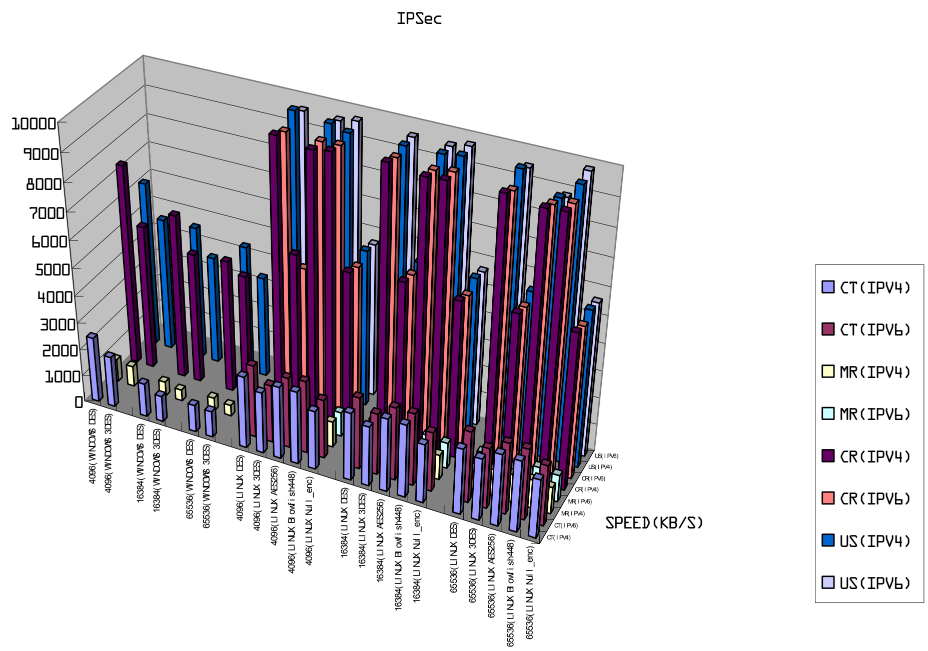


Figure 5. The IPv6/IPv4 DICOM image communication with IPSec security.

5.3 The performance evaluation of DICOM image communications with SSL/TLS security over TCP/IPv4 in different operating systems

Figure 6 shows the performance of DICOM image transmission with different SSL/TLS security algorithms over TCP/IPv4 in different operating systems.

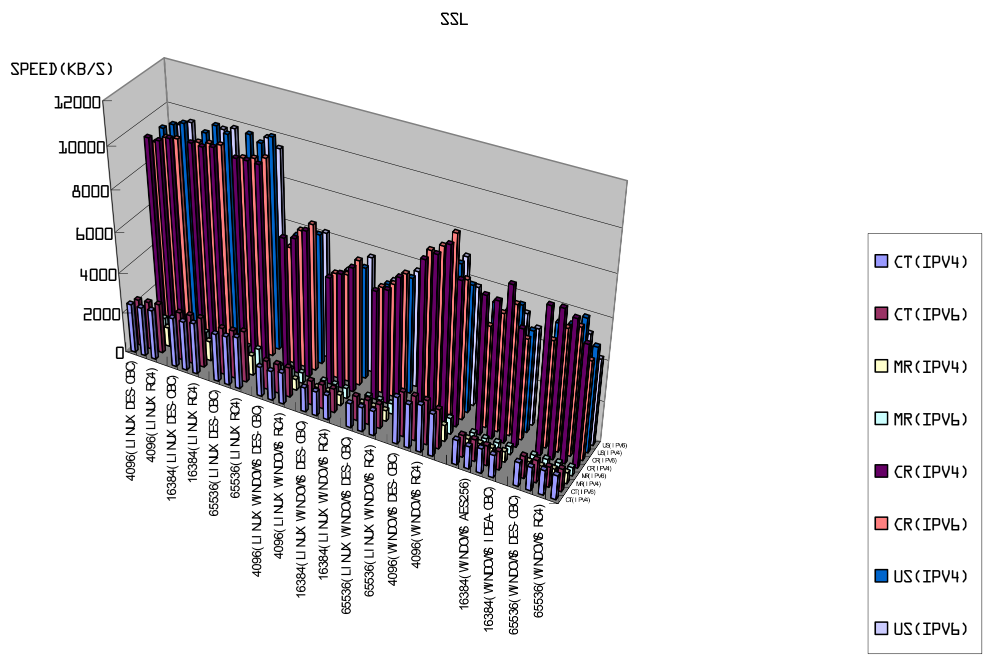


Figure 6. The performance of DICOM image transmission with different SSL/TLS security algorithms in different operating systems

From Figure 6, we see that:

(1) Transmission speeds of DICOM images with different SSL/TLS security algorithms over TCP/IPv4 in Linux and Windows are faster than that with same security setting but cross-platforms such as SCU (Linux) to SCP (Windows);

(2) All the encryption algorithms have stronger decreasing effect on CR and US(DSA) images, while they are weaker on CT and MR images. The explanation is same as we achieved in Section 5.2;

(3) The security algorithms of DES-CBC and IDEA-CBC have stronger decreasing effect on the performance of DICOM image transmission than that of AES and RC4.

5.4 The Performance comparison of DICOM image transmission with IPsec and SSL/TLS

Figure 7 shows the performance comparison of IPSec and SSL. From this figure, we see that, for CT and MR images, the speeds of DICOM image communication using IPSec are faster than that of using SSL. But when the max PDU size is 16384 and 65536, as well as the algorithm is AES and the DICOM images are CR or US, the speeds of DICOM image communication using SSL are faster than that of using IPSec. But in general speaking, they are almost same.

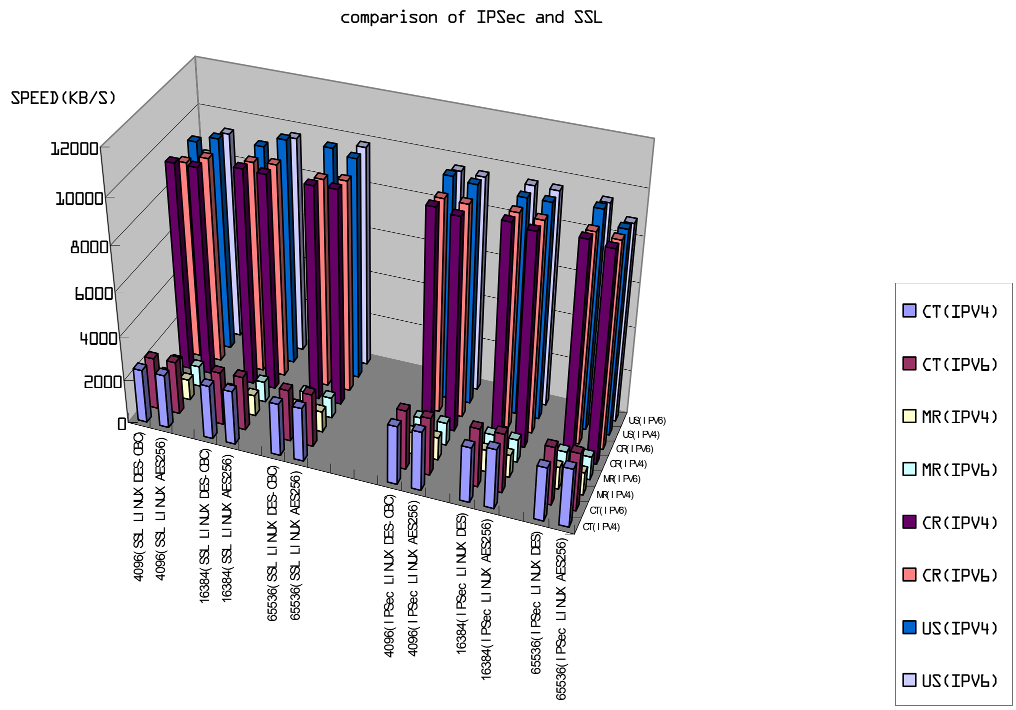


Figure 5 the performance comparison of IPSec and SSL

6. CONCLUSIONS

It is straight forward to use open source codes of DICOM software to develop IPv6 and IPv4 enabled DICOM communication software and libraries. The evaluation measurement results showed that: (1) there were some overheads in using IPv6 DICOM image communication compared to IPv4 since more advanced features were achieved in IPv6; (2) All the encryption algorithms used in IPsec and SSL/TLS security have stronger decreasing effect on CR and US(DSA) images than that on CT and MR images over IPsec/SSL security-integrated TCP/IPv6/IPv4 networking; (3) The performance of DICOM image transmission is faster over TCP/IPV4 than that over TCP/IPv6 if IPsec was not enabled, but it is not always true when the IPsec security was enabled, and the integration of IPv6 with IPSec may be better than that of IPv4 with IPSec; (4) Transmission speeds of DICOM images with different SSL/TLS security algorithms over TCP/IPv4 in Linux and Windows are faster than that with same security setting but cross-platforms such as SCU (Linux) to SCP (Windows).

There are some trade-off to choose security solution between IPsec and SSL/TLS in the security implementation of IPv6/IPv4 protocols. If the WAN networks only use IPv6, the choice is IPsec, since IPv6 and IPsec are implemented in IP layer, it has not to change application software. But, the operating systems have to do some configurations to enable the IPv6 protocols with IPsec-integrated. If the networks are IPv4 or the combination of IPv6 and IPv4, it is better to use SSL/TLS security by modifying the TCP/IP APIs of application software, since the integration of IPv4 with IPsec is not so good as IPv6 with IPsec, or to find a VPN product but there are still some limitation in network deployment.

7. ACKNOWLEDGMENT

This research was supported in part by the National Nature Science Foundation of China (Grant No. 39925009), China High Tech. Plan (Contract No. 2003AA0010), and Shanghai Sci.&Tech. Plan (Contract No. 03DZ19709), and Chinese Academy of Sciences. The authors would like to thank the colleagues of Shanghai Infection Hospital for data acquisition, transmission, and system testing.

8. REFERENCES

1. National Electrical Manufacturers Association Digital Imaging and Communications in Medicine (DICOM)□NEMA Standards Publication PS3.1, 1998.
2. Christian Huitema IPV6 The New Internet Protocol Prentice Hall Nov 1999 ISBN 7-302-03547-4
3. Carlton R. Davis, IPSec: Securing VPNs, McGraw-Hill, April 2001, ISBN 0-07-212757
4. IPSec HOWTO Ralf Spenneberg ralf (at) spenneberg.net
5. SSL and TLS: Designing and Building Secure Systems Oct 2000 by Addison-Wesley ISBN: 0201615983
6. Implementing and Configuring IPv6 in Windows 2003 and XP SP1 Keith H. Irby (Assignment 1.4b, Option 1) 12/22/03
7. Setting up IPSec between Windows 2000/XP and FreeBSD 4 Tim J. Robbins 20 January 2002.
8. White Paper Introduction to Secure Sockets Layer Cisco Systems, Inc

Evaluation of Security Algorithms used for Security Processing on DICOM Images

Xiaomeng Chen*, Chen*, Jie Shuai*, Jianguo Zhang*+ , and H. K. Huang*+
*Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai
+Image Processing and Informatics Laboratory, Departments of Radiology and Biomedical Engineering, University of Southern California, Los Angeles

ABSTRACT

In this paper, we developed security approach to provide security measures and features in PACS image acquisition and Tele-radiology image transmission. The security processing on medical images was based on public key infrastructure (PKI) and including digital signature and data encryption to achieve the security features of confidentiality, privacy, authenticity, integrity, and non-repudiation. There are many algorithms which can be used in PKI for data encryption and digital signature. In this research, we select several algorithms to perform security processing on different DICOM images in PACS environment, evaluate the security processing performance of these algorithms, and find the relationship between performance with image types, sizes and the implementation methods.

Key Words: Security architecture, Medical information systems, PACS

1. INTRODUCTION

The Health Insurance Portability and Accountability Act (HIPAA) of 1996 authorized the Secretary of Health and Humans Service to provide Congress with regulations mandating standards for the security and privacy of patient medical records, and require that the Secretary of Health and Human Services (HHS) establish security standards for healthcare information systems and standards for electronic signatures .The standard (45 CFR 142) is a single standard that integrates all possible components of medical security protocols (administrative procedures, physical safeguard ,technical security services , technical mechanisms, and electronic signature). These technical security mechanisms include access control, audit control, authorization control, data authentication, entity authentication, alarm and event reporting, data encryption and electronic signatures. We have developed security components, certificate authoring (CA) system and patient record digital signature (DSPR)management system, as well as electronic envelope technology, and designed to integrate these security components into the current hospital healthcare information infrastructure to provide security measures and functions such as confidential or privacy, authenticity, integrity, reliability, non-repudiation, and authentication for in-house healthcare information systems daily operating, and EPR exchanging among the hospitals or healthcare providers. In this paper, we first outline current work flow and data flow in an integrated healthcare information system such as HIS/RIS integrated PACS, and analyze the security measures and functions required in some components and procedures. Then, we introduce security services, based on public key infrastructure (PKI), in in-House operation of HIS/RIS integrated PACS and in EPR exchanging between medical institutions. Finally, we evaluate the performance of security algorithms used in these security services.

2. WORK AND DATA FLOW IN IN-HOUSE OPERATION INTEGRATED PACS

The major security components and dataflow integrated in PACS related to the work flow are shown in Figure 1. In order to provide the data security services such as confidential or privacy, authenticity, integrity, non-repudiation and authentication integrated PACS workflow operation, e.g., generating or creating images, accessing and retrieving images, the related security measures and functions should be embedded into PACS components or designed in the system architectures,

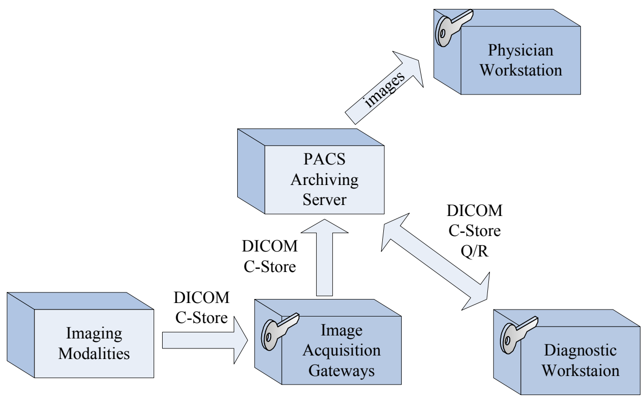


Figure 1. The data flow of PACS components integrated with security services of data encryption and digital signature

In the following, we will describe the security architectur integrated PACS. This architecture provides security functions and features by using the security information and add some security processing logics in the data flow of three components of the Image Acquisition Gateway (IAG), Diagnostic Workstations (DWS), and Physician Workstations (PWS), as indicated in Figure 1 with **key icons.** After the IAG receives images from the modalities, it creates digital signature of the image files and encrypts them. The purpose of creating Digital Signature by IAG is to validate the source of images and verify the integration of the image data in the inter-operation between the workstations and the PACS Server, so the signature of IAG mainly aims at pixel data of medical images. The signature component installed in the DWS and PWS is served for the doctors that can make use of it to create digital signature of diagnostic reports, key images, and verify the reviewed reports or images signed by others.

3. SECURITY SERVICES IN IN-HOUSE OPERATION OF HIS/RIS INTEGRATED PACS

Since the security usually was concerned from the medical record generation to final applications, the security measures and mechanisms should exist in the chain from data acquisition to end user applications. In a HIS/RIS integrated PACS, the major components involved in security protection and verification are IAG of imaging modalities, PACS Archiving server, diagnostic/reporting workstations and physician viewing workstations. Here, we have designed and developed DSPR and CA components to implement security in a HIS/RIS integrated PACS about the privacy, authenticity, integrity, reliability, non-repudiation, and authentication [1]. First, We recognize that the IAGs are the signers of Digital signature (DS) of medical images, and the radiologists and physicians are the both signers and users of DS of radiological or medical reports. We regard that the radiological workstations (diagnostic/reporting), physician workstations, and applications servers are also the users of security services. Secondly, we separate the security information of a record from the record file, and store the extracted or computed security information into the DSPR component, e.g., ID of signer, UID of record, DS and HV of record, Tag of item encrypted in record, and ID of encryption algorithm applied to the record, etc. All these security processing and computation are done in the IAG for images, application workstations or servers for text reports, with the private-keys of the owners contained in the digital certificates. The Data flow of computation and transmission of computed security information of images in IAG of PACS, as shown in Figure 2.

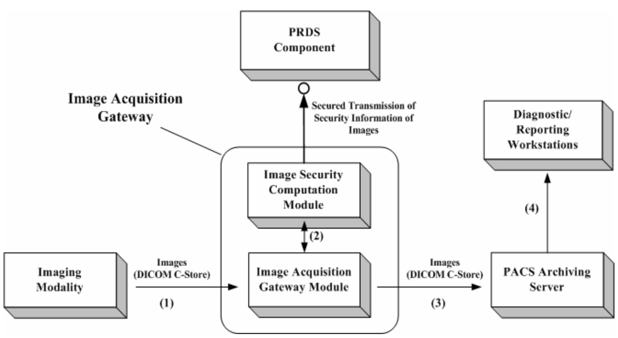


Figure 2 Data flow of computation and transmission of computed security information of images in IAG of PACS

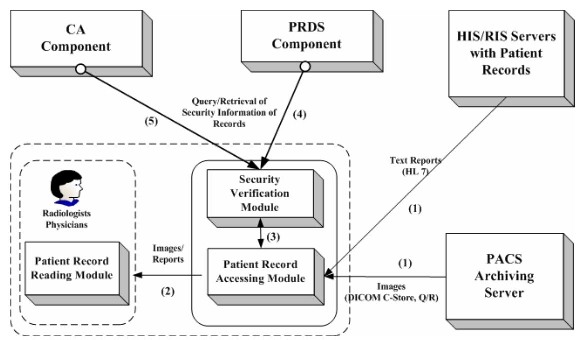


Figure 3 Data flow of security verification in reviewing patient records of images

Since all necessary security information of medical records are acquired from acquisition gateway and diagnostic/reporting workstations, stored and managed in DSPR and CA components, it is easy to achieve required security services with these information in daily clinical operation. Figure 3 shows how to use these security information and build the input/output interfaces of the security component to the PACS components to provide the security features and mechanisms in in-house HIS/RIS integrated PACS operation. The security processing logics are: (1) If the healthcare information systems are two tiers of client/server architecture, the **Patient Record Reading Module**, **Patient Record Accessing Module** and **Security Verification Module** will stay in application workstations, e.g., radiological workstations, or physician workstations; (2) If the healthcare information systems are three tiers of architecture, **the Patient Record Reading Module** will stay in the client side, and **Patient Record Accessing Module** and **Security Verification Module** will stay in the middleware application server. Following, we will describe how to use **CA** and **DSPR** components and **Security Verification Module** to perform the security services

Privacy: The sensitive information of patient records can be encrypted and stored in the record file, and the related **tags** of the information and **ID of encryption algorithm** are stored in DSPR. This process can be done in IAG for images, and diagnostic/reporting WS for text reports, as indicated in Figure 2. So, it will not disclose the sensitive information when user read the record in non-clinical services. But, it is easy to decrypt the sensitive information with the retrieved Tags and ID of encryption/decryption algorithm from DSPR by using **Security Verification Module** installed in the workstation or application server if it is required by authorized users.

Authenticity and Integrity: The authenticity and integrity of a record can be easily done by comparing the hash values (HV) stored in DSPR and retrieved record from PACS/HIS/RIS with same UID of the record. The steps to perform this comparison are: (1) when an user read a record retrieved from a server, and want to verify the authenticity and integrity of this record, he or she triggers the **Security Verification Module** to compute the HV of the record and compare it with the that got from DSPR. If they are same, the record is correct, otherwise, the authenticity and integrity of record was broken.

Non-repudiation and Authentication: When user want to verify the non-repudiation and authentication of a record, (1) he or she calculates the HV of this record and uses the UID of this record to get the ID of the signer and DS of this record from DSPR through the **Security Verification Module**; (2) use the ID of the signer to get the public key of the signer stored in CA component as indicated in Figure 3; (3) if the user can decrypts the DS by use the public key of the signer and the extracted HV from DS is same with that calculated from the record, the security verification is passed, otherwise, it is failed.

The **Image Security Component Module** and **Security Verification Module** contain the security services of digital signature and verifying signature, which use security algorithms including SHA1, MD5, RipeMD160 associated with RSA, DSA. As the P12 certificate including Public key and private key coming from certificate authority (CA) system, it accessed by password can be adopted into our signature module. One signer loads his certificate and reads the whole file or only pixel data of it, then calculates the hash value of the to-be-signed object, finally calculates the signature via RSA algorithm and stores it into the file. While being verified, the part of signed from signed DICOM file will be hashed, we can use the hash value to verify the digital signature. The digital signature and verification are illustrated in Figure 4 and 5.

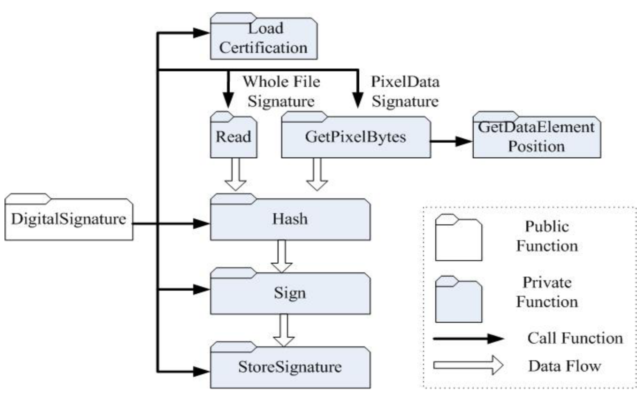


Figure.4 Dicom Images Signature Function

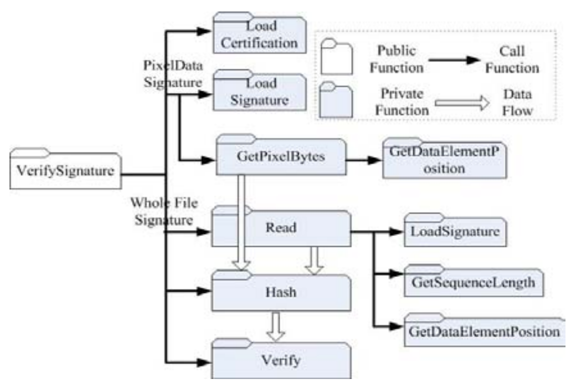


Figure.5 Dicom Images Verification Function

4. SECURITY SERVICES FOR EPR EXCHANGING BETWEEN MEDICAL INSTITUTIONS

In telemedicine and enterprise healthcare operation, the electronic patient record (EPR) will be transferred from one hospital or clinic to another through public wide area network. The security issue is very important in these kinds of applications. Usually, in order to safely transfer the EPR, electronic envelopes [2] are needed to convey the EPR and related security information from one side to another. There are two situations: (1) Both sending and receiving sides have CA components and their digital certificates were authorized from same public root CA system, also both sending and receiving computers get digital certificates from their CA components and can use SSL technology to transfer the EPR; (2) Both sending and receiving sides may have CA components, but can not communicate each other by using of SSL in EPR exchanging. The electronic envelopes will be constructed differently with the security information from DSPR and CA for these two situations and the communication efficiency is also different.

In EPR data (the major challenges are mostly on medical images) secure transmission, we adopt symmetric crypto algorithms that include DES, DESede (TripleDES), Blowfish, Rijndael□AES□,IDEA,RC4, (others are block ciphers except that RC4 is stream cipher),etc, and encrypt the EPR data or whole DICOM images as indicated in Figure 6 and only pixel data (Figure 7) of it. The key length of DES is 56 bits, and other algorithms are 128 bits. The whole DICOM file encrypted can't be read, and must be parsed and displayed after being decrypted. When the only pixel data of DICOM file is encrypted, value of DICOM tags can be parsed, however pixel data is displayed into grey.

Figure.6 Whole Dicom Images Encryption/Decryption

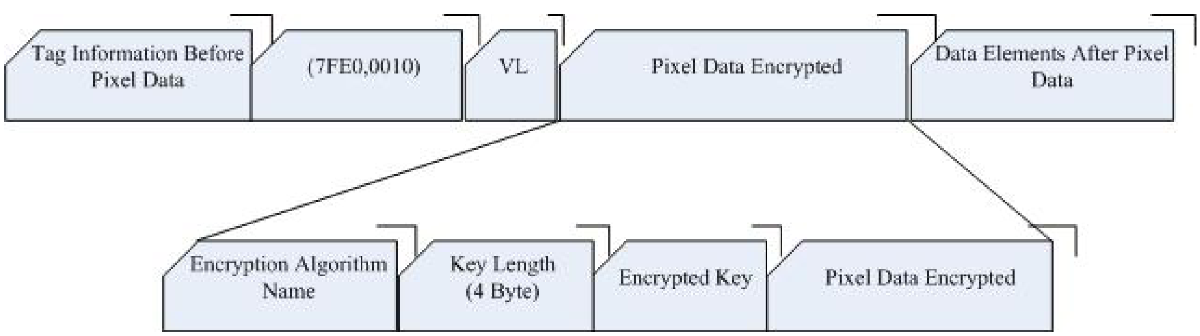


Figure.7 Only Pixel Data Encryption/Decryption

5. RESULTS OF ALGORITHM EVALUATION

We used the Java Cryptography Extension (JCE) provided by Bouncy Castle and the CTN DICOM Toolkit to develop data encryption software for DICOM image data encryption and decryption processing, and used the RSA package with different hash computing algorithms to do the digital signature of DICOM image data and authentication. These two security processing modules were integrated into the PACS DICOM gateway and display workstations. We select one series of DICOM CT images with 50 slices (512x512x2 bytes/image), one series of DICOM MR images with 100 slices (256x256x2 bytes/image), 10 CR images (8 MB/image), and two DICOM multi-frame Ultrasound images (512x512x3bytes, 6.2MB), to test the impact of security processing on various DICOM images in PACS data flow. We used the computer (Dell Dimension 8100 with Pentium III: single CPU 866MH, RAM 512MB) to test the security computation on various images.

5.1 Time required for performing encryption and decryption on DICOM image files

The times required for performing security processing on various DICOM images are presented in Table I.

Table I. The times required for performing security processing on different DICOM images with different encrypt/decrypt algorithms.

File Type	Crypto Algorithm	Whole File Encrypt Time □Second/10^7 Byte□	Whole File Decrypt Time □Second/10^7 Byte□	Only Pixel Encrypt Time □Second/10^7 Byte□	Only Pixel Decrypt Time □Second/10^7 Byte□
CR Whole File 71.24MB Only Pixel□ 64.52MB	DES	2.94778	3.44329	3.18661	3.14321
	DESede	6.34475	6.77148	7.07843	6.87384
	Blowfish	2.24312	2.86216	2.56975	2.5031
	Rijndael	11.1468	13.7086	12.2768	14.5614
	IDEA	3.45171	4.08899	3.79107	3.75387
	RC4	1.06261	1.56373	1.15623	1.16398
CT Whole File□ 26.41MB Only Pixel□ 26.21MB	DES	2.93828	3.59712	2.94544	2.91492
	DESede	6.35365	6.5354	6.56238	6.45937
	Blowfish	2.61265	2.82469	2.47234	2.54101
	Rijndael	11.4578	13.6161	11.5681	13.1973
	IDEA	3.71829	4.38092	3.68943	3.42236
	RC4	1.30632	1.80235	1.28195	1.27432
MR Whole File□ 26.65MB Only	DES	3.26454	3.84991	3.32439	2.80291
	DESede	6.46154	7.07317	6.6411	6.37653
	Blowfish	2.57411	3.22702	2.82209	2.45399
	Rijndael	11.3508	13.8424	11.8098	13.2515
	IDEA	3.53096	4.33771	3.94172	3.42791

Pixel□ 26.08MB	RC4	1.54597	2.15009	1.70629	1.43021
US Whole File□ 6.19MB Only Pixel□ 6.14MB	DES	3.11793	3.53796	2.71987	2.89902
	DESede	6.33279	7.59289	5.86319	5.7329
	Blowfish	2.43942	2.47173	1.40549	2.50814
	Rijndael	11.2763	13.6672	10.6678	13.5179
	IDEA	3.50565	4.23263	2.96417	3.5342
	RC4	1.22779	2.90792	1.04235	1.30293

5.2 Time required for performing digital signature and verification on security standardized DICOM image files

We calculated the times required for both the whole DICOM image file and that of only pixel data part, and compared them with different kinds of medical images. Table II gives the measured times required in these computation of digital signature and verification.

Table II. The times required for digital signature and verification processing on different DICOM images with different security algorithms.

File Type	Crypto Algorithm	Signing Time of Whole file □Second/10^7 Byte□	Verifying Time of Whole file □Second/10^7 Byte□	Signing Time of Pixel Data □Second/10^7 Byte□	Verifying Time of Pixel Data □Second/10^7 Byte□
CR	SHA1/RSA	3.95564	1.21421	5.3797	2.38376
	MD5/RSA	5.72431	1.22824	6.2554	2.65654
	RipeMD160/RSA	5.77625	1.29281	6.1299	2.47675
CT	SHA1/RSA	5.86899	2.54828	6.5433	2.81572
	MD5/RSA	6.70201	2.72624	6.4556	3.18962
	RipeMD160/RSA	5.3919	2.74517	6.6768	3.39565
MR	SHA1/RSA	8.74672	5.11069	11.277	5.57515
	MD5/RSA	9.17824	5.18574	10.817	5.46012
	RipeMD160/RSA	9.14447	5.51595	10.234	5.66334
US	SHA1/RSA	8.28756	3.19871	8.3225	3.87622
	MD5/RSA	7.68982	3.03716	8.4365	3.92508
	RipeMD160/RSA	7.65751	3.08562	8.5505	3.84365

We also calculated consuming times of encryption/decryption processing on different amount of different modality images with different algorithms and drawn the relationship curves of them in Figure 8, 9. From these curves, we see that different encryption/decryption algorithms on different modality medical images are different, and Blowfish and IDEA algorithms performed best on most DICOM images in encryption and decryption. Also, the processing times are linearly related to the total volume of image pixel data

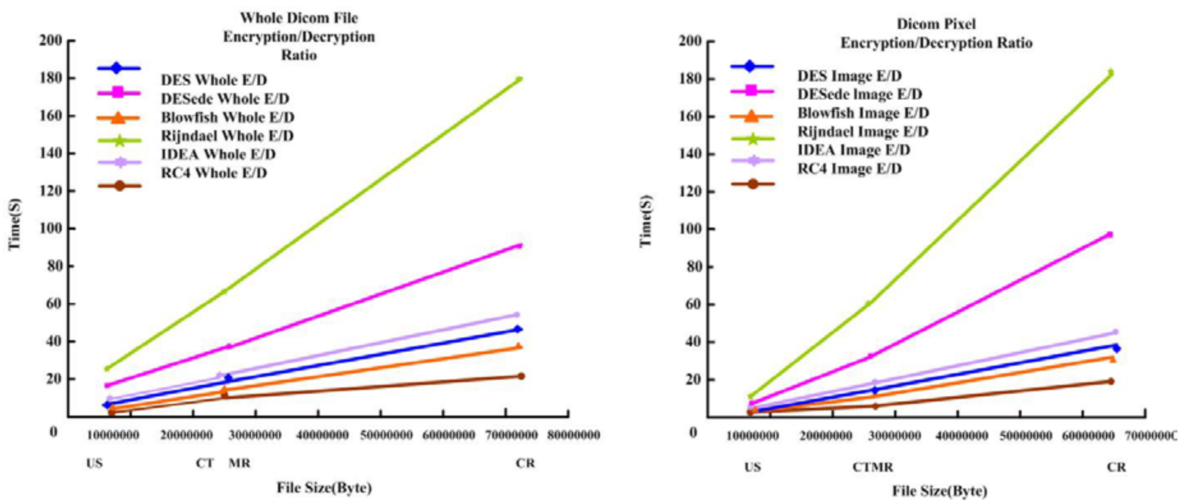


Fig.10 DICOM Images Signature/Verification Ratio

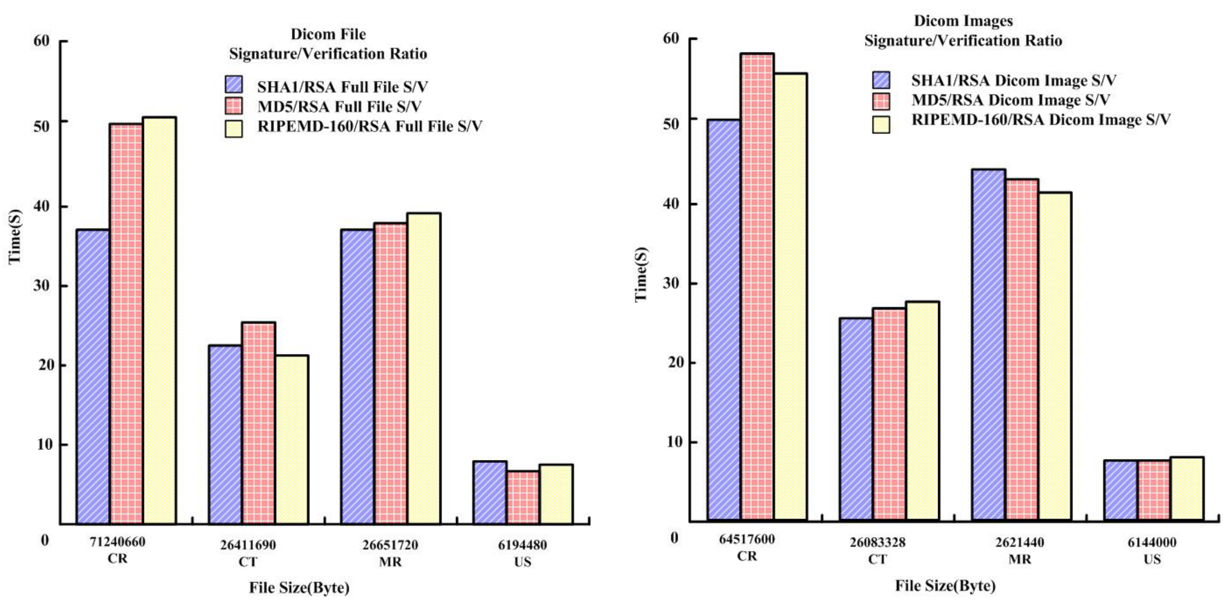


Fig.11 DICOM Images Signature/Verification Ratio

For security processing of the digital signature and verification on different kinds of modality images with different security algorithms, we drawn two histograms as shown in Figure 10 and 11. From Fig.10 and Fig.11, we see that it is time consuming to do digital signature or verification processing on large size of image files, and also different algorithms have different efficiency impacts on processing performance. All three algorithms of SHA1/RSA, MD5/RSA, RipeMD160 have similar performance on selected images. Since SHA1 is widely used, it may be considered the better choice in algorithm selection in security implementation.

6. CONCLUSIONS

Integrating security processing modules based on PKI in PACS acquisition gateways and display workstations in PACS data flow can achieve the security features of confidential, privacy, authenticity, integrity, and non-repudiation in digital radiology work flow. In this paper, we developed security approach to provide security measures and features in PACS image acquisition and Tele-radiology image transmission. The security processing on medical images was based on public key infrastructure (PKI) and including digital signature and data encryption to achieve the security features of confidentiality, privacy, authenticity, integrity, and non-repudiation. There are many algorithms which can be used in PKI for data encryption and digital signature. The encryption and decryption algorithms of Blowfish and IDEA perform better on most DICOM images, and the required processing times are linearly related to the volumes of image pixel data. The algorithms of SHA1/RSA, MD5/RSA, RipeMD160 used in digital signature and authentication on DICOM images have similar performance, but, SHA1 is widely used, it may be considered the better choice in security implementation.

7. ACKNOWLEDGMENT

This research was supported in part by the National Nature Science Foundation of China (Grant No. 39925009), China High Tech. Plan (Contract No. 2003AA0010), and Shanghai Sci.&Tech. Plan (Contract No. 03DZ19709), and Chinese Academy of Sciences.

8. REFERENCES

[1]. Jo Sammuel Dwyer III, “HIPPA and Implementation Security of Medicine Data”, Chapter 2, pp9-14, Security Issue in the Digital Medicine Enterprise, Edited By B. Reiner, E. Siegel, SCAR Press, 2000.

[2]. X. Zhou and H.K. Huang, “Authenticity and Integrity of Digital mammography Images”, IEEE Trans. Medical Imaging, Vol. 20, pp748-791, 2001.

[3]. “Develop Security Architecture for Both in-House Healthcare Information Systems and Electronic Patient Record”, SPIE, 2003

SELECTED PEER-REVIEWED REPRINTS AND PREPRINTS

RadioGraphics

Informatics in Radiology (*info*RAD)

A Complete Continuous-Availability PACS Archive Server[1]

Brent J. Liu, PhD • H. K. Huang, DSc • Fei Cao, PhD • Michael Z. Zhou, BS • Jianguo Zhang, PhD • Greg Mogel, MD

The operational reliability of the picture archiving and communication system (PACS) server in a filmless hospital environment is always a major concern because server failure could cripple the entire PACS operation. A simple, low-cost, continuous-availability (CA) PACS archive server was designed and developed. The server makes use of a triple modular redundancy (TMR) system with a simple majority voting logic that automatically identifies a faulty module and removes it from service. The remaining two modules continue normal operation with no adverse effects on data flow or system performance. In addition, the server is integrated with two external mass storage devices for short- and long-term storage. Evaluation and testing of the server were conducted with laboratory experiments in which hardware failures were simulated to observe recovery time and the resumption of normal data flow. The server provides maximum uptime (99.999%) for end users while ensuring the transactional integrity of all clinical PACS data. Hardware failure has only minimal impact on performance, with no interruption of clinical data flow or loss of data. As hospital PACS become more widespread, the need for CA PACS solutions will increase. A TMR CA PACS archive server can reliably help achieve CA in this setting.

Abbreviations: ASP = application service provider, CA = continuous availability, DLT = digital linear tape, DMR = double modular redundant, FT = fault tolerance, PACS = picture archiving and communication system, RAID = redundant arrays of independent disks, SJHC = Saint John's Health Center, SPOF = single point of failure, TMR = triple modular redundant

Index terms: Picture archiving and communication system (PACS) • Radiology and radiologists, design of radiological facilities

RadioGraphics 2004; 24:1203–1209 • **Published online** 10.1148/rg.244035051 • **Content Code:** HP

[1]From the Department of Radiology, University of Southern California, Los Angeles (B.J.L., H.K.H., F.C., M.Z.Z., G.M.); and the Shanghai Institute of Technical Physics, Shanghai, China (J.Z.). Presented as an *info*RAD exhibit at the 2002 RSNA scientific assembly. Received March 5, 2003; revision requested May 15 and final revision received September 18; accepted October 16. Supported in part by U.S. Army Medical Research and Materiel Command Contract No. DAMD17–99-P-3732, NIH National Library of Medicine Contract No. N01-LM-9–3538, and Reliant Health Care, Inc. **Address correspondence to** B.J.L., 4676 Admiralty Way, Ste 1001, Marina del Rey, CA 90292 (e-mail: *brentliu@usc.edu*).

Introduction

Any large medical imaging database with clinical, research, or educational applications requires a robust archive server to receive images and deliver them to users in a reliable and timely manner. With respect to the operation of large, computer-networked server systems, *high availability* and *continuous availability* (CA) (also called fault tolerance [FT]) are the two terms commonly used to describe the degree of system reliability (1–3). We will use the abbreviations *FT* and *CA* interchangeably in this article. There is a continuum of high-availability solutions, with system availability rates ranging from 99% (88 hours of downtime per year), achieved with simple "hot spare" technology, to 99.99% (1 hour of downtime per year), achieved with clustered server–based solutions that make use of hardware and software to maintain uptime. Although the technology is available, these solutions are expensive and challenging to implement. Systems with fully FT hardware are known as CA solutions, with the highest availability rate (99.999%) and a minimal downtime of 5 minutes per year (4–6). Unfortunately, to purchase and support this highest level of FT would require close to $1 million (7).

A picture archiving and communication system (PACS) (1) is an integrated system consisting of medical image and data acquisition components involving different imaging modalities (eg, computed tomography, computed radiography, magnetic resonance imaging, ultrasonography) and storage and display subsystems, all integrated with various digital networks. PACS are widely used in hospitals and are considered mission critical for around-the-clock daily clinical operation. Whereas failure of individual workstations or acquisition components will affect functions and data flow in the local PACS branch, failure of the PACS controller or main PACS archive server could cripple the entire PACS operation. The operational reliability of the PACS server is always a major concern in a filmless hospital environment and is becoming even more important as digital techniques are increasingly being used for critical tasks within hospitals. The design and implementation of a reliable PACS should provide maximum uptime for end users while ensuring transactional integrity. In the event of a PACS component failure, users should notice only a minimal performance impact with no interruption of clinical data flow or loss of data. CA PACS design is intended to minimize manual intervention and management changes in daily workstation operation. These recovery procedures and the steps required to resume normal operation after downtime are tedious and labor intensive. At present, however, PACS vendors do not feature such a CA solution for their PACS archive servers.

In this article, we describe the design, implementation, evaluation, and clinical applications of a CA PACS archive server. In addition, we provide performance measurements that can be used as a benchmark and reference standard for future FT implementations in hospital PACS. We also discuss additional applications of a CA PACS server.

TAKE-HOME POINTS

- Continuous availability is becoming increasingly important for mission-critical clinical systems like PACS.
- Continuous availability in the clinical environment must balance cost with ease of operation.

Archive Server

For the purposes of our study, we considered possible component failures in the clinical PACS archive server at the Saint John's Health Center (SJHC) in Santa Monica, California (8). SJHC is a 224-bed community hospital that performs about 120,000 radiology examinations annually. Approximately 90% of these examinations are digitally acquired and stored in a PACS archive server. The archive server is also responsible for prefetching and distributing any current and previous examinations. In addition, it provides patient location information for automatic image distribution to specific review workstations located on the hospital floors. As with all traditional local area network PACS, the archive server is the command center of the SJHC PACS.

Downtime Experience

During a 1½-year period of continuous operation (24 h/d, 7 d/wk), the archive server encountered a total of approximately 4 days of downtime resulting from three separate incidents. The first two incidents were due to system hard disk failures, with 1.5 days of downtime occurring for each incident. The third incident was due to a CPU motherboard failure: Nonvolatile RAM on the motherboard failed and needed replacement. Because the hard disk was not corrupted or damaged, the server encountered only about 1 day of downtime.

Normal operations and routines that were performed automatically had to be adjusted and readjusted once the server was brought back up. These adjustments had a major impact on clinical workflow. Some of the contingency steps and adjustments to normal clinical workflow were as follows:

1. Staff personnel communicated to inform all parties of downtime and implement contingency plans.
2. Technologists manually distributed PACS examinations to all necessary workstations.
3. Historical examinations for comparative study were not available.
4. Query and retrieval were not available for clinical users.
5. Sending jobs were cleared from the workstation queue before they affected workstation performance.
6. Automatic archiving rules were turned off on workstations so that additional examinations would not be sent.
7. Provision was made for the deletion of historical examinations on workstations to make room for newly acquired examinations if downtime were to last longer than 24 hours.
8. Hardcopy reprints for off-site clinics were not available.

An additional 2 hours may be necessary to bring the system to full operational status once the archive server is back online. An ideal FT design prevents the existence of a single point of failure (SPOF) within the system—especially an SPOF that critically impacts normal operation. The archive server is considered one of the major SPOFs of a PACS. The CA PACS archive server described in this article can be used to replace this SPOF.

CA PACS Server Design

A CA PACS server is designed with both hardware and system software redundancy to automatically detect and recover any hardware failure instantaneously. In this article, we address hardware-related failure throughout the image server system. We also discuss "failover" procedures for achieving a CA PACS, whereby induced system failure causes the system to automatically begin running on a secondary system. Application software failures and human errors are not considered.

The key to designing a CA PACS server is to implement redundant systems such that, in the event of a system failure, maximum availability and reliability without the loss of transactional data can be achieved (9–11). Features to consider include *(a)* highly reliable PACS operations (ie, no loss of data or workflow interruptions), *(b)* acceptable system performance (ie, no performance degradation in daily PACS operations or only occasional glitches over relatively long periods of time, and *(c)* low cost and easy implementation (12) (ie, system portability, scalability, and affordability).

CA PACS Server System Architecture

The CA PACS server developed at the Image Processing and Informatics Laboratory at Children's Hospital Los Angeles/University of Southern California is the first low-cost implementation of CA architecture. The server makes use of a triple modular redundant (TMR) system with a high level of integration and an elegantly simple voting mechanism to achieve CA (13). In addition, the CA PACS server has been equipped with two external mass storage devices: a redundant arrays of independent disks (RAID) system and a digital linear tape (DLT) library.

TMR Architecture

Figure 1 shows the core of the TMR PACS controller, which consists of three identically configured UltraSPARC-based modules (Sun Microsystems, Santa Clara, Calif) (6). The three modules are tightly synchronized and interconnected through a high-speed backplane for intermodule communications. Each module is a complete computer running the Solaris Unix operating system (Sun Microsystems) with its own UltraSPARC CPU, memory, input/output interfaces, bridge logic, and power supply (14). The hardware in the backplane bridge compares all data that enter or exit the synchronous part of the system to diagnose any problems. The TMR voting system uses a simple majority voting logic to mask

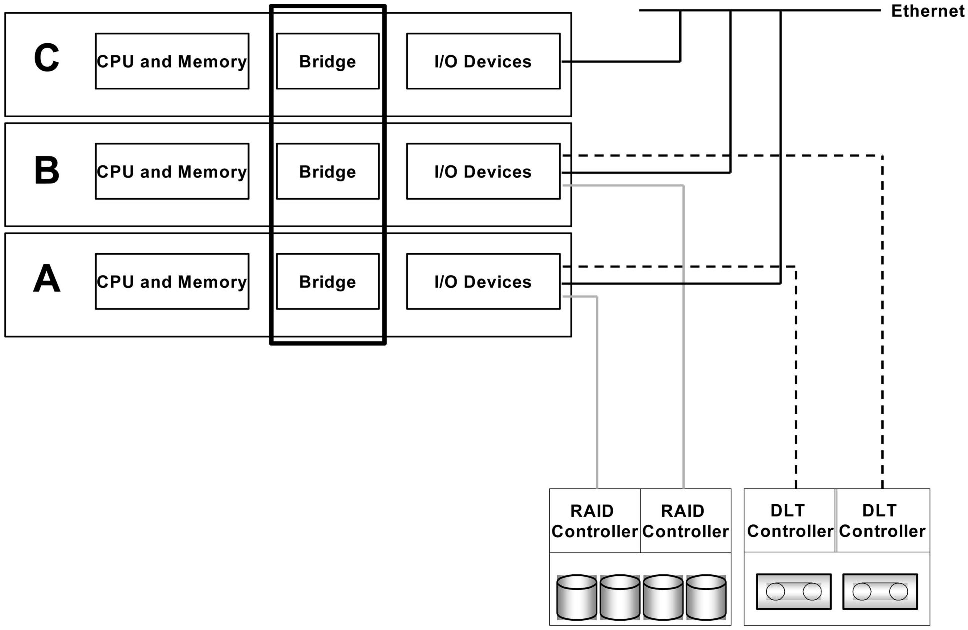


Figure 1. Diagram illustrates a PACS archive server system with TMR architecture (CPU, memory, input/output [I/O] devices, Ethernet). The server is connected to two peripheral storage devices: a dual RAID controller system for short-term storage and a dual-drive DLT library for long-term storage. Note the dual redundant paths for both types of storage.

out the failed module. This model is conceptually simple because the logic that compares the operation of the three modules need not know what faulty behavior looks like. The only requirement is that the failed module be recognized as deviant from the others. The system pauses (typically for 5–30 seconds, depending on memory size) while the diagnostic software identifies and disables the faulty module. Then, the memories of the two remaining modules are synchronized by performing a full memory copy. Once the copy is complete, the system resumes processing with the two functioning modules. During a module failure, the TMR configuration operates as a double modular redundancy (DMR) configuration until the third module is repaired. With DMR, the system can still detect a fault with use of comparison, but voting fails to determine which module is faulty. A DMR configuration generally requires the use of additional algorithms to perform self-diagnosis. The low-cost TMR system implemented in our laboratory does not support self-diagnosis in the DMR configuration. Because the comparisons occur within the hardware, the operating system incurs no overhead and performance is not affected. The system responds in a totally automatic, seamless manner, allowing uninterrupted operation of the system. The system administrator is notified of any failure within a module, which can then be replaced as needed without having to shut down the entire system (Fig 1).

Additional PACS Server Components

Ethernet and mass storage are considered important components of a CA PACS archive server and were designed and developed with CA as well (15).

Ethernet.—Each of the three modules contains its own 10/100baseT Ethernet interface. The three interfaces are connected via independent paths to the local network backbone (Fig 1) and form a single software interface with one Internet protocol and one media access control address. One interface acts as the active interface, and the others are on stand-by. Should the module containing the active interface fail or some element of its connection to the backbone fail, the interface is disabled and a stand-by unit becomes active in its place. Normal network retry mechanisms hide the failure from applications.

Short-term PACS Archive Storage.— Each of the three modules contains a fast, wide small computer system interface. To automatically survive a single disk failure, all storage must be RAID configured. Storage management software is used to implement two different FT strategies. The first strategy is the use of disk mirroring for system boot and applications storage. Identical external disk drives are attached to two or three modules, and the same data are written to each drive. The second strategy is the use of a RAID subsystem for nonboot storage. The RAID system must be cluster ready, whereby two or more modules are given their own direct link to the storage system. The RAID system is automatically detached from a failed module and reattached to a healthy module that has a functional path to the data. The RAID system must also have at least dual ports to which multiple modules can be attached. For the RAID system to be fully FT, a second hardware controller should be used. This dual-controller RAID system allows for continuous operation of the system should one of the controllers encounter a failure. In a typical configuration, two modules are connected to each of the dual-controller interfaces for redundancy (Fig 1). RAID controller failures and disk drive failures in the RAID system are handled by the system itself.

A Hitachi 9200 RAID system (Hitachi Data Systems, Santa Clara, Calif) with Veritas Volume manager software (Veritas, Mountain View, Calif) has been implemented with the FT PACS controller. The 325-Gbyte Hitachi 9200 RAID system with dual hardware controllers connected to both module A and module B provides a redundant path to the TMR PACS server. Veritas software dynamically monitors the two paths and automatically switches from one to the other if a path is disconnected due to physical damage or a failure in the hardware or connections.

Long-term PACS Archive Storage.— For long-term PACS archive storage, a StorageTeK L40 DLT library (Storage Technology, Louisville, Colo) with a storage capacity of 3.2 terabytes was implemented. The library comes with two drives, each of which is connected to one of the TMR controller modules. This configuration provides redundant paths to the FT PACS server (Fig 1). Veritas Netbackup and Migrator software is installed to automatically migrate the data from the short-term RAID system to the long-term tape library archive. The Veritas Netbackup software monitors the multiple paths and automatically causes the system to shift from one path to another should a path encounter a failure. We are currently implementing this feature in our laboratory. Another SPOF in long-term archive storage is the mechanical robotic arm of the tape library. Discussion of this SPOF is beyond the scope of this article; however, it does impact downtime in long-term storage, and future work will focus on addressing this issue. At present, the fallback plan for long-term storage is to use short-term storage during downtime. It is crucial to configure short-term storage with the proper storage space for a specified time period to accommodate potential long-term storage downtime.

The total cost of hardware for the CA PACS server is only approximately $165,000, which includes both short-term and long-term storage. The Veritas software used in the server design is an additional $70,000; however, there are other less expensive software options available that can perform similar functions. These costs do not include the PACS archive server software applications.

Testing and Performance Measurements

To evaluate the robustness and effectiveness of an FT PACS server design, a test bed environment has been developed to allow observation and gathering of results.

PACS Simulator

A system that simulates key components and the normal clinical data flow of a typical clinical PACS was developed to evaluate the FT PACS design (16). This system, called a PACS simulator, consists of a radiology information system simulator, an acquisition modality simulator, a gateway, a PACS controller, two clinical viewing workstations, and a network infrastructure. Clinical PACS images are sent manually or automatically from the modality simulator through the various components of the PACS simulator during testing and evaluation.

Evaluation Protocol

Development of a protocol for testing and evaluation is centered around creating clinical scenarios in which hardware failure can occur. In addition, simulations of hardware failure are created within these scenarios to test the FT abilities of the CA PACS server. Clinical scenarios are categorized according to the perspective of the end user, which is important because ultimately the success of the FT of the system benefits the end user the most. The two types of scenarios are clinical background (“passive”) scenarios (eg, storage, automatic distribution of PACS examinations) and clinical on-demand (“active”) scenarios (eg, query and retrieval).

Failover Procedures

Three types of hardware failures were simulated on the PACS controller running on the CA PACS server: *(a)* failure of network devices in the PACS controller or of the network switch connected to the controller, *(b)* failure of the CPU, memory, or entire motherboard of the PACS controller, and

(c) failure of the external hard disk or some other storage device. The three types of failure were initiated during the query and retrieval of a PACS examination from the CA PACS archive server to the viewing workstation. Each of these scenarios was tested a total of 182 times. Ethernet failure required 3–15 seconds of recovery time, CPU module failure required 20–40 seconds, and disk failure required 35–75 seconds. The PACS data flow continued automatically after system recovery. No data loss or interruption of data flow was observed, and all patient examinations were successfully transferred. From the perspective of the user at the display workstation, the delay was equal to the system failover time.

Background Burn-in Testing

To create an effective clinical simulation, the PACS archive server should perform background functions continuously. An automatic loop was created to continuously simulate clinical scenarios for long periods of time in the evaluation of the FT PACS design, and any hardware failures were noted and recorded. The automatic feedback distribution loop between the PACS controller, gateway, and viewing workstation was activated continuously for 3 months. The burn-in period simulated a PACS server operating continuously in a clinical environment. No system failure was observed.

Other Clinical Applications

Another application of the CA PACS archive server is its use as an FT solution for disaster recovery of short-term image data with an application service provider (ASP) model (17,18). The ASP short-term image archive provides instantaneous, automatic off-site backup of acquired image data and instantaneous recovery of stored image data with CA quality and at low cost. Such an application has been implemented in our laboratory. The CA image server has both a RAID system and a DLT library and serves as a short-term off-site backup archive server for SJHC. All clinical image data are sent to this ASP CA image server in parallel to the radiologic examinations that were archived in the main server at SJHC. Currently, connectivity between the main archive and the ASP storage server is established via a T-1 (1.5 Mbits/sec) connection. In the near future, Internet 2 will replace the T-1 connection (Fig 2), allowing a transfer rate of 155 Mbits/sec.

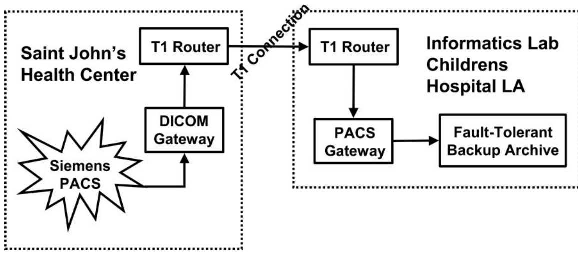


Figure 2. Diagram illustrates the configuration of the ASP model off-site backup CA PACS archive server between SJHC (clinical site) and the Image Processing and Informatics Laboratory at Children's Hospital Los Angeles/University of Southern California (backup server site). Note the Digital Imaging and Communications in Medicine (DICOM) gateways serving as transmission buffers between the two sites to minimize impact on clinical PACS performance.

A disaster scenario was initiated during the implementation stage, and the disaster recovery process with use of the ASP archive server was successful in repopulating the clinical system on-site within a short period of time (a function of the amount of data and the transfer rate). The ASP archive was able to recover 2 months of image data with no complex operational procedures. Furthermore, no image data loss was encountered during the recovery. This ASP off-site backup archive has been integrated into the SJHC PACS for daily clinical operation.

The CA image server has also been implemented as a backup archive solution for downtime events scheduled for the main server located on-site. Software upgrades and preventive maintenance are routinely performed for the main image server in the hospital. Although these downtime events are scheduled, they still affect normal clinical workflow because these main image servers are mission-critical systems. Such an event occurred recently at SJHC, where the main image server on-site was scheduled for software upgrades (19). The CA image server, which was off-site, was used to provide image data for emergency operation during this downtime. A total of 100 images (87.5 MBytes) were transmitted to the SJHC clinical system directly from our laboratory during a 4-hour period.

Summary and Conclusions

We have implemented a TMR CA PACS archive server with integrated external RAID and DLT storage devices. The server can automatically detect failures by comparing the activity of three synchronized modules. Each module performs exactly the same activities, and when a module fails, the system detects the deviation, identifies the faulty module by simple majority vote, and removes it from service. The remaining two modules continue normal operation, with no adverse effect on PACS data flow or system performance. The CA PACS archive server has been evaluated by conducting a series of manually induced failover tests of its reliability, functionality, and performance. We believe that a simple, low-cost TMR CA PACS archive server can be implemented and used as a reliable PACS server to achieve CA. Meanwhile, our performance measurements provide a valuable benchmark for CA PACS design and implementation in clinical PACS environments and for PACS applications.

As PACS become more widely used in hospitals, there will be a greater need for CA PACS solutions because a PACS is considered to be mission critical for around-the-clock daily clinical operation. As discussed earlier, failure of the main PACS archive server can cripple an entire PACS operation. The operational reliability of the PACS server is always a major concern in a filmless hospital environment. The design and implementation of the CA PACS archive server described in this article provide maximum uptime for end users while ensuring the transactional integrity of all clinical PACS data. In the event of a hardware failure in the CA PACS, users will notice only minimal performance impact without interruption of clinical data flow or loss of data. To help prevent a disaster that could physically destroy even a CA PACS archive server located on-site, an off-site backup CA PACS archive server can be implemented to store the PACS data for recovery and use. This backup server can also be used as a temporary archive server during downtime events scheduled for the clinical PACS server on-site. These applications underscore the importance of a CA PACS archive server solution within a filmless environment.

References

1. Huang HK. PACS: basic principles and applications. New York, NY: Wiley, 1999.
2. Institute of Electrical and Electronics Engineers. A compilation of IEEE standard computer glossaries: IEEE standard computer dictionary. New York, NY: Institute of Electrical and Electronics Engineers, 1990.
3. Reiner BI, Siegel EL, Dwyer SJ III, eds. Security issues in the digital medical enterprise. Vol. 1. Great Falls, Va: Society for Computer Applications in Radiology, 2000.
4. Compaq/Tandem's "non stop Himalaya." Available at: *http://www.tandem.com.* Accessed February 2001.
5. IBM S/390 parallel sysplex. Available at: *http://ibm.com.* Accessed February 2001.
6. Sun Microsystems. Available at: *http://www.sun.com.* Accessed February 2001.
7. Available at: *http://www.highavailabilitycenter.com.* Accessed February 2001.
8. Huang HK, Cao F, Zhang J, Liu BJ, Tsai ML. Fault tolerant picture archiving and communication system (PACS) and teleradiology design. In: Reiner BI, Siegel EL, Dwyer SJ III, eds. Security issues in the digital medical enterprise. Vol. 1. Great Falls, Va: Society for Computer Applications in Radiology, 2000; 57–64.
9. Carrasco JA. Computationally efficient and numerically stable reliability bounds for repairable fault-tolerant systems. IEEE Trans Comput 2002; 51:254–268.
10. Xu J, Randell B, Romanovsky A, et al. Rigorous development of an embedded fault-tolerant system based on coordinated atomic actions. IEEE Trans Comput 2002; 51:164–179,402–409.
11. Bajpai G, Chang BC, Kwatny HG. Design of fault-tolerant systems for actuator failures in nonlinear systems. Proc Am Control Conf 2002; 5:3618–3623.
12. Muller G, Banatre M, Peyrouze N, Rochat B. Lessons from FTM: an experiment in design and implementation of a low-cost fault tolerant system. IEEE Trans Reliability 1996; 45:332–340.
13. Huang HK, Cao F, Liu BJ, et al. Fault-tolerant PACS server design. SPIE Med Imaging 2001; 4323:83–92.
14. Resilience Company. Ultra2 Solaris servers. Available at: *http://www.resilience.com.* Accessed February 2001.
15. Avizienis A. Toward systematic design of fault-tolerant systems. Computer 1997; 30:51–58.
16. Zhou Z, Law M, Huang HK, et al. Educational RIS/PACS simulator. SPIE Med Imaging 2003; 5003:139–147.
17. Liu BJ, Huang HK, Cao F, Documet L, Sarti DA. A fault-tolerant back-up archive using an ASP model for disaster recovery (abstr). Radiology 2001; 221(P):741.
18. Liu BJ, Huang HK, Cao F, Documet L, Muldoon J. A fault-tolerant back-up archive using an ASP model for disaster recovery. SPIE Med Imaging 2002; 4685:89–95.
19. Liu BJ, Huang HK, Cao F, Documet L, Muldoon J. Clinical experiences with an ASP model backup archive for PACS images (abstr). Radiology 2002; 225(P):313.

Original Investigations

A Continuous Available (CA) Server for Medical Imaging Applications[1]

H. K. Huang, DSc, FRCR(Hon), Brent J. Liu, PhD, Z. Zhou, MS

Rationale and Objectives. The trend of medical imaging research and application is toward large database management and manipulation, which requires a robust image server to receive image data from sources and to deliver them to users reliably and in a timely fashion. This article describes the design, implementation, and clinical applications of a continuous available (CA) image server for these purposes.

Materials and Methods. The design of the CA image server is based on the concept of a triple modular redundancy server with three redundant server modules. Coupled with a majority voting mechanism in the three modules and failover software, the triple modular redundancy server takes care of all single points of failure hardware components in the CA image server automatically to achieve fault tolerance. Methods and procedures of evaluating the fault tolerance system reliability caused by network connectivity, motherboard, and disk storage failures are described.

Results. Thorough experimental results in laboratory and clinical environments verify that the image server achieves 99.999% hardware up time (or 5 minutes/year down time), satisfying the industrial terminology of hardware continuous availability. Performance of failover of the CA image server is automatically tabulated during these procedures.

Conclusion. Applications of CA image server are extensive. Two examples are given including Picture Archiving and Communication System, and off-site back-up archive using the Application Service Provider model. As designed, the CA image server is portable, scalable, affordable, easy to install, and requires no human intervention during failover and system recovery.

Key Words. Archive server; continuous availability; fault tolerance; medical imaging; picture archiving and communication system (PACS); system failover.

CONTINUOUS AVAILABILITY AND FAULT-TOLERANCE

The trend of medical imaging research and application is toward large database management and manipulation. Any large medical imaging database for clinical, research, or education application requires a robust archive server to receive image data from sources and deliver them to users reliably and timely (1). In consideration of the operation of a large integrated computer networked server system, high availability and continuous availability (CA) are the two commonly used terms categorizing the degree of system reliability (2,3). There is a continuum of high-availability solutions, ranging from a 99% system availability rate (88 hours downtime/year) by using the simple "hot spare" technology to 99.99% availability (1 hour downtime/year) with clustered server-based solutions using hardware and fail-over software. The fully hardware "fault-tolerance" (FT) systems are known as CA solutions

Acad Radiol 2004; 11:767–778

[1] From the Division of Imaging Informatics, Department of Radiology, The Keck School of Medicine, University of Southern California, Information Sciences Institute, 4676 Admiralty Way, Suite 601, Marina del Ray, CA 90292. Received March 22, 2004; accepted March 26. Supported in part by the US Army Medical Research and Material Command Contract No. DAMD17-99-P-372, NIH National Library of Medicine Contract No. N01-LM-9-3538, and Reliant Care, Inc No. 30-8034-400391. **Address correspondence to** H.K.H. e-mail: hkhuang@aol.com

doi:10.1016/j.acra.2004.03.053

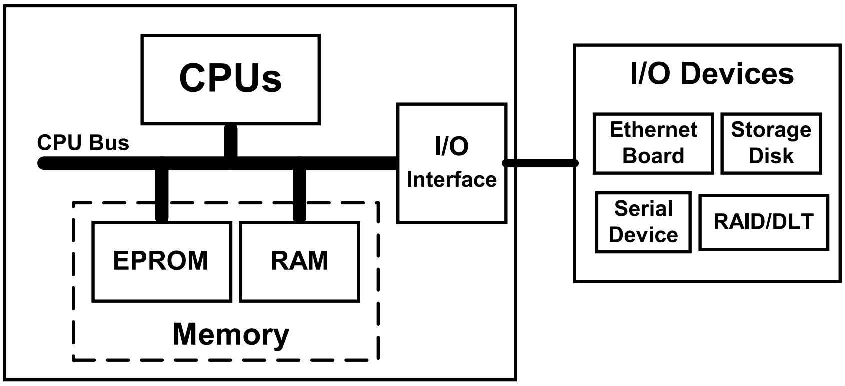


Figure 1. General computer system architecture with a short-term (RAID) and long-term (DLT) archive.

with the highest 99.999% availability rate and the minimum downtime of 5 minutes/year (4–6). We will use the terms FT and CA interchangeably in this article. Examples of these systems used in medical imaging applications are PACS (picture archiving and communication system), off-site back-up archive, and on-line computer-aided detection and diagnosis system (CAD). PACS and off-site back-up archive are both mission-critical operations, continuing 24 hours a day and 7 days a week, and it requires CA. In CAD, remote access of preliminary indication or second opinion of the patient's condition using CAD server based on imaging examination is time critical (1). This article describes the design and implementation of, and clinical experience with, a CA image server for large-scale medical imaging applications.

HARDWARE COMPONENTS IN AN IMAGE SERVER

Basic hardware components in an image server consist of the central processing units (CPUs) and memory, I/O ports and devices, and storage devices shown in Figure 1. Any of these hardware components can fail and if it is not addressed immediately, operation of the server will be compromised. The design of the CA image server is to develop both hardware and system software redundancy to automatically detect and recover any hardware failure instantaneously. Table 1 shows a current survey on the causes of computer and network system downtime in which computer hardware, hard disk drive, communication processor, and data communication network account for approximately 71% of the system failure (7). In this article, we address the issue of this 71% hardware-related failure in the complete image server system, and methods of circumventing the failure with failover to achieve the definition of a CA system. Application software failure and human errors would not be considered in this CA image server.

DESIGN CRITERIA

Fault tolerance is the implementation of redundant hardware components with software control in the server, such that in the event of a component failure, maximum availability and reliability (without the loss of transactional data) can be achieved (8–13). The requirements for an FT image server are to support:

High Reliable Server System Operations

- No loss of data
- No workflow interruptions

Image server reliability means server uptime and FT. In the event of a server hardware component failure, users might notice a minimal performance impact during the

Table 1
Causes of Computer and Network System Downtime

Computer hardware	24%
Hard disk drive	26%
Communication processor	11%
Data communication network	10%
Software	22%
Human error	6%
Others	1%
Total	100%

server failover process, but all server functions such as image archive, retrieval, distribution, and display must be continuously available. No loss of data and no work interruptions are permitted. All current server processes and transactions should automatically resume with no interruptions.

Acceptable System Performance

- No performance degradation in daily routine operations
- Occasional glitches in an acceptable amount of time

Performance inclusively measures how well hardware, operating systems, network, and application software perform together. Ultimately, the image server performance affects the end users and the response time of the applications. For FT image servers, the redundant server hardware once taking over the failed hardware should be able to handle the same workload in network speed, CPU power, and archive storage as in normal server operations so that the user will not have noticeable performance degradation. Normally, server operations and user sessions halt momentary (about 30 seconds, see *Results* section) until the FT image server successfully fails over and resumes in the event of a system glitch. A longer delay in minutes is acceptable for non-interactive background processes such as image archiving.

Low-Cost and Easy Implementation

- Portable
- Scalable
- Affordable

Portability means that existing server software should be able to run on the FT image server without any major changes (14). Scalability tests how additional hardware, system and application software work with the FT server, and impact its ability to handle the workload. High-end million-dollar FT machines such as the Tandem system (Hewlett-Packard, Palo Alto, CA), which uses a sophisticated system design and can recover from system failure in milliseconds, is too expensive for most large-scale medical imaging application, for example in PACS, and is often used for short-transaction types of applications in banking, security and stock exchange, and telecommunication industries. Our design of the FT server using the concept of the triple modular redundancy (TMR) UNIX server is affordable, the prototype hardware costs about three times the amount of a comparable Unix machine but has much longer failover times (in seconds) than the Tandem system.

MATERIALS AND METHODS

In this section, we describe the general architecture of the CA image server by presenting the triple modular redundant server followed by its peripheral components.

The Triple Modular Redundant Server

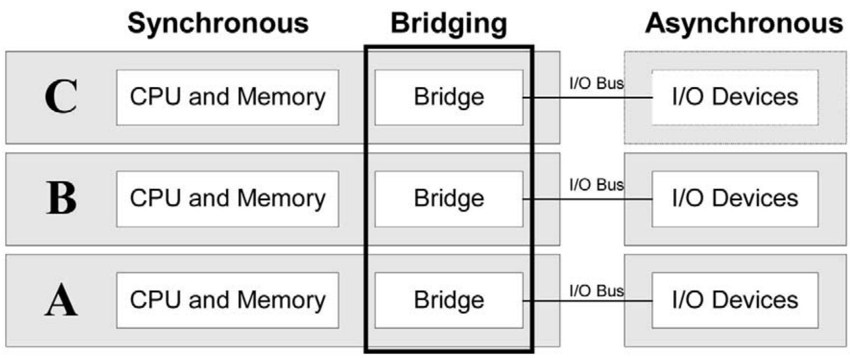
The CA image server uses TMR to achieve FT at the CPU/memory level. Figure 2 shows the core of the TMR server made up of three identically configured UltraSPARC-based modules (Sun Microsystems, Santa Clara, CA) (6). The three modules are tightly synchronized and interconnected through a high-speed backplane for intermodule communications. Each module is a complete, operational computer running Sun's Solaris Unix OS server, with its own UltraSPARC CPU, memory, I/O interfaces, bridge logic, and power supply. Each module runs all software applications independently and synchronously under the standard Solaris operating environment (15,16).

To the UltraSPARC core, a programmable application-specific integrated circuit technology is used to build TMR bridge logic that keeps the three modules synchronized, continuously monitors and compares their operation, and exchanges I/O data among the three modules. The bridge logic reads S-bus transactions within each module and compares them across the modules. If the logic detects a variation between transactions, it assumes an error. The system then pauses (typically for 5–30 seconds depending on memory size, see *Results* section) while the diagnostic software determines the faulty module and disables it. Next, memories of the two remaining modules are synchronized by performing a full memory copy. Once the copy is complete, the system resumes processing by using the remaining two modules (17,18).

Other Server Components–Fault-Tolerance I/O and Storage Subsystem

Complete System Architecture

Other server components in addition to the TMR in the server system are I/O buses and devices (19). Operating I/O devices such as SCSI or Ethernet interfaces in TMR synchronization is not only extremely difficult, it does not achieve the required system-wide CA. Each I/O sub-



a.

b.

Figure 2. **(a)** Fault-tolerant (CA) image server with the TMR architecture. Each module contains the basic computer system architecture shown in Fig 1 (left). The bridges are used to synchronize the three modules. **(b)** The CA TMR with three modules (Ultra Sparc II-like architecture) with three separate power supplies (bottom). The console switch (middle) is for managing the three module displays during the debugging stage.

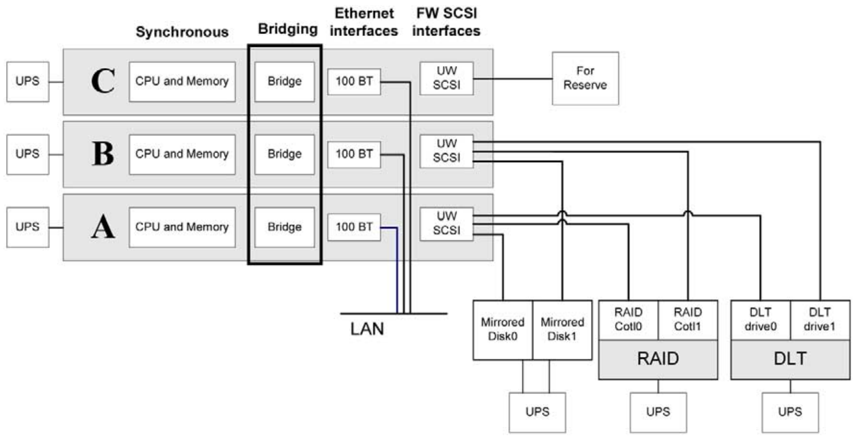
system must be regarded individually, with CA implemented in a manner appropriate to each subsystem. Figure 3 shows the complete architecture of the TMR and I/O buses and devices.

Ethernet

Each of the three modules contains its own 100baseT Ethernet interface, each of which is connected, via independent paths, to the local network backbone. The three interfaces form a single software interface with one IP and media access control (a unique hardware number) address. One interface acts as the active interface, while the others stand by. Should the module containing the active interface fail, or some element of its connection to the backbone fail, that interface is disabled and a standby unit becomes active in its place. Normal network retry mechanisms hide the failure from applications.

Fault-Tolerant Storage System

There are two storage subsystems, RAID (redundant arrays of independent disks) for short-term and DLT (digital linear tape) library for long-term storage. Because the former is mission-critical in a complete CA image server system, we have designed FT in the RAID system. DLT in the CA image server is used as a peripheral storage



a.

b.

Figure 3. **(a)** The complete CA image server system architecture. Three Ethernet interfaces from each of the three modules connected to the LAN, while only one Ethernet interface is active and forms only one IP address for the application. Two mirrored disks (mirrored disk0 and mirrored disk1) are connected to module A and module B through UW (ultra-wide) SCSI interface with failover mode. Two RAID controllers (RAID Cotl0 and RAID Cotl1) are connected to module A and module B through UW SCSI interface with failover mode. Two DLT controllers (DLT drive0 and DLT drive1) are connected to module A and module B through UW SCSI interface with failover mode. In Section III: system evaluation explains how this fault-tolerant architecture performs failover under various conditions. **(b)** The CA image server connected to the PACS simulator shown during the meeting of the Radiological Society of North America, 2001 (20,24), and SPIE 2002 (26). The simulator consists of the modularity simulator (left), Gateway, TMR (see Fig 2b), and two workstations (right). The RAID and LTD library are not shown in the figure. UPS, uninterruptible power supply.

subsystem, FT is designed and implemented up to the connectivity level with the CA image server.

Redundant arrays of independent disks as a short-term storage.—Although RAID has its own built-in FT mechanism in handling disk failure, its single points of failure (SPOF) are in the RAID controller, and its connection to the CA image server. In our design, a dual-controller RAID is used for short-term storage. The TMR module A and B are connected to each of the two RAID controllers as shown in Figure 4. Disk drive or RAID controller failures are handled by the RAID mechanism. The two redundant connections shown in Figure 4 provide a full FT short-term storage solution, which guarantees the system survival in an event of failure of one module, one RAID controller, or any combination of both at the same time.

A Hitachi 9200 (325 Gbytes) dual-controller RAID (Hitachi Data Systems, Santa Clara, CA) with Veritas Volume manager software (Mountain View, CA) has been implemented with the CA image server shown in Figure 4. The dual hardware controllers connected to the module A and B, respectively, provides the two paths to the TMR server while Veritas software dynamically monitors the two paths and switches auto-

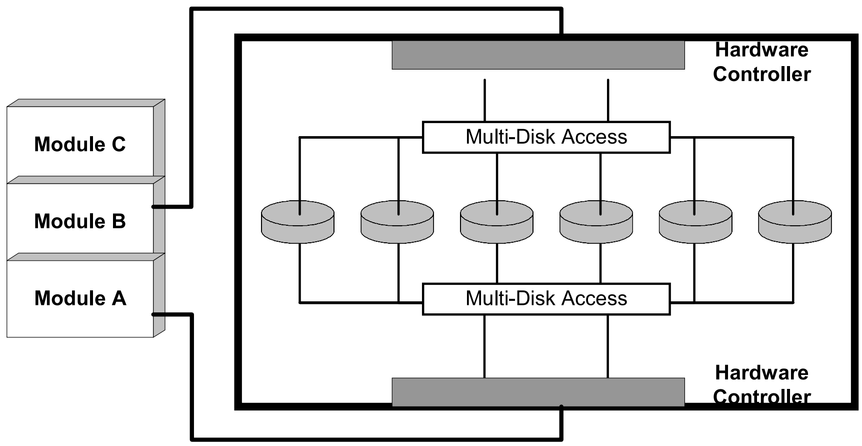


Figure 4. A RAID with two hardware (HW) controllers. Module A SCSI port connects to one controller and the module B SCSI port connects to the other controller providing redundancy for system reliability.

matically from one to another in case one path is disconnected.

Digital Linear Tape Library for Long-term Archive

For long-term archive in the CA image server, a StorageTeK L40 DLT library with 3.2TB (Storage Technology Corp, Louisville, CO) is used. The library has two drives, each of which is connected to one of the TMR controller modules, providing the redundant paths to the CA image server, and hence it provides FT connectivity to the Server. Meanwhile, Veritas Storage Migrator and Netbackup software are installed to automatically migrate and back-up the data from short-term RAID to long-term tape library archive. The Veritas Migrator software has a built-in feature to monitor the multiple paths and to fail-over from one to another in case of a failure of any single CPU module or tape driver. But the DLT library itself is not designed as fault-tolerant. Its controller and robot arm are still the SPOF. The tape library is used for secondary archive and the FT can be tolerated. The main concern is to preserve the data instead of real-time recovery of library system failure.

RESULTS

System Evaluation

Testbed Development

To evaluate the robustness and effectiveness of a CA image server design, two key components are crucial to the process. First, a testbed environment was developed to allow for observation and results gathering. Second, the key SPOF were realized and defined as targets for replacement within the new design. In the specific case of a CA image server, these (SPOF) are manifested in system components or devices.

The Image Server Simulator System.

A system that simulates some of the functions of an image server was developed to evaluate the CA design (20). It is comprised of:

1. Acquisition modality simulator: Simulates a device that acquires medical images
2. Gateway: Receives images from acquisition modality and verifies no image loss
3. Image server: Receives images from gateway for storage and distribution (Sun Ultra Sparc II workstation running Solaris v 2.6) (Sun Microsystems, Santa Clara, CA)
4. Two workstations: Receives images from image server and displays them (Cedara Display software, Toronto, Canada)
5. Network infrastructure

The simulator is a closed network using TCP/IP Fast Ethernet consisting of CAT5 cables and two portable 8-port 100 Mbit/sec switches. Each of the system components resides on separate computer devices (see Fig 5). The modality, gateway, and two workstations are running each on four separate Pentium III Windows 2000 PC workstations. The image server is running on a Sun Ultra

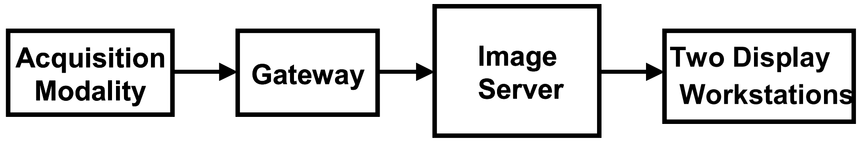


Figure 5. Image server simulator system components and data flow used in the CA image server system evaluation.

Sparc II Solaris v 2.6 workstation. In addition, the simulator has an automatic software package where images are sent from the gateway to the image server, then deleted and resent again in a continuous loop. This function aids the burn-in evaluation to be discussed later.

The CA image server simulator system

Initially, the simulator was designed with the image server running on a Sun Ultra Sparc II Solaris v 2.6 workstation (21–23). The image server is the main engine for the system and crucial for image data flow. Therefore, it is considered a SPOF component within the system. Before implementing the CA image server, all components in the simulator were tested to assure that the image data flow was complete. This was accomplished by executing the image data flow multiple times. Results were verified at the image server. To evaluate the CA image server, the image server in the simulator system was replaced with the CA image server described in the *Materials and Methods* section (Fig 2). Identical software running on the image server was installed on the CA image server. The same testing procedures and data used to verify the simulator were applied to the CA image server to insure that there were no software, setup, or configuration differences while implementing the evaluation protocol.

Continuous Available Image Server Testing and Performance Measurement

Two sets of testing were used to evaluate the reliability and performance of the CA imaging server during failover. Details and results are described in this section.

Evaluation Protocol for Continuous Available Image Server Reliability and Functionality Tests

Define and create clinical test scenarios.—Development of an evaluation protocol is anchored around creating operational situations with which hardware failure can occur. Simulations of hardware failures are created within the operation scenarios to test the abilities of the CA image server. Operation scenarios simulated in the evaluation can be broken down into two types based on the perspective of the end-user: operational background or "passive" scenarios, and operational on-demand or "active" scenarios.

Operational background or "passive" scenarios are automatic functions of the CA image server. These include storage and archival of image data and automatic distribution of this data to workstations. An effective operational simulation would be to have the CA image server perform these background functions on a continuous basis for 24 hours a day, 7 days a week. An automatic loop was created within the simulator testbed. The loop involves the gateway that automatically sends image data to the CA image server that in turn automatically distributes the image data to the workstations. After a set amount of time elapses, the CA image server automatically deletes image data from its storage and database and the data are resent. This loop can be executed continuously for extended periods of time to simulate a true operational environment. This scenario was performed in the evaluation protocol of the CA image server design and any hardware failure occurrences were noted and recorded.

Operational on-demand or "active" scenarios are functions of the image server executed based on requests made by another device. For example, a workstation can query the image server for specific image data. A retrieve request is initiated for the image data to be sent to the requesting device. This scenario is executed with the CA image server. Failover procedures are described in greater detail in the forthcoming paragraphs.

Failover procedures for evaluation.—Three types of hardware failures were simulated on the CA image server: (1) network devices/components failure; (2) CPU, memory, or entire motherboard failure; and (3) hard disk/storage failure.

Scenario 1: Ethernet connection failure.—The first failover procedure was evaluated during both receiving and transmitting of image data from the CA image server. During transmission of image data to the workstation, the Ethernet cable connecting the CA image server to the network switch is removed, simulating a network component failure. Figure 6 shows this failover procedure. The

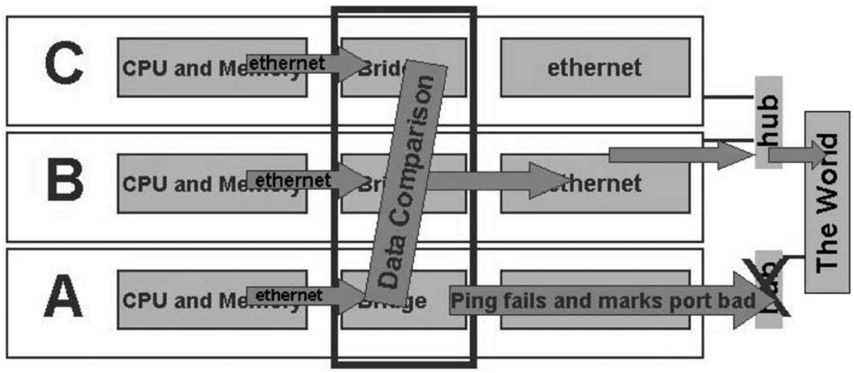


Figure 6. Ethernet failover procedure on module A where CA image server marks port bad and arrows signify the selected data transmission goes through module B.

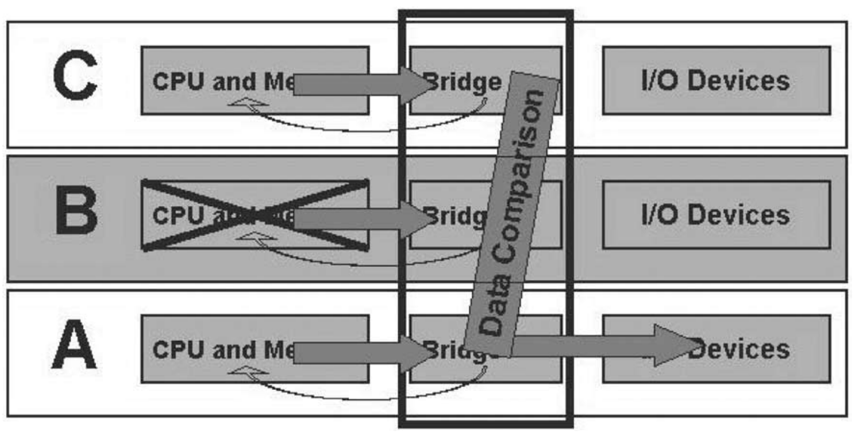


Figure 7. CPU failover procedure on module B where CA image server compares data and marks module B as failed and arrows signify data transmission goes through module A.

image data should continue to transfer after a few seconds once the CA image server has completed its Ethernet failover automatically. Successful transmission of image data during this failover procedure is verified by displaying the image data on the workstation and confirming all data is present and not corrupted. Successful receiving of image data is verified by sending the image data to the workstation for verification.

Scenario 2: CPU, memory, motherboard failure.—The same failover procedure was applied to the simulated hardware failure of the CPU, memory, the entire motherboard, or the power supply. Power to one of the three modules is shut off during the transmitting or receiving of image data either by turning the power switch off or pulling the power plug. Figure 7 shows this failover procedure. The image data should continue to transfer to once the CA image server has completed its CPU module failover. Again, successful transmission of image data during this failover procedure is verified by displaying the image data on the workstation and confirming all data is present and not corrupted. Successful receiving of image data is verified by sending the image data to the workstation for verification.

Scenario 3: Hard disk or storage device failure.—The same failover procedure was applied to the simulated hardware failure of the external hard disk or other storage device. Power to one of the hard disks is shut off during the transmitting or receiving of image data either by turning the power switch off or pulling the power plug on the external hard disk. Figure 8 shows this failover procedure. The image data should continue to transfer to the proper destination once the CA image server has completed its hard disk failover. An additional step is involved where if the hard disks are mirrored, the data needs to be re-synced once the hard disk is recovered after failover. The same verification procedures were performed as above.

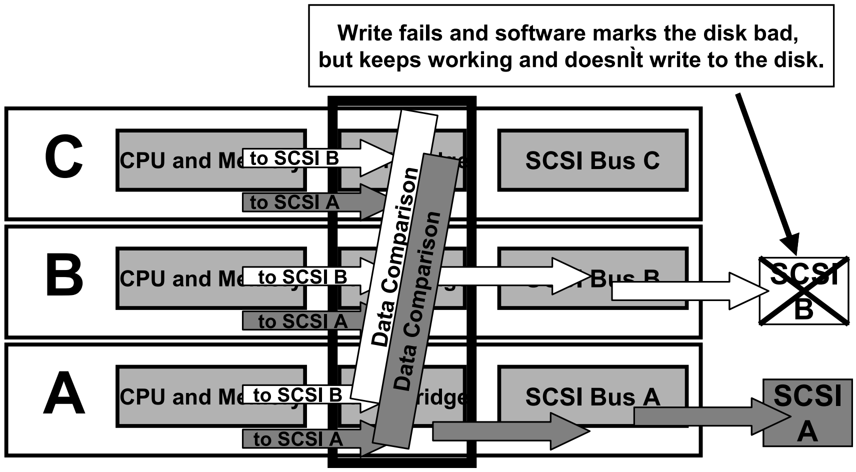


Figure 8. Hard disk failover procedure. In a normal operation, CA image server writes simultaneously to SCSI disks A and B. When the disk attached to module B fails, CA image server marks the disk B bad and arrows signify data written instead to the hard disk connected to module A only.

Fault-Tolerant Server Performance

Interactive test scenarios.—Performance test measures the failover time of CA image server in an event of system component failures that occur during standard image server operations–receiving and distributing image data. The same three scenarios described previously were used.

Test scenario 1: Ethernet failure while receiving and distributing image data.
Test scenario 2: CPU module failure while receiving and distributing image data.
Test scenario 3: Disk failure while receiving and distributing image data.

Table 2 summarizes the recovery time for the scenarios listed above. The image dataflow continues automatically after the system recovery. No data loss and no interruption of data flow had been observed. All image data were successfully transferred. Effect from the user's perspective is a delay equal to the system failover time. The three scenarios were each tested 182 times using test data of 111 images totaling 28 MB in size. The recovery time (T_r in seconds) was measured:

$$T_r = T_t - T_f$$

Where T_t is the time measured when image data transmission is continued and $T_f$is the time measured at the onset of the failure.

Standard deviations (SD) were calculated based on the sample size of the 182 executed failures for each scenario. Note that for Scenario 3, the high SD is because of two different types of failure times: (1) SCSI failure captured by the system immediately; and (2) SCSI failure captured by a SCSI ping timeout, which is the longer failover time.

Background burn-in test.—The automatic feedback distribution loop described in the *Materials and Methods* section between the CA image server, gateway, and workstation was activated for a continuous period of 3 months (2,160 hour). This is the burn-in period that simulates an image server in a clinical setting. No system failure was observed.

With the CA image server, applications execute concurrently on highly replicated hardware. In the event of a failure, work continues on the remaining and still fully functional hardware. The system and network performance are not affected except during the failover process that users will experience a delay equal to the system re-

Table 2
Average System Recovery Time for Common Faults. Total tests = 182 for Each Scenario. Unit: Second

Common Faults	Ethernet	CPU Module	Disks
Range of recovery time	3–15	20–40	35–75
Average	5.8	29.8	42.2
Standard deviation	1.05	4.66	32.86

cover time. While the fault recovery may be effectively addressed by the CA design of hardware and system, the time required to recover from errors can vary from couple of seconds to 75 seconds depending on one or more of the following:

1. The nature of the failure
2. The elapsed time to discovery of the failure
3. The time required to resynchronize the functional hardware
4. The time required to re-establish network communications with switches
5. The time waiting for a SCSI timeout and then taking the alternative path to the external storage device

The CA image server with the simulator (Figs 3B and 5) was demonstrated live at the Annual Conference of the Radiological Society of North America, in Chicago, IL during 2000 (21,22), 2001 (23,24), 2002 (27), and SPIE (The International Society of Optical Engineering), Medical Imaging Conference, San Diego, CA, 2001 (25) and 2002 (26).

Applications of the Continuous Available Image Server

We give two medical imaging applications of the CA image server. The first is in PACS and teleradiology (21,23–27,31) operation, and the second is short-term off-site image back-up archive using the Application Service Provider (ASP) model (28–30).

Picture Archiving and Communication System and Teleradiology

Use CA image server to replace single points of failure in PACS imaging components. —PACS and teleradiology are very large-scale system integrations within the health care information technology regime. Multiple computer devices, imaging devices, network technologies, display workstations, and information systems have to be seamlessly integrated together to provide effective and efficient operation. Among these components, many have their own FT design, but some do not. Many PACS image servers still do not have an industry-standard FT design because it is very expensive and technically tedious to implement involving many components. Some PAC systems use Tandem architecture for its archive, but it is very costly and the failure recovery procedure is tedious and manually intensive. Other PAC systems use redundant components, but without elegant failover architecture. The CA image server described here can be used to replace SPOF in the image components of the PACS operation (31). The CA image server simulator described in the *Results* section is actually a generic PACS designed for PACS training classes which has not experienced failure for over 1 year because of the FT architecture.

Application Service Provider Mode Image Archive

Off-site back-up archive–Application service provider model for disaster recovery.—Another application of the CA image server is used as a fault-tolerant solution for disaster recovery of short-term image data using an ASP model. The ASP short-term image archive provides instantaneous, off-site automatic back-up of acquired image data and instantaneous recovery of stored image data with CA quality and low operational cost. Such an application has been implemented in our laboratory. The CA image server with both RAID and DLT library as shown in Fig 3 located in our laboratory serves as a short-term off-site back-up archive server for the St John's Health Center in Los Angeles, CA. One hundred percent clinical image data is sent to this ASP CA image server in parallel to the exams being acquired from modalities and archived in the main server from St John's. Currently, connectivity between the main archive and the ASP storage server is established via a T-1 (1.5 Mbits/sec) connection.

During the implementation stage, a disaster scenario was initiated and the disaster recovery process using the ASP archive server was successful in repopulating the clinical system onsite within a short period of time (a function of the data size and data transfer rate). The ASP archive was able to recover 2 months of image data with no complex operational procedures. Furthermore, no image data loss was encountered during the recovery. Table 3 shows the number of image data in terms of clinical image exams that were tested on this system. A total of approximately 447 exams comprising of 29,000 images of various types or 9 GB of total data were tested. The average T1 performance bandwidth was measured as 179 Kbytes/sec using the FTP protocol and 168 Kbytes/sec using the DICOM transfer protocol.

This ASP off-site back-up archive has been integrated with the St John's PACS for daily clinical operation.

Offsite back-up image archive–Scheduled downtime service.—The CA image server has also been implemented as a back-up archive solution for scheduled downtime events that occur to the main server located onsite.

Table 3
Results of Image Data Stored Using the ASP Backup Archive during a Disaster Recovery Operation

	CT Exams	MR Exams	CR Exams	US Exams	RF/Angio Exams	Total
No. of exams	111	110	109	61	56	447
Total images	10,385	15,977	164	1,981	497	29,004
Average no. of images per exam	93.6	145	1.5	32.5	8.9	
Data size (MB)	5,192.5	998.6	1,312	495.3	994	8,992.4

Routinely, the main image server within the hospital will undergo software upgrades as well as preventative maintenance. Although, these downtime events are scheduled, they still effect normal clinical workflow as these main image servers are mission-critical systems. Recently, such an event was performed at St John's where the main image server onsite was scheduled for software upgrades. The CA image server, which was offsite, was used to provide image data for emergency operation during this downtime period. A total of 100 images (87.5 MB) were transmitted to St John's clinical system directly from our laboratory during a 4-hour period.

DISCUSSION

We have described the design and implementation of a CA image server that can achieve 99.999% hardware up time. The design concept is based on a TMR server with three redundant server modules. The FT and failover is based on coupling the TMR with majority vote mechanism and failover software architecture. The majority votes detect the faulty component in cycle time, and the software takes care of the automatic failover. Depending on the types of hardware failure and the state of the execution in the component, the failover can take from 3 to 75 seconds (SD = 32.86 seconds). The longer failover time and large SD happen in the computer disks where the SCSI failure can either be captured by the system immediately or by a SCSI ping timeout which yields the large SD. In the case of a mirrored disk failure, the CA image server continues to run smoothly, but to reconstruct the mirrored disk in the background does require more time depending on the size of the disk under consideration.

Although the concept of TMR logic has been used in other types of FT design, the TMR CA image server described and implemented is a technology innovation in its handling of failover. The CA image server connected with the FT storage as a total system for medical image applications has several main advantages over other current FT server designs. It is CA, lower cost to implemented, portable, scalable, affordable, easy to install without extensive change in the application software, and not manpower-intensive during failover and system recovery. The TMR CA image server is very suitable for large-scale medical image database application like PACS and ASP back-up archive.

REFERENCES

1. Huang HK. PACS and imaging informatics. Hoboken, NJ: John Wiley & Sons, 2004.
2. Institute of Electrical and Electronics Engineers. A compilation of IEEE standard computer glossaries. IEEE Standard Computer Dictionary. New York, NY: 1990.
3. Reiner BI, Siegel EL, Dwyer SJ III. Security issues in the digital medical enterprise. Society for Computer Applications in Radiology, November 2000.
4. Compaq/Tandem's non stop Himalaya. Available at: http://www.tandem.com.
5. IBM S/390 Parallel Sysplex. Available at: http://ibm.com.
6. Sun Microsystem. Available at: http://www.sun.com.
7. Resilience Corporation Technical Report, 2000. IMEX Research.com, 2000.
8. Carrasco JA. Computationally efficient and numerically stable reliability bounds for repairable fault-tolerant systems. IEEE Trans Comput 2002; 51:254–268.
9. Xu J, Randell B, Romanovsky A, Stroud RJ, Zorzo AF, Canver E, Von-Henke F. Rigorous development of an embedded fault-tolerant system based on coordinated atomic actions. IEEE Trans Comput 2002; 51: 164–179 and 402–409.
10. Bajpai G, Chang BC, Kwatny HG. Design of fault-tolerant systems for actuator failures in nonlinear systems. Proc Am Control Conference 2002; 5:3618–3623.
11. Some RR, Beahan J, Khanoyan G, Callum LN, Agrawal A. Fault-tolerant systems design–estimating cache contents and usage. Proc IEEE Aerospace Conf 2002; 5:2149–2158.
12. del Gobbo D, Cukic B, Napolitano R, Easterbrook S. Fault detectability analysis for requirements validation of fault tolerant systems. High-assurance systems engineering. Proceedings of 4th IEEE International Symposium, 1999; 231–238.
13. Fabre JC, Perennou T. A meta object architecture for fault-tolerant distributed systems: the FRIENDS approach. IEEE Trans Comput 1998; 47:78–95.

14. Muller G, Banatre M, Peyrouze N, Rochat B. Lessons from FTM: an experiment in design and implementation of a low-cost fault tolerant system. IEEE Trans Reliability 1996; 45:332–340.
15. Kanoun K, Ortalo-Borrel M. Fault-tolerant system dependability-explicit modeling of hardware and software component-interactions. IEEE Trans Reliability 2000; 49:363–376.
16. Resilience Company. Ultra2 Solaris servers. Available at: http://www.resilience.com.
17. Latif-Shabgahi G, Bass JM, Bennett S. History-based weighted average voter: a novel software voting algorithm for fault-tolerant computer systems. Parallel and distributed processing. Proceedings of Ninth Euromicro Workshop on Parallel and Distributed Processing, Mantoua, Italy, February 7-9 2001, February 2001. New York, NY: IEEE Computer Society, 2001; 402-409.
18. Trivedi K, Dugan JB, Geist R, Smotherman M. Modeling imperfect coverage in fault-tolerant systems. Fault-tolerant computing. Highlights from Twenty-Five Years. Twenty-Fifth International Symposium. Pasadena, CA: IEEE Publication, 1995; 176.
19. Avizienis A. Toward systematic design of fault-tolerant systems. Computer 1997; 30:51–58.
20. Huang HK, Cao F, Liu BJ, Zhou MZ, Zhang J, Mogel GT. PACS simulator: a standalone educational tool. Radiology 2001; 221:688.
21. Huang HK, Cao F, Zhang J, Liu B, Tsai ML. Fault tolerant picture archiving and communication system and teleradiology design. In Reiner BI, ed: Security issues in the digital medical enterprise. Society for Computer Applications in Radiology, November 2000.
22. Huang HK, Liu BJ. Coverage by diagnostic imaging. Available at: http://www.dimag.com/webcast00/archives.shtml (Monday, November 27, 2000).
23. Huang HK, Cao F, Zhang JZ. Fault-tolerant design and implementation of the PACS controller. Radiology 2000; 217:519, 709.
24. Cao F, Huang HK, Liu BJ, Zhou MZ, Zhang J, Mogel GT. Fault-tolerant PACS server. Radiology 2001; 221:737.
25. Huang HK, Cao F, Liu BJ, Zhang J, Zhou Z, Tsai A, Mogel G. Fault-tolerant PACS server design. SPIE Medical Imaging 2001; 4323:83–92.
26. Cao F, Liu BJ, Huang HK, Zhou Z, Zhang J, Zhang X, Mogel G. Fault-tolerant PACS server. SPIE Medical Imaging 2002; 4685:316–325.
27. Huang HK, Cao F, Liu BJ, Zhou MZ, Zhang J, Mogel G. A complete continuous-availability PACS archive server solution 2002;225:692.
28. Liu BJ, Huang HK, Cao F, Documet L, Sarti DA. A fault-tolerant back-up archive using an ASP model for disaster recovery. Radiology 2001; 221:741.
29. Liu BJ, Huang HK, Cao F, Documet L, Sarti DA. A fault-tolerant back-up archive using an ASP model for disaster recovery. SPIE Medical Imaging 2002; 4685:89–95.
30. Liu BJ, Huang HK, Cao F, Documet L, Muldoon J. Clinical experiences with an ASP model backup archive for PACS images. Radiology 2002; 225:313.
31. Liu BJ, Huang HK, Cao F, Zhou M, Zhang J, Mogel G. A complete continuous-availability (CA) PACS archive server solution. Radiographics 2004, in print.

Journal of Digital Imaging

Computer-Assisted Bone Age Assessment: Graphical User Interface for Image Processing and Comparison

Ewa Pietka, PhD, DSc,[1] Arkadiusz Gertych, MS,[1] Sylwia Pospiech–Kurkowska, MS,[1]
Fei Cao, PhD,[2] H.K. Huang, DSc,[2] and Vincente Gilzanz, MD, PhD[2]

The current study is part of a project resulting in a computer-assisted analysis of a hand radiograph yielding an assessment of skeletal maturity. The image analysis is based on features selected from six regions of interest. At various stages of skeletal development different image processing problems have to be addressed. At the early stage, feature extraction is based on Lee filtering followed by the random Gibbs fields and mathematical morphology. Once the fusion starts, wavelet decomposition methods are implemented. The user interface displays the closest neighbors to each image under consideration. Results show the sensitivity of different regions to both stages of development and certain feature sensitivity within each region. At the early stage of development, the distal features are more reliable indicators, whereas at the stage of epiphyseal fusion, a larger dynamic range of middle features makes them more sensitive. In the current study, a graphical user interface has been designed and implemented for testing the image processing routines and comparing the results of quantitative image analysis with the visual interpretation of extracted regions of interest. The user interface may also serve as a teaching tool. At the later stage of the project it will be used as a classification tool.

KEY WORDS: Skeletal maturity, bone age assessment, computer-aided diagnosis, feature extraction, image analysis, graphical user interface

BONE AGE assessment is a radiological procedure performed for pediatric patients. A left-hand-wrist radiograph is acquired and compared with standard patterns. The diagnostic procedure is based on an overall comparison of the entire hand with an atlas pattern[1] or a detailed analysis of selected regions of interest.[2] Simplicity makes the first approach be used 76% of the time.[3] The disadvantage of this method is the subjective nature of the analysis performed by various observers with different levels of training. Studies have shown[4] interobserver differences ranging from 0.37 to 0.6 year, whereas the intraobserver differences range from 0.25 to 0.47 year. Atlas-matching methods have such high discrepancies because of a general comparison of the radiograph with the atlas pattern. By a more detailed comparison of individual bones, ambiguous results may be obtained.

Another technique of bone age assessment relies on a trained observer applying the Tanner and Whitehouse (TW2) method.[2] This method uses a detailed analysis of each individual bone (epiphysis – metaphysis complex of each tubular bone and carpal bones). Each complex is assigned to one of nine classes reflecting nine developmental stage[2]:

stage A: lack of calcium deposit
stage B: single deposit of calcium
stage C: center is distinct in appearances
stage D: maximum diameter is half or more the width of metaphysis
stage E: border of the epiphysis is concave

[1]From the Division of Biomedical Electronics, Institute of Electronics, Silesian University of Technology, Gliwice, Poland.

[2]From the Department of Radiology, Children's Hospital Los Angeles, Los Angeles, CA, 90027, USA.

Correspondence to: Ewa Pietka, PhD, DSc, tel: +48 609 855 512; e-mail: pietka@polsl.gliwice.pl

Online publication 13 July 2004
doi: 10.1007/s10278-004-1006-6

stage F: epiphysis is as wide as the metaphysis
stage G: epiphysis caps the metaphysis
stage H: fusion of epiphysis and metaphysis has begun
stage I: epiphyseal fusion completed

Assignment of each bony structure to a certain developmental stage is described in terms of scores. The sum of all scores reflects the overall stage of the skeletal development.

Several approaches have been made to the problem of computerized assistance in image diagnosis. A computer-assisted Tanner–Whitehouse (TW) method requires[5] an operator to position each epiphysis beneath a video camera, view the image on the computer screen, and match the image to the template of the developmental stages displayed on the screen. A Fourier transform yields coefficients that are compared with those generated by each stage of the TW standards and the closest match is sought. The computerized TW method has been evaluated in assessment of bone age[6] well as patients with Turner syndrome and constitutionally tall stature.[7]

Studies on certain hand regions have also been performed. Duryea et al[8] have developed an algorithm for the extraction of joint space and location of phalanx edges. Their approach[9] uses a cross-correlation as a measure of similarity between an analyzed region and a prototype. Another approach[10] performs the analysis of the middle region of the III phalanx. The method is based on a point distribution model derived by several points marked manually on the bones. This model is then compared with pattern models describing each developmental stage. A segmentation method based on the knowledge of anatomical structures has been suggested.[11]

This study is part of a project whose overall goal is to develop a computer-aided methodology able to assist the radiologist in assessment of skeletal maturity. The goal of this study is to develop a graphical user interface (GUI) able to view all stages of the image analysis. The GUI is used for testing purposes in order to analyze the reason for failures of certain image processing procedures and in clinical evaluation. The GUI is also applied in teaching and training programs.

DATA

The analysis was performed on left-hand-wrist radiograms selected from a normal population and organized into four blocks: African–American male and female and European male and female. Based on preliminary results,[12] for prepubertal children (0 – 9 years old) 5 images for each age group were collected, whereas for children in puberty (10 – 18 years old) 10 images for each age group were collected. This yielded 135 images per block and a total of 540 images. These images were acquired at the USC Children's Hospital in Los Angeles.

A separate group of 1,000 clinical images (normal and pathological) was collected for evaluation. These images were acquired at the University of California, San Francisco. A Fuji FCR-9000 computed radiography (CR) system was used to acquire full-resolution CR images (2 kb × 2 kb × 10 bits) in a DICOM 3.0 standard. If analog technology was used, films were digitized with a Lumiscan 200 digitizer to 2 kb × 2 kb × 12 bits.

All images were stored in a Web-based environment database.[13,14] It included the hand images and their relevant patient data. The system architecture was separated into two parts: One served as a data collection part and the other was a hand atlas server. Reference and clinical hand image were stored on an Oracle Image Database Server.

IMAGE PREPROCESSING PROCEDURES

Graphical user interface views three stages of the image analysis. The first stage performs the image standardization including the background suppression and removal of markers, labels, and noisy objects. Then, features sensitive to a certain stage of development are extracted. Finally, images from the reference database, whose features fulfill the closest match, are selected and corresponding regions of interest are displayed.

Background Removal

Background extraction[15] is based on a joint analysis of a set of histograms. The tallest peak of a global histogram reflecting the soft and bony structure is mapped onto four local histograms of four image quarters. Locations of the tallest local peaks correspond to the average intensity of the background level. Because of background nonuniformity, these levels have to be found independently. They serve as a basis for the overall background estimation found by the means of a bilinear interpolation.

After the background removal a global histogram features a narrow peak reflecting the hand area. A triangle algorithm points to the threshold value which permits the hand to be extracted. Morphological operators remove isolated points.

Extraction of the Epiphyseal Region of Interest

After the hand area has been extracted, the epiphyseal regions of interest (EROI) of the II, III, and IV phalanges are located (Fig. 1). Along phalangeal axes first-order derivative vectors are searched for three local maxima reflecting the gaps between epiphyses and metaphyses. An insensitivity sector which covers the entire area surrounding the gap prevents two neighbor maxima from being selected within the same epiphyseal region.[15]

Once the gaps are located, the EROI edges are defined. The width is estimated by the distance between the phalangeal edges and the height is set on the basis of the neighbor gap location.

GRAPHICAL USER INTERFACE

Because of the changes in bony structures, skeletal development is divided into two stages. In the first stage the epiphyses change in size filling up the gap between epiphyses and metaphyses. Once the mature size is reached, both bones start to fuse. The process starts from the middle and expands toward the edges.

Features describing the developmental stage differ from one stage to the other. As long as the epiphyseal size is changing, the diameters can be used as sensitive parameters. Once the size does not differ, features measuring the stage of fusion have to be extracted. Since the nature of the features differs significantly, the extraction techniques vary as well. In order to view the image function performed in both developmental stages, the GUI is designed independently for both phases.

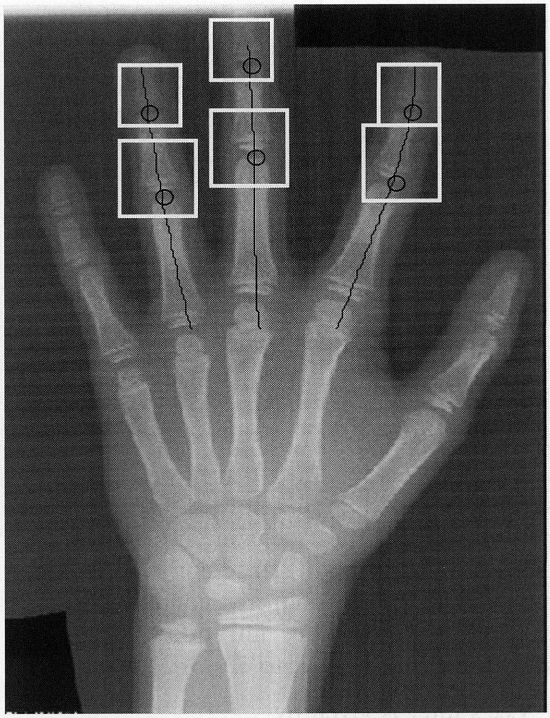


Fig 1. Epiphyseal regions of interest of the II, III, and IV phalanges.

Early Stage of Skeletal Development

The early stage of skeletal maturity requires a detailed analysis of the size and shape of epiphyses and metaphyses. The flowchart of the EROI processing performed by the GUI is shown in (Figure 2). The main menu of the GUI is divided into two sections (Fig. 3). First, the filtration permits testing various filters and viewing the processed region next to each other for comparison. The second section turns on the segmentation and features extraction stages. After the region is viewed by using the **Load Image** button, various filtering techniques may be tested.

Filtration

The median filter replaces each pixel by the median of its neighborhoods. All eight neighborhoods may be chosen or a star-shaped area may be tested. This is defined as:

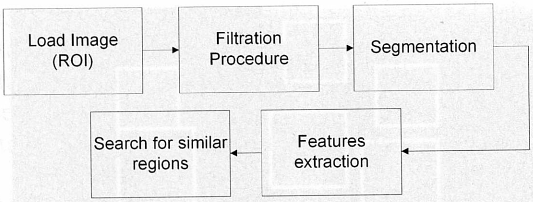


Fig 2. Flowchart of EMROI processing.

$$y(i,j) = \text{median}|x(i,j), x(i,j+1), x(i,j-1), x(i-1,j), x(i+1,j)|, \quad (1)$$

where i, j are pixel coordinates, x is the input image, and y is the output image. The function suppresses the impulse noise ("salt and pepper") and smoothes the nonuniformity within the soft tissue and bony structure. At this step, a median filter or star-shaped median filter can be tested followed by Lee filtration.[16] The mask size of each filter is set to 3 × 3.

Lee's filter is given by

$$p(i,j) = m_f(i,j) + \frac{\sigma_f^2(i,j)}{\sigma_f^2(i,j) + \sigma_v^2}[g(i,j) - m_f(i,j)], \quad (2)$$

where $p(i,j)$ is a processed image (output image), $g(i,j)$ is a corrupted image (input image), $m_f(i,j)$ and $\sigma_f(i,j)$ are local mean and standard deviation of original image, σ_v^2 is a noise variance, and i,j are pixel coordinates.

The filtration procedure requires the noise variance σ_v^2 to be known. If the noise variance is not given, the algorithm uses the average of all local estimated variances. Other parameters such as $m_f(i,j)$ and $\sigma_f(i,j)$ are estimated from the input image and delimited by a 3 × 3 mask. A larger mask smoothes the edges. This is particularly undesirable in images of optimally chosen exposure parameters yielding sharp variation edges.

Segmentation and Features Extraction

Both filtered and unfiltered (original) regions can be subjected to further analysis. Its goal is to separate the bony structure from soft tissue. The segmentation procedure consists of two steps: (1) a preliminary clustering step applying a c-means algorithm and (2) segmentation with Gibbs random fields and estimation of intensity function.

A c-means clustering algorithm[17] uses a similarity measure based on the Euclidean distance between samples and cluster prototypes. Each kth sample (pixel) has three components: image coordinates and gray level. The samples are grouped into a predefined number of classes. A criterion of partitioning is based on the minimization of the relative difference between prototypes in two iterations.

The goal of the second step of the segmentation procedure is to assign the image pixels to a certain class on the basis of their intensity and relative location. In our model, a pixel y of the image is at a location s and its intensity (gray level) is y_s. A segmentation (subdivision) of the image into classes will be denoted by x, where $x_s = i$ means that a pixel belongs to the class i. In our approach the number of classes is set to 2. Each class can be described by a slowly varying intensity function plus white Gaussian noise having a $\sigma_n{}^2$ noise variance. In order to make use of the spatial information, a Gibbs random field describing local distribution of gray levels is used. Given the intensity function, a probability density function having two components can be defined. One constrains the region intensity to be close to the data, the other imposes spatial continuity. The algorithm maximizes the probability density function and estimates the intensity functions alternately. As the algorithm progresses, the intensities are updated by averaging gray level values of pixels belonging to the considered classes over the sliding window whose size progressively decreases. Thus, the algorithm starts with global estimates and slowly adapts to the local characteristics of each class[18,19] (Fig. 4a). In our model only one-point and two-point cliques are applied (Fig. 4b). The noise parameter is calculated by a distinct procedure however, the user is able to set its value manually (Fig. 3).

The c-means clustering algorithm initially estimates x. Beginning with segmentation (Fig. 4a), the algorithm estimates intensity functions $\mu_s^{(i)}$ and x interchangeably. Each iteration consists of one update of x and one up-

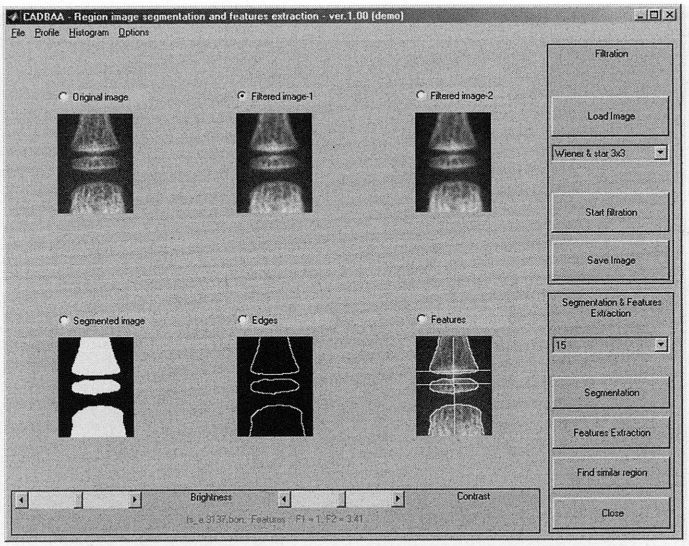


Fig 3. Program main window.

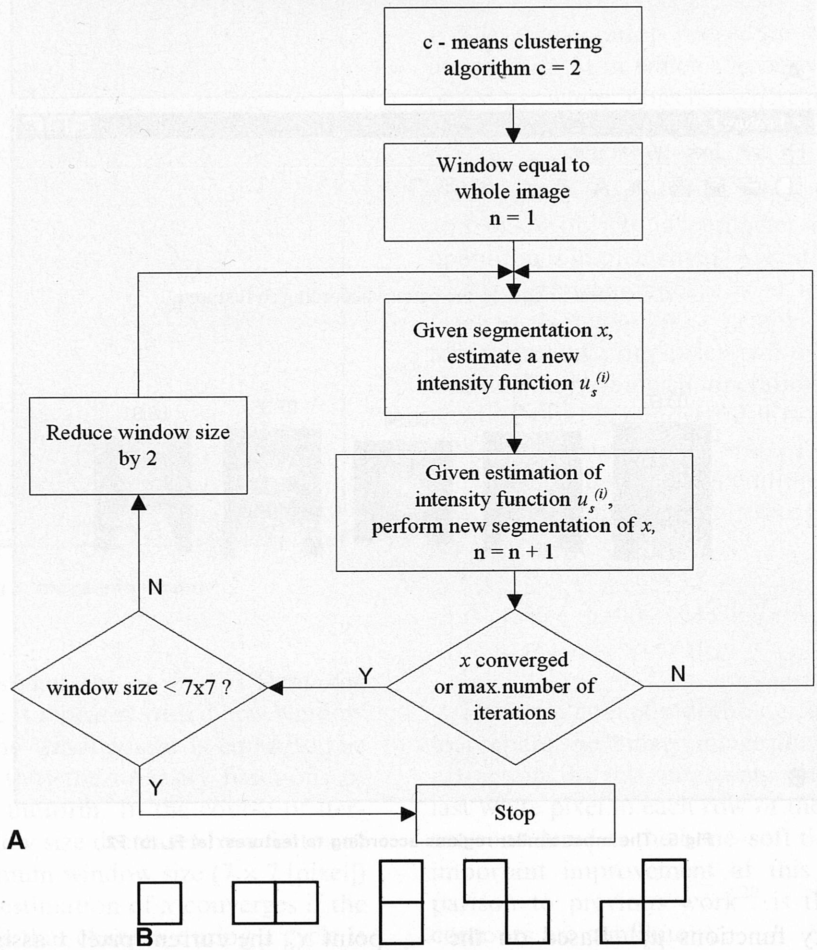


Fig 4. Segmentation procedure: (a) the flowchart, (b) one-point and two-point cliques.

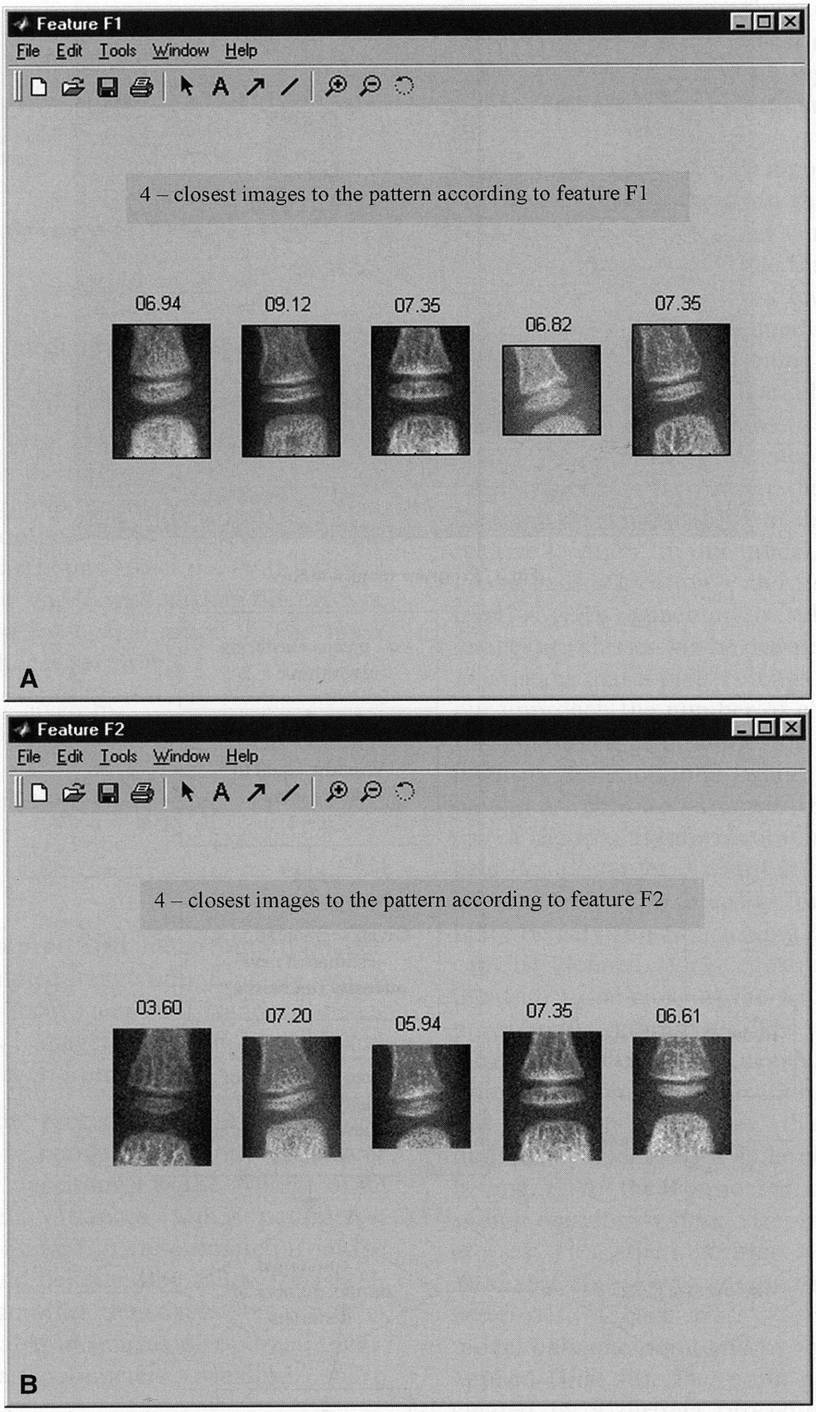


Fig 5. The most similar regions according to features: (a) Fl, (b) F2.

date of intensity functions $\mu_s^{(i)}$. Based on the estimation of intensity functions $\mu_s^{(i)}$ and maximization of the conditional density at each point x_s, the current pixel x assignment to one of the classes is updated. Until the procedure converges, the window size W for calculating

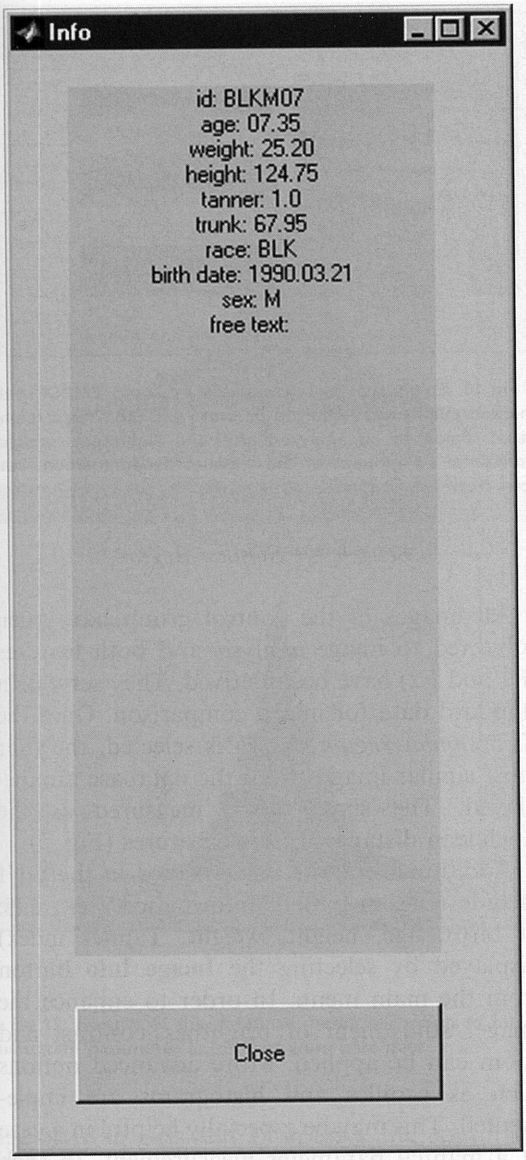


Fig 6. Image info window.

the intensity estimations is fixed. Then the whole procedure is repeated with a new window size. Initially, the window size is equal to the image size and thus the intensity functions of each region are uniform. In the course of iterations, the window size decreases by a factor of 2 until the minimum window size (7×7 [pixel]) is reached. The estimation of x converges if the number of pixels that changes during a cycle is less than a threshold, typically $M/10$ (where M is the smallest image dimension).

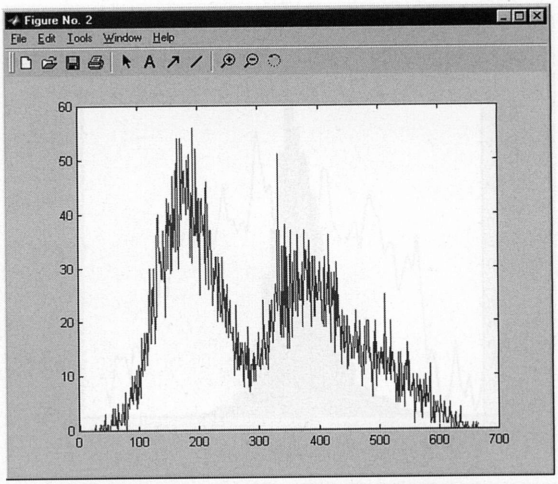


Fig 7. Sample histogram of original image.

Bone Edge and Feature Extraction

The segmentation procedure yields a binary image (Fig. 3) in which the bony structure appears as white objects, whereas other areas (background and soft tissue) turn to black. This binary image is subjected to a further image analysis. In order to draw edges between white and black objects, a perimeter morphological operation is implemented. A white pixel belongs to the perimeter if there is at least one black pixel in its eight-closest neighborhoods. Next, for removing one-pixel white objects, the cleaning morphological operation is performed. A white pixel is removed if it is surrounded by eight black pixels.

Two morphological operations, a perimeter PM and a cleaning pixel operate CP, are defined as:

$$PM(i,j) = X \cap (\overline{X_0} \cup \overline{X_1} \cup \ldots \cup \overline{X_7})$$
$$CP(i,j) = X \cap (X_0 \cup X_1 \cup \ldots \cup X_7)$$

The main goal of morphological operations is to prepare the binary image for the main axis extraction. In this step, only the first and the last white pixel in each row of the binary image are considered as the bone–soft tissue edge. The important improvement at this step in comparison to previous work[20] is that the object contours are continuous.

Once the *Features extraction* option from the menu (Fig. 3) is selected, three steps are laun-

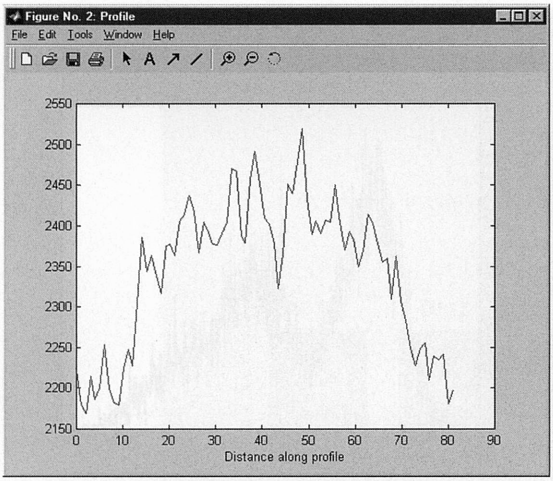


Fig 8. Distance along profile.

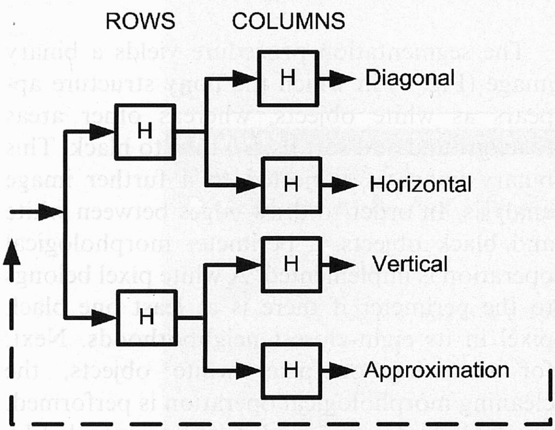


Fig 9. Wavelet decomposition algorithm scheme (first iteration).

ched. First, the main axis of each phalanx (metaphyses) is found. Then, the image is scanned perpendicular to the main axis. At this point the transverse dimensions of the metaphysis and epiphysis are located. If the segmented image contains three distinct objects (area of metaphysis, epiphysis, and diaphysis), immediately the extraction of epiphyses is proven. Finally, the epiphyseal vertical diameter is found along the main axis. Based on these findings, two features are extracted: the ratio of both transverse dimensions and the ratio of both epiphyseal dimensions marked F1 and F2, respectively.

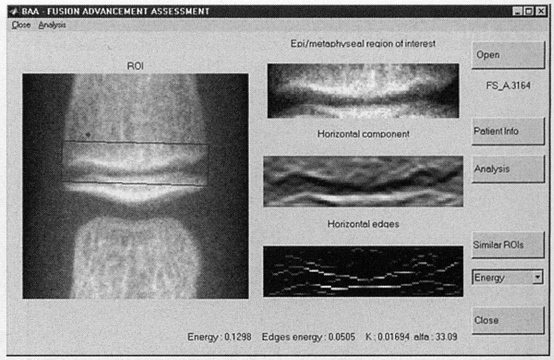


Fig 10. Epiphyseal fusion assessment. Left: EMROI with marked epiphyseal region to be analyzed. Right top: epiphyseal region to be analyzed after the alignment; middle: horizontal component of the wavelet decomposition, bottom: horizontal edges.

Search for Similar Region

All images of the control group have been subjected to image analysis and both features (F1 and F2) have been derived. They serve as a standard data for image comparison. Once the *Find similar region* option is selected, the four most similar images from the database are displayed. The similarity is measured as the Euclidean distance between features (Fig. 5).

Additional options implemented in the GUI include (Fig. 6) patient information (sex, date of birth, age, height, weight, Tanner index) displayed by selecting the **Image Info** button from the main menu. In order to enhance the image, adjustment of brightness/contrast and zoom can be applied. More advanced options such as profiles and histograms are implemented. This may be especially helpful in a case of a manual parameter measurement. In each case a new window appears and a histogram (Fig. 7) or gray level values along the chosen profile (Fig. 8) are drawn. At any time the user may save the image for further analysis by clicking the **Save image** button or selecting *Save* from the *File* menu (Fig. 3).

Later Stage of Skeletal Development

Methodology of Feature Development

In our study the late stage of skeletal development is the stage in which the epiphyseal fusion progresses. It starts approximately at the

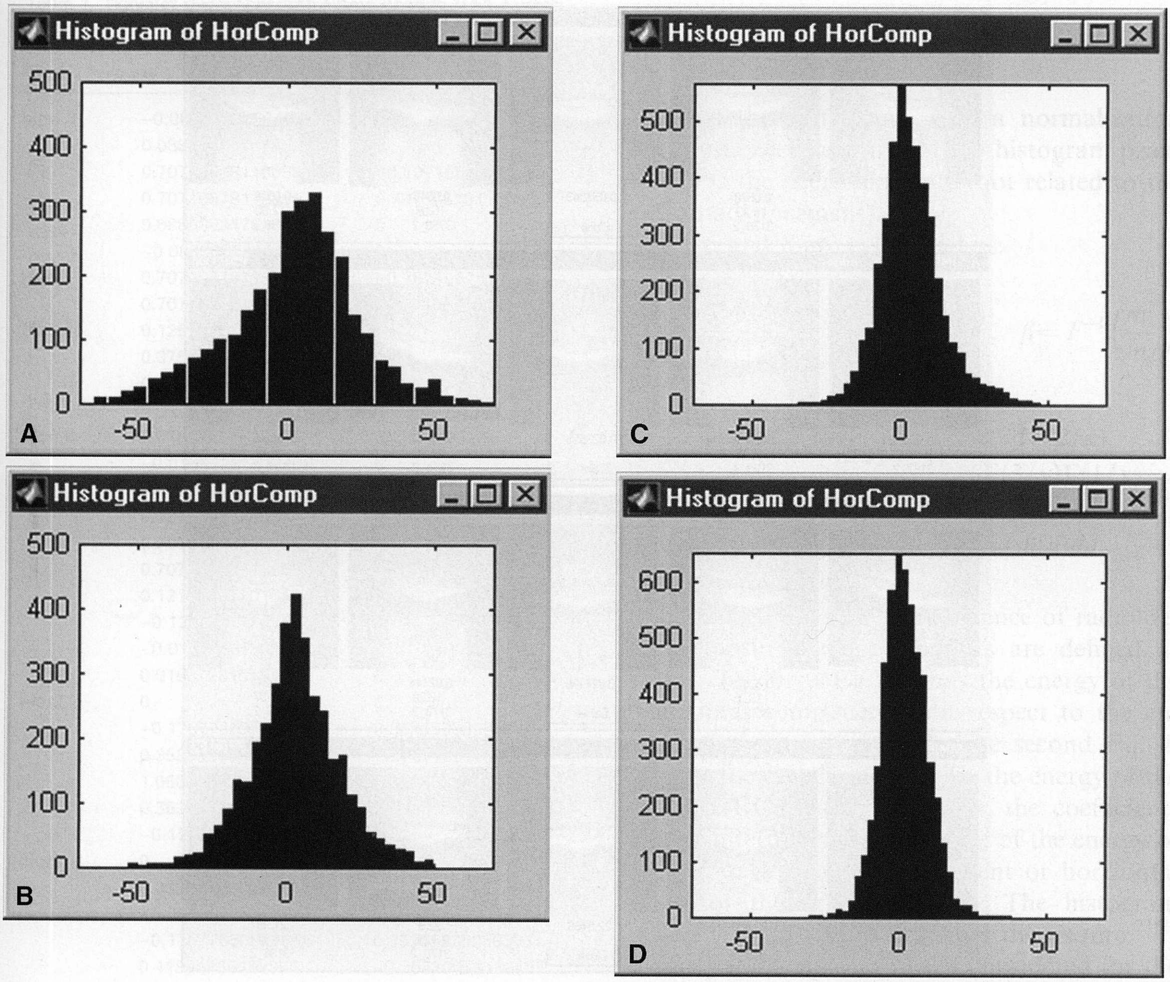


Fig 11. Results of the database match according to (a) energy of horizontal component; (b) energy of horizontal edges; (c) histogram parameter *K*; (d) histogram parameter α.

age of 10 in girls and 12 in boys and continues until the epiphysis and metaphysis become one adult bone. Features measuring the size of bones are no longer pertinent at this developmental stage. The assessment of the fusion stage is performed by means of the wavelet decomposition. In our study, the most important property of the 2-dimensional separable wavelet transform is its ability to separate vertical, diagonal, and horizontal structures of an image. The separation is achieved by convolving the image with low- and high-pass wavelet filters along the rows and columns.[21] The decomposition yields an approximation component and three detailed components: vertical, horizontal, diagonal (Fig. 9).

It can be noticed that the advancement of epiphyseal fusion suppresses horizontal edges of bones. At this stage features reflecting the presence of edges have to be defined and extracted. The analysis is performed on the horizontal component of the wavelet decomposition. Unlike in the standard decomposition scheme, in the current study the image is not subsampled after each level.

An analysis of the epiphyseal – metaphyseal ROI (EMROI) begins with an alignment approval. If the angle of the phalangeal axis and the vertical direction exceeds 5°, a rotation procedure is applied. Quantitative analysis requires a precise definition of the region: It is limited to the upper part of the epiphysis and metaphysis

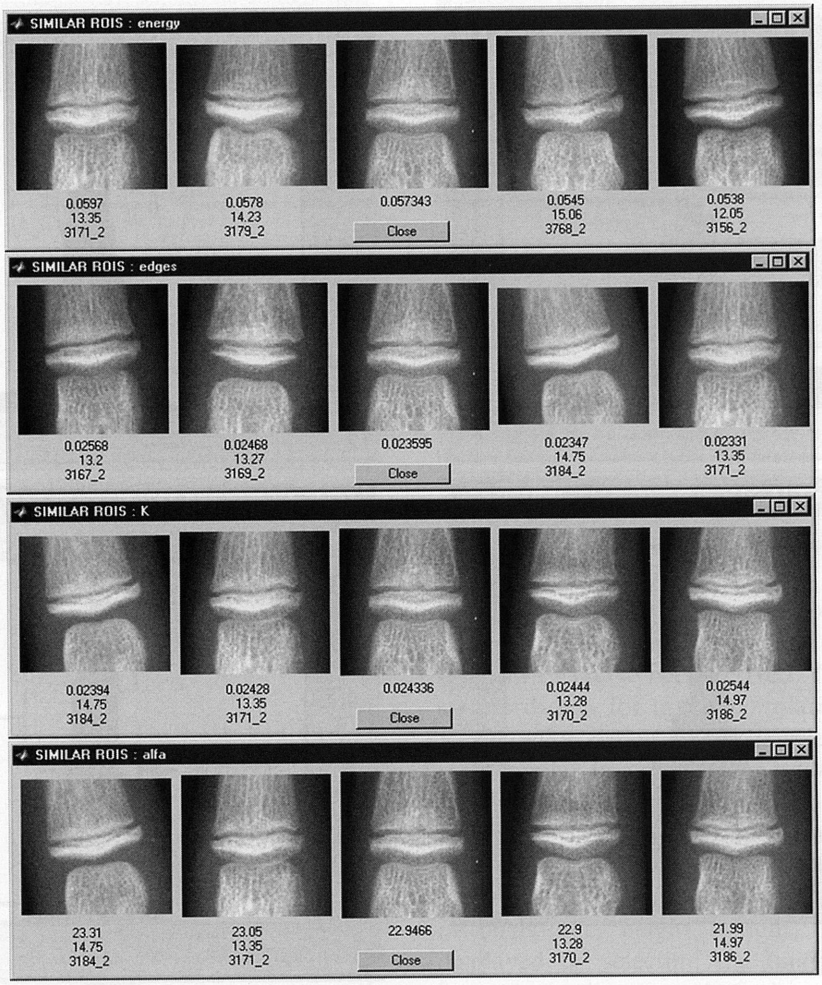


Fig 12. Histogram of horizontal component at different stages of fusion: from epiphysis and metaphysis not fused to a complete fusion.

(Fig. 10). The already marked transverse dimensions delimit the region. Next, the decomposition algorithm is applied. It is based on the bioorthogonal, Haar, and Mallat wavelet filters.[22] Coefficients of original filters L_1 and H_1 are shown in Table 1. Because of the size of the regions subjected to the wavelet transform (an average of 30–40 rows), using short high-pass filters is an important advantage.

Features applied to the estimation of the stage of fusion are calculated for the horizontal detail component at the second level of decomposition.[23] In order to extract the horizontal edges, local maxima of the horizontal component along columns are found. Then, a thresholding procedure is applied. The threshold value is estimated on the basis of the standard deviation of the combination of the horizontal and vertical components. This procedure enhances the main vertical edges.

Two coefficients sensitive to the developmental process have been chosen.[24] They are derived from the energy of horizontal details:

$$C1 = E(H)/E(\mathrm{ROI}) \quad (3)$$

and the energy of horizontal edges, defined as

$$C2 = E(Y)/E(\mathrm{ROI}) \quad (4)$$

Parameters K and α of a histogram model also reflect the advancement of the epiphyseal fusion:

Table 1. Wavelet decomposition filters used in BAA–Fusion Advancement Unit.

Filter[a]	L_1	H_1
biorl.3	−0.08838834764832	0
	0.08838834764832	0
	0.70710678118655	−0.70710678118655
	0.70710678118655	0.70710678118655
	0.08838834764832	0
	−0.08838834764832	0
Haar	0.70710678118655	−0.70710678118655
	0.70710678118655	0.70710678118655
Mallat	0.12500	0
	0.37500	−2.0
	0.37500	2.0
	012500	0
biorl.5	0.01657281518406	0
	−0.01657281518406	0
	−0.12153397801644	0
	0.12153397801644	0
	0.70710678118655	−0.70710678118655
	0.70710678118655	0.70710678118655
	0.12153397801644	0
	−0.12153397801644	0
	−0.01657281518406	0
	0.01657281518406	0
bior2.2	0	0
	−0.17677669529664	0.35355339059327
	0.35355339059327	−0.70710678118655
	1.06066017177982	0.35355339059327
	0.35355339059327	0
	−0.17677669529664	0
bior2.4	0	0
	0.03314563036812	0
	−0.06629126073624	0
	−0.17677669529664	0.35355339059327
	0.41984465132951	−0.70710678118655
	0.99436891104358	0.35355339059327
	0.41984465132951	0
	−0.17677669529664	0
	−0.06629126073624	0
	0.03314563036812	0
db2	−0.12940952255092	−0.48296291314469
	0.22414386804186	0.83651630373747
	0.83651630373747	−0.22414386804186
	0.48296291314469	−0.12940952255092

[a] Filters: Biorl.3, Haar and Mallat are implemented for fusion assessment, the others are implemented for comparison.

$$h(u) = Ke^{-(|u|/\alpha)^{\beta}} \quad (5)$$

where H is the horizontal component at the second level of decomposition, Y is the horizontal edge at the second level of decomposition, ROI is the epiphyseal region of interest, E is the energy defined as

$$E(X) = \sum_{i,j} x_{i,j}^2$$

$h(u)$is the histogram, K is a normalization constant, α is the width of the histogram peak, and β is the decreasing rate (not related to the fusion advancement).

K, α, and β are calculated as:

$$K = \frac{\beta}{2\alpha\Gamma(1/\beta)}, \quad \alpha = m_1\frac{\Gamma(1/\beta)}{\Gamma(2/\beta)}, \quad \beta = F^{-1}\left(\frac{m_1^2}{m_2}\right)$$

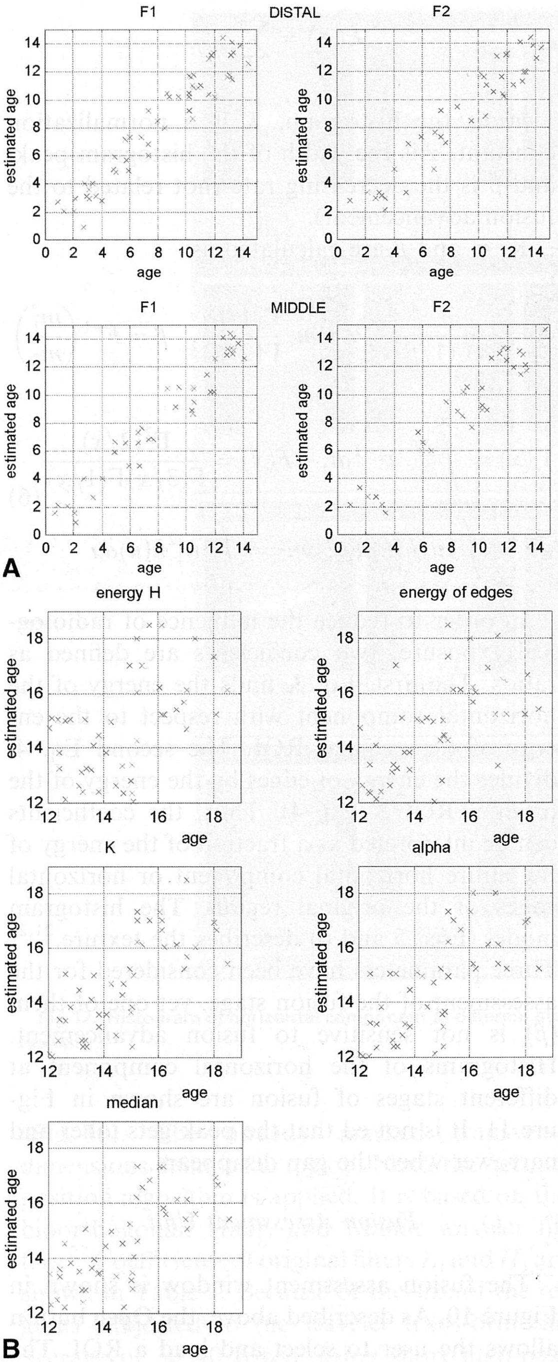
$$\Gamma(x) = \int_0^{\infty} e^{-t}t^{x-1}dt, \quad F(x) = \frac{\Gamma^2(2/x)}{\Gamma(3/x)\Gamma(1/x)} \quad (6)$$

$$m_1 = \int |u|h(u)du, \quad m_2 = \int |u|^2h(u)du$$

In order to reduce the influence of radiological exposure, two coefficients are defined as ratios. The first, Eq. 3, finds the energy of the horizontal component with respect to the energy of the original ROI. The second Eq. 4, divides the energy of edges by the energy of the original ROI 2 (Eq. 4). Thus, the coefficients can be interpreted as a fraction of the energy of the entire horizontal component or horizontal edges of the original region. The histogram model (Eqs, 5 and 6) describes the texture.[21,25] Three parameters have been considered for the assessment of the fusion stage, yet one of them (β) is not sensitive to fusion advancement. Histograms of the horizontal component at different stages of fusion are shown in Figure 11. It is not ed that the peak gets taller and narrower when the gap disappears.

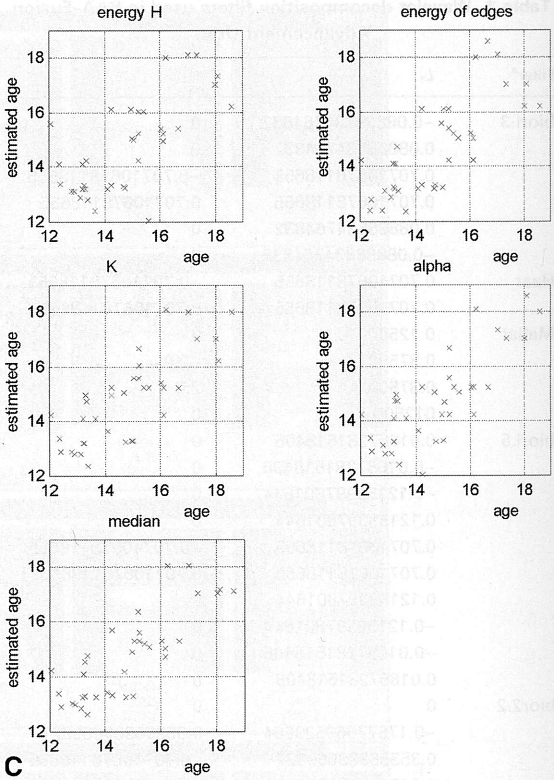
Fusion Assessment Unit

The fusion assessment window is shown in Figure 10. As described above, the **Open** button allows the user to select and load a ROI. The **Analysis** button invokes the wavelet decomposition and feature extraction. This results in the following images being displayed: aligned region to be analyzed, horizontal component, and horizontal edges. Values of features are listed below. Three types of wavelets filters (biorthogonal, Haar, Mallat) at different decomposition levels can be tested by changing (in the



A

B



C

Fig 13. Distribution of the measurement distance between a selected image and its closest neighbor for (a) distal features and (b) middle features.

menu) the default settings. After the analysis is completed the user can select one of the features and search for similar regions in the database. On default four regions are displayed, two on both sides. The feature value, age of a child, and ID of the displayed image are listed below each of the reference regions (Fig. 12). The user is able to compare the analyzed region with the patterns and assign it to the bone age. Results of the database search according to each feature are displayed simultaneously.

Since the bioorthogonal spline wavelet filters (bior1.3 in Table 1) at the second level of decomposition appear to feature the best discriminatory power in the assessment of fusion,[22] they are set on defaults. One can also perform the analysis for any level from 1 to 3 and for other wavelet filters (as shown in Table 1). However, when the default setting is changed a **similar ROIs** search becomes inactive as comparison of parameters calculated for different levels or filters makes no sense. The **Patient Info** button displays the information from the ROI file header. After displaying a new ROI, **Patient Info** and **Analysis** buttons are active and **Feature Selection** and **Similar ROIs** buttons are inactive.

RESULTS AND CONCLUSION

The GUI has been developed as a tool for the comparison of quantitative image analysis with the appearance of changes in the image. The analysis has been performed for one ethnic group and one gender. Two steps of the evaluation have been performed. First, a quantitative analysis of features has shown the accuracy of the assessment of the image similarity. Then, the four closest neighbor images have been selected and displayed.

In the current study images from the normally developed population have been considered and the similarity of features has served as a basis for the image comparison. Therefore, the chronological age corresponds to the skeletal age. For each image the features have been found. Then, the image of the closest feature has been selected from the database. By selecting the closest match, plots (Fig. 13) show the deviation of the skeletal age within the standard database. The distribution varies for different features. A percentage of images with 3 standard deviations serves as an indicator in the analysis of feature sensitivity. For the younger population, the ratio of horizontal diameters (F1) matches images with a better similarity (Fig. 13a) than the ratio of epiphyseal diameters in vertical and horizontal diameters (F2). The distal regions show a higher discriminatory power than the middle regions. For the older population (Fig. 13b, c), features extracted from the middle regions are better indicators for image matching than the corresponding features of the distal regions.

As far as the visual comparison is concerned, there is no significant difference between the analyzed image and two closest neighbors, yet there is a visible difference if four closest neighbors are displayed.

Results have shown various sensitivity of features extracted from both regions. The distal ROIs yield better results in the earlier stage of development. A closer match is obtained and the mean distance between an image and its closest neighbor is smaller for distal features. It differs at the latter stage of development. Features extracted from the middle regions show a better sensitivity in differentiating various stages of development. These conclusions are in accordance with the medical statement; "The more distal, the better"[26]. The distal areas feature a higher sensitivity to the developmental changes and reach the mature size and shape earlier than middle regions. Thus, after a certain stage of development, the middle regions show a larger dynamic range than the more mature distal areas.

The user interface is implemented for testing and teaching purposes well. It permits certain stages of the image analysis to be visualized. The influence of processing parameters can be examined and their adjustment improves the performance of the overall analysis.

Finally, the GUI will be applied at the classification stage when clinical images will be subjected to bone age assessment. In the final clinical implementation of the entire computer-aided system, it will be used as an assistance procedure in performing radiological diagnosis. Access to the closest image and its display will improve the objectivity and speed up the radiological procedure.

ACKNOWLEDGMENT

This work was supported in part by KBN 7711C 03320 and in part by NIH Grant No.ROI-LM06270.

REFERENCES

1. Greulich, WW, Pyle, SI: Radiographic Atlas of Skeletal Development of Hand Wrist 2nd ed. Stanford, CA: Standford University Press, 1971

2. Tanner, JM, Whitehouse, RH: Assessment of Skeletal Maturity and Prediction of Adult Height (TW2 Method). London: Academic Press, 1975

3. Miler, GR, Levick, RK, Kay, R: Assessment of bone age: A comparison of the Greulich and Pyle and the Tanner and Whitehouse methods. Clin Radiol 37:119-121, 1986

4. Roch, AF, Rohmann, CG, Davila, GH: Effect of training on replicability of assessment of skeletal maturity (Greulich–Pyle). AJR Am J Roentgenol, 108:511-515, 1970

5. Tanner, JM, Gibbons, RD: Automatic bone age measurement using computerized image analysis. J Pediator Endocrinol 7:141-145, 1994

6. Albanese, A, Hall, C, Stanhope, R: The use of a computerized method of bone age assessment in clinical practice. Horm Res 3:2-7, 1995

7. van Teunenbroek, A, de Waal, W, Rock, A, et al: Computer-aided skeletal age scores in healthy children, girls with Turner syndrome, and in children with constitutionally tall stature. Pediatr Res 39:360-367, 1996

8. Duryea, J, Jiang, Y, Countryman, P, Genant, HK: Automated algorithm for the identification of joint space and phalanx margin location on digital hand radiographs. Med, Phys 26:453-461, 1999

9. Dickhaus H, Habich R, Wastl S, et al: A PC-based system for bone age assessment. In proceedings of the EMBEC '99, Vienna, Medical Imaging IV, 1999, pp 1008–1009

10. Al-Taani AT, Ricketts IW, Cairns AY: Classification of Hand Bones for Bone Age Assessment. In Proceedings of the 3rd IEEE International Conference on Electronics, Circuits and Systems (ICECS '96), October 1996, Greece, Vol 2, pp 1088–1091 (ISBN 0-7803-3650-X (softbound) and ISBN 0-7803-3651-8 (microfiche)

11. Efford, ND: Knowledge-based segmentation and feature analysis of hand and wrist radiographs. proc SPIE 1905:596-608, 1993

12. Pietka, E, Huang, HK: Image processing techniques in bone age assessment In: Leondes, CT ed. Image Processing Techniques and Applications Gordon & Breach Publishers, Inc., London, 1997, pp 221-272

13. Cao, F, Huang, HK, Pietka, E, et al: Design and implementation of a digital hand atlas for Web-based bone age assessment. Radiology 217(P):685, 2000

14. Cao, F, Huang, HK, Pietka, E, et al: A digital hand atlas for Web-based bone age assessment: System design and implementation. SPIE Med Imaging 3976:297-307, 2000

15. Pietka, E, Poépiech, S, Gertych, A, et al: Computerized approach to the extraction of epiphyseal regions in hand radiographs. J Digital Imaging 14(14):165-172, 2001

16. Lim, JS: Two Dimensional Signal & Image processing Prentice-Hall, Englewood Cliffs, UJ, 1990, pp 536-540 Image Restoration

17. Banks, S: Signal Processing, Image Processing and Pattern Recognition Englewood Cliffs, NJ: Prentice-Hall, 1990

18. Pappas, T: An adaptive segmentation algorithm for image segmentation. IEEE Trans, Signal Proc. 40(4):901-914, 1992

19. Saber, E, Murat Tekalp, A, Bozdagi, G: Fusion of color and edge information for improved segmentation and edge linking. Image Vision Comput 15:769-780, 1997

20. Pietka, E, Gertych, A, Poépiech, S, et al: Computer assisted bone age assessment: Image preprocessing and ROI extraction, IEEE Trans. Med. Imaging 20(8):715-729, 2001

21. Mallat, SG: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Patt Anal Mach. Int 11(7):674-693, 1989

22. Mallat, S, Zhong, S: Characterization of signals from multiscale edges. IEEE Trans-Patt, Anal, Mach, Int. 14(7):710-732, 1992

23. Poépiech S: Assessment of epiphyseal fusion by means of wavelet transform. Proceedings of the International Workshop on Control and Information Technology, IWCIT'99 Ostrava, 1999, pp 248–254

24. Poépiech S, Gertych A, Pietka E, et al: Wavelet decomposition based features in description of epiphyseal fusion. In: Analysis of biomedical signals and images. Proceedings of BIOSIGNAL, Brno, 2000, pp 246–248 Vutium Press

25. Wouwer, G Van de , Scheunders, P, Dyck, D Van : Statistical texture characterization from discrete wavelet representations. IEEE Trans Med, Imaging 8(4):592-598, 1999

26. Kirks, DR: Practical Pediatric Imaging. Diagnostic Radiology of Infants and Children ILittle, Brown, Boston/ Toronto, 1984, Chap. 6 pp 198-201

ELSEVIER

Computerized Medical Imaging and Graphics xx (2004) 1–12

Computerized Medical Imaging and Graphics

www.elsevier.com/locate/compmedimag

International Internet-2 performance and automatic tuning protocol for medical imaging applications

Lawrence W.C. Chan[a,*], Michael Z. Zhou[b], S.K. Hau[c], Maria Y.Y. Law[a], F.H. Tang[a], J. Documet[b]

[a]*Department of Optometry and Radiography, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China*
[b]*Department of Radiology, The University of Southern California, USA*
[c]*Information Technology Services Office, The Hong Kong Polytechnic University, Hong Kong, China*

Received 15 July 2004; revised 19 August 2004; accepted 30 September 2004

Abstract

Internet-2 is an advanced computer network, which has been widely used for medical imaging applications such as teleradiology and teleconsultation, since Internet-2 can fulfill the requirements for high-speed data transmission and short turn-around time with low operation cost once installed. However, such high performance of Internet-2 may not be retained for global access from international network peers. Considering the international Internet-2 connection between the PolyU and the IPI/USC, there exist two major factors, network looping in the US and bottleneck of the connection, raising the round-trip time and limiting the available bandwidth, respectively. The available bandwidth will be further underutilized if the TCP/IP parameters at the sending and receiving computers are not appropriately chosen. This paper proposes a repeatable and consistent protocol to automatically tune these parameters for the clinical applications.

Keywords: International Internet-2; Bandwidth; Automatic tuning protocol; PACS; DICOM compliance; Teleconsultation

1. Introduction

Internet-2 is the advanced computer network technology supported by high performance gigabit backbones and isolated from the commercial payload of the public ordinary Internet. To create the tomorrow's Internet among academia, industry and government, the University Corporation for Advanced Internet Development (UCAID), Internet-2 consortium being led by US universities working in partnership with industry and government, was established to develop and deploy advanced network applications and technologies based on the Internet-2 platform [1,3,5,16]. Over 200 research universities and laboratories in North America are connected by the advanced network of Internet-2. Abilene is one of the major high performance backbones of Internet-2, which currently supports a bandwidth of 10 Gigabit-per-second (Gbps) and has about 50 millisecond (ms) round-trip response time across the US continent. Because of its high bandwidth, low latency and low operating cost, Internet-2 provides proving ground of new technology standards and applications such as IPv6, Quality of Service (QoS), Multicast and Network Security. With the effort of its members, Internet-2 makes the difference in terms of data transmission, streaming, security, etc. in many academic and research disciplines, especially medical imaging, comparing with the public ordinary Internet.

Medical imaging applications including large-volume image transmission and tele-imaging consultation requires advanced network facilities that provide high network throughput, short response time, low operating cost and security assurance. The key features of Internet-2 can fulfill all these requirements by the fact that it takes less than 80 s to transmit a set of 40 Computed Tomography (CT) images from Los Angeles to Washington DC and facilitate more than 10 cursor movements per second for synchronized image manipulation between these two remote sites [3]. Though Internet-2 technology is ideal for medical imaging

* Corresponding author. Tel.: +852 9860 0416.
E-mail address: orlchan@polyu.edu.hk (L.W.C. Chan).

0895-6111/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compmedimag.2004.09.006

applications in the US, it is not easy to use Internet-2 globally. The international partnership is a way for the non-US research and education institutions around the world to join the Internet-2 consortium.

The Hong Kong Polytechnic University (PolyU) is one of the eight tertiary institutions in Hong Kong, which is able to use Internet-2 and has been carrying out collaborative research on its applications with other collaborating institutions in the US since October 2002. For instance, this international Internet-2 connection has been used for daily medical image exchange between the Image Processing and Informatics Laboratory (IPI) of the University of Southern California (USC) and the PolyU. At the 88th Annual Meeting of the Radiological Society of North America (RSNA), which took place at Chicago from November 28 to December 3, 2002, the PolyU was acting as an international site in the demonstration of end-to-end performance using Internet-2, yielding 1.5 Megabit-per-second (Mbps) throughput [11]. Similar performance was also demonstrated in the Internet-2 Fall 2002 Member Meeting at USC. In October 2003, a multiple end-point video-teleconference connecting the PolyU, Hospital Authority in Hong Kong, University of Hawaii and University of Pittsburgh was successfully held. It was shown by Chan et al. [1] that the end-to-end performance of international Internet-2 in medical image transmission and video streaming is better than that of the public ordinary Internet.

Applications such as teleconsultation [2, 19,20], teleconferencing with remote medical image presentation [21] and image-assisted surgery/therapy [23] require a network connection with higher bandwidth and lower latency to facilitate synchronization at remote sites. In Hong Kong, the research and development of such applications are normally supported by the Hong Kong Government and the donation from companies. The larger the financial support we get, the higher the bandwidth we may have. The bandwidth of the international connection is limited by the financial support. On the other hand, an application can only utilize a small portion of bandwidth since the TCP/IP transmission rate is bounded by the TCP/IP parameters set at the operating systems of computers involved. To increase the network throughput for the international Internet-2, our Hong Kong Internet-2 community and Web100 project [12] have been designated to derive automatic tuning protocols to optimize the TCP/IP parameters. This paper proposes a consistent and repeatable tuning protocol for international Internet-2 applications in the clinical environment. The experimental results of the non-DICOM and DICOM transmission using the automatic tuning protocol at both sending and receiving ends will be presented and compared with the results for the combinations where this protocol is used at either one end only and where this protocol is not used at both ends, showing the improvement in overall network throughput by the proposed tuning method.

2. Connectivity

2.1. Internet-2 backbones: past and present; major and regional

In 1995, MCI/Worldcom ran the first Internet-2 backbone, vBNS (*v*ery high performance *B*ackbone *N*etwork *S*ervice) with the support of National Science Foundation (NSF). Under the development by UCAID in partnership with Qwest Communication, Nortel and Cisco Systems in 1999, Abilene becomes the current major backbone of Internet-2 [6]. Most of the backbone links of Abilene have the bandwidth of 10 Gbps and there are 11 fast response core router nodes interconnect the backbone links. Most universities aggregate their connections to Internet-2 through GigaPOPs (gigabit point-of-presence) directly or through regional backbone networks, as shown in Fig. 1. GigaPOPs are regional network aggregation points being formed by Internet-2 member universities to connect to

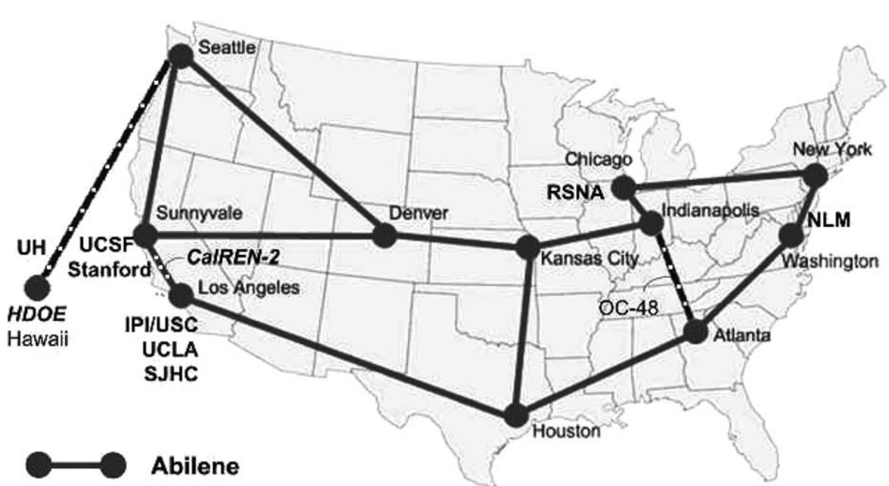


Fig. 1. Network connection map of Internet-2 backbones in the US.

a variety of high performance, and other types of networks [4]. CalREN-2 (California Research and Education Network) is one of the regional backbones, which is using regional optical network (RON) [8,10], a model of facility-based network built with owned assets supported by FiberCo. The network connection topology of CalREN-2 is illustrated in Fig. 2. The bandwidth of CalREN-2 and its uplink to Abilene are 2.4 Gbps and 622 Mbps, respectively. Hawaii State Department of Education (HDOE) is a state education network, which is sponsored by the University of Hawaii. The bandwidth of HDOE backbone is 2.4 Gbps and the speed of its connection to Abilene is 100–155 Mbps. Fig. 1 shows the Internet-2 backbones and how the Internet-2 sites are connected to the corresponding backbones. The US Internet-2 sites considered in this paper include the IPI/USC, University of California at Los Angeles (UCLA), St. John's Healthcare Center (SJHC), University of California at San Francisco (UCSF), Virtual Lab of Stanford University (Stanford), University of Hawaii (UH), National Library of Medicine (NLM) and RSNA. In Fig. 1, the backbone between the GigaPOPs at Sunnyvale and Los Angeles is CalREN-2 and the connection between Hawaii and Seattle is the uplink from HDOE to Abilene. It is also found that most of the carriers of Abilene have been upgraded to OC-192, i.e. 10 Gbps, except the OC-48 connection between Indianapolis and Atlanta.

2.2. International Internet-2 connection

As mentioned in Section 1, the non-US national research and education networks (NRENs) can peer with Internet-2 through international partnership, which enable and prompt collaboration between the US researchers, faculty, students and their overseas counterparts. Internet-2 has partnered with over 40 peer-level international organizations and networks. The Hong Kong Academic and Research Network (HARNET) is one of the peer-level networks. Since first established in 1985, HARNET has been connecting eight tertiary institutions, including the PolyU, in Hong Kong. The Joint University Computer Centre (JUCC), which is formed by these institutions, is responsible for managing and maintaining the HARNET. In September 2000, the JUCC signed a Memorandum of Understanding (MoU) with the UCAID. Based on this MoU, the tertiary institutions in Hong Kong can use Internet-2 for academic and research purposes through advanced applications in a global scale.

The peering between the Internet-2 backbones and the NRENs is done via the international interconnection points. Fig. 3 shows the international interconnection points at Abilene backbone and their corresponding NRENs. StarLight is one of the international interconnection points peering HARNET with Internet-2 ('StarLight' and

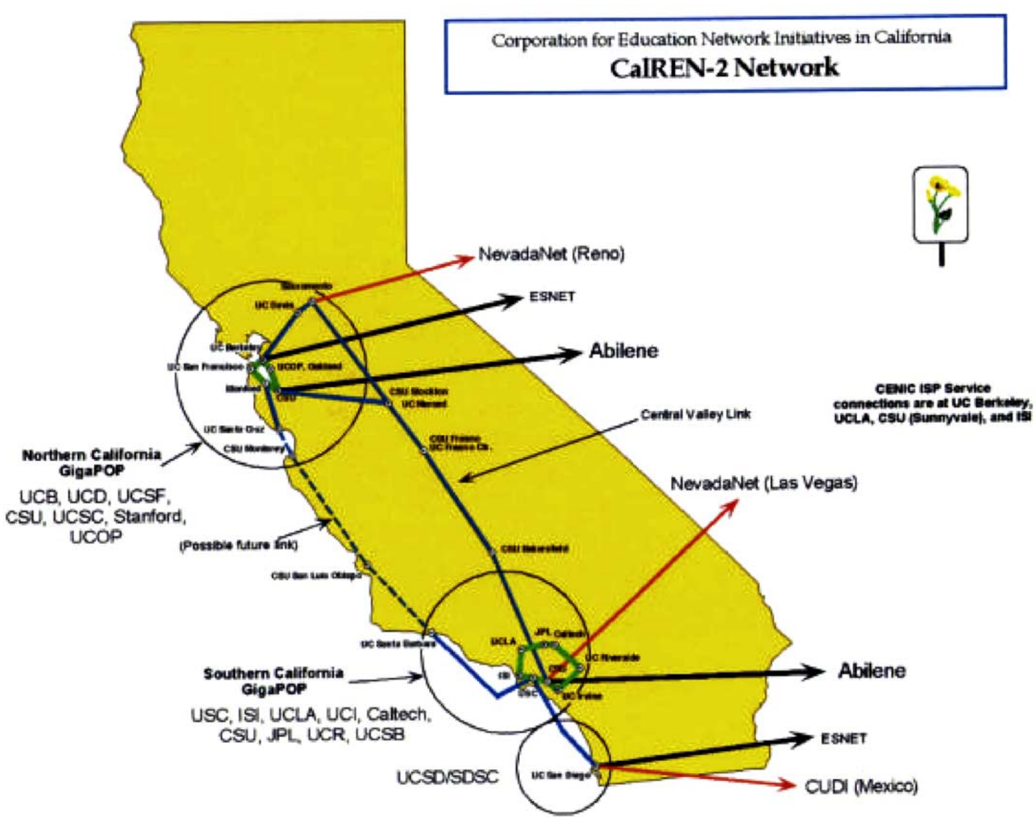


Fig. 2. Network connection topology of CalREN-2.

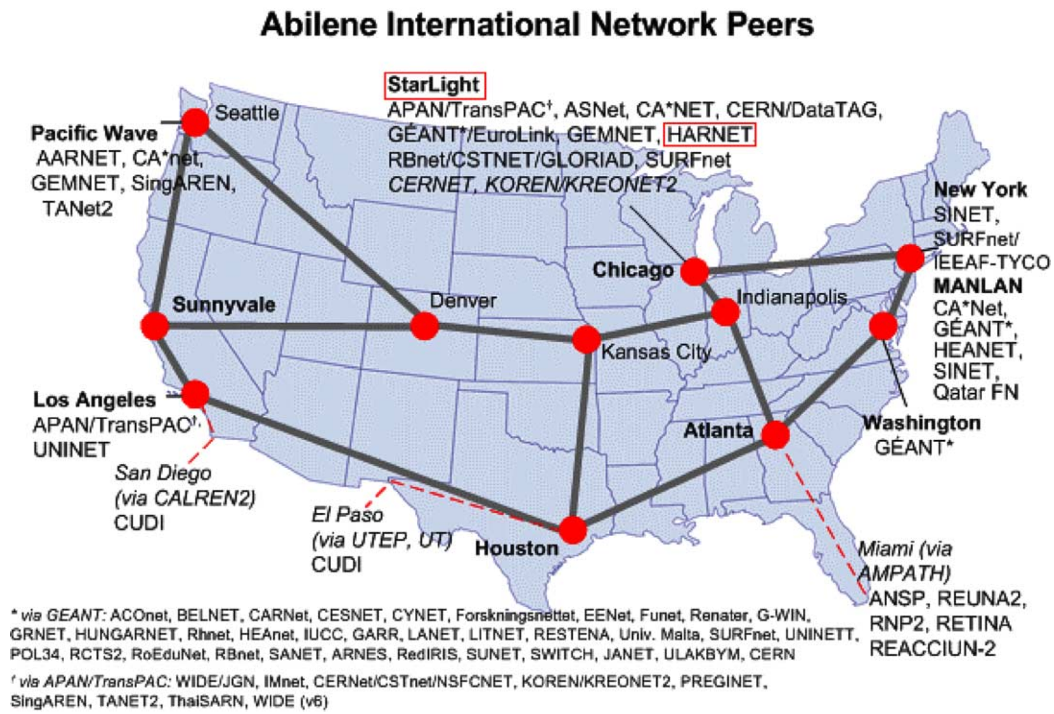


Fig. 3. Major international interconnection points of Internet-2 at Abilene backbone.

'HARNET': enclosed by rectangles in Fig. 3). As an optical version of StarTAP founded in 1997, StarLight is an advanced optical infrastructure at Chicago, facilitating a global proving ground for network services optimized for high-performance applications [13]. In Hong Kong, Pacific Century Cyber Works (PCCW), an Internet Service Provider (ISP), provides HARNET with routers and an international link between HARNET and StarLight starting October 2002.

Due to the limited funding support, the international link is a DS-3 connection with a bandwidth of 45 Mbps only. One of the international Internet-2 connection considered in this paper is the connection between the PolyU and the IPI/USC. Fig. 4 shows the network routing path of this connection, in which every individual node is identified. The topology of such routing path can influence the performance of the international connection with respect to bandwidth and latency, which will be further discussed in Section 4.

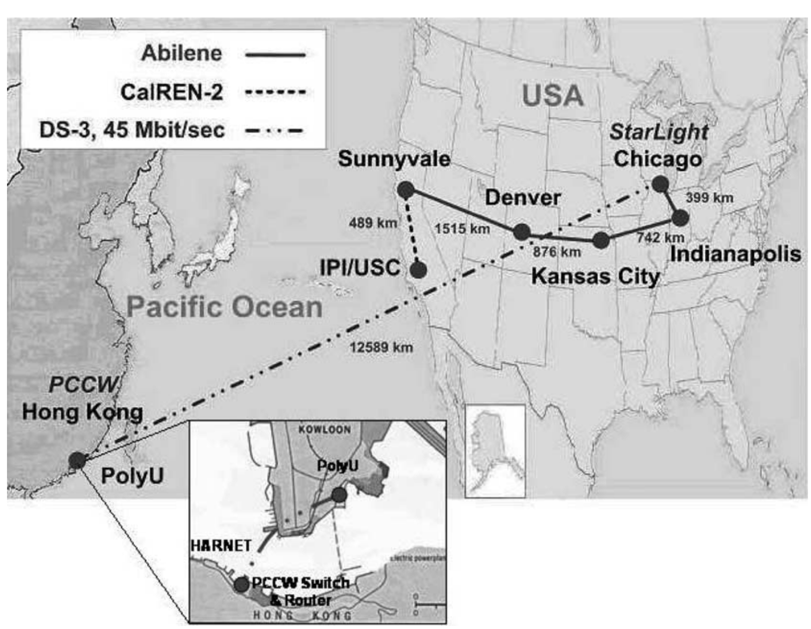


Fig. 4. Routing path between the IPI/USC and the PolyU through international Internet-2 connection (not to scale).

3. Medical imaging applications and important events

3.1. Demonstrations in Fall 2002, Internet-2 member meeting and RSNA

Through the Next Generation Internet support since 1999, the IPI/USC have been pioneering and have accumulated substantial performance data in using the Internet-2 in medical imaging applications for several years. As an international Internet-2 peer, the PACS team of the PolyU participated in the Internet2/NLM infoRAD Demos and Tutorials at RSNA 2002 at Chicago [11] and the Internet-2 Demos for the Fall of 2002 at USC. The demonstrations include the PACS ASP (Application Service Provider) image off-site back-up archive for Disaster Recovery [22], the PACS DICOM [7] remote image query, retrieval and display and wireless and web applications through several Internet-2 connections with the remote sites: UCLA, UCSF, SJHC, Stanford, UH, NLM, and the PolyU. These remote sites are shown in Figs. 1 and 4. In the RSNA demo site, two 2000-line LCD display workstations and a PACS simulator [18] were set up. At the remote sites, additional display workstations and/or PACS servers were set up. For the connection to the PolyU, both demonstrations showed well-functioning DICOM communication without any interruption of image data flow. It took about 1 min to download and display a chest image and 10–20 min to retrieve and display a set of 100× 512 kB CT or 100×130 kB MR examination. The average throughput is about 1.4 Mbps and the latency was about 300 ms. The results show the underutilization of bandwidth and the need for tuning the TCP/IP parameters.

3.2. Video-teleconferencing

The Integrated Services Digital Network (ISDN) has long been a traditional digital phone connection for the video-conferencing over a decade. Using end-to-end digital connectivity, ISDN allows audio and video to be transmitted simultaneously across the world. The deployment cost and service charge of ISDN, comparing with the connection through Internet Protocol (IP), are very high, making it less feasible to be used for research and academic purposes. Moreover, the ordinary Internet has very limited bandwidth and high latency that the audio and video quality of international video-conferencing cannot be acceptable. Internet-2 has combined the advanced network technology with H.323 audio and video (AV) standard to provide IP-based communication in an affordable and accessible way. The IP-based video-teleconferencing (VTC) application, which can connect seamlessly and compatibly the VTC machines of various vendors at multiple end-points, plays a very important role in teleconsultation.

A video-teleconference on infectious disease was held among the University of Pittsburgh, Hospital Authority Head Office (HAHO) in Hong Kong, UH and the PolyU on 10 October, 2003. The required multiple end-point network connections among the four sites are shown in Fig. 5. The STAN Hub at Honolulu acted as a coordinating node, integrating all the input AV signals from the remote sites and controlling the output AV signals to every remote site. The State of Hawaii Telehealth Access Network, known as STAN, is a state-of-the-art network that today connects not only the 12 hospitals that make up the Hawaii Health Systems Corp. (HHSC) but also other hospitals throughout Hawaii and the South Pacific. This VTC event experienced the excellent performance of AV streaming ability of international Internet-2 connection, comparing with that using the ordinary Internet.

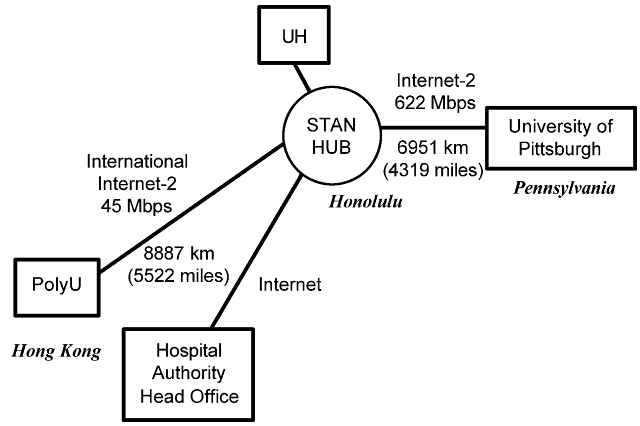


Fig. 5. Multiple end-point network connections for the Video-Teleconference on Infectious Disease held on October 10, 2003.

3.3. Daily operation: DICOM and non-DICOM image transmission

The PACS in the PolyU was installed in 2000 and built on a generic PACS simulator, which is able to accept medical images acquired from various kinds of imaging modalities [9]. A number of modalities such as Computed Radiography (CR), Ultrasound (US) and high-resolution laser scanner for Secondary Capture (SC) of hardcopy films in the PolyU radiography clinic are connected to the PACS for clinical, research and academic uses. For other kinds of medical images such as CT and MR, a modality simulator provides these images a way to enter the PACS. These images, obtained from the IPI/USC and hospitals, are used to enrich the range of medical images from various vendors that the PACS can accept and to test daily the connectivity of every component. The international Internet-2 connection has been used for daily medical image exchange between the PolyU and the IPI/USC in both DICOM and non-DICOM communications since October 2002.

4. Limitations on performance of international Internet-2

It is not necessarily true that the global access of Internet-2 retains the high performance of Internet-2 in the US.

Table 1
The detailed information of each connection in the routing path between the PolyU and the IPI/USC

Connection between		Technology	Bandwidth	Distance (km)	Round-trip response time (ms)
PolyU	PCCW, Hong Kong	ATM	45 Mbps	3	11
PCCW, Hong Kong	StarLight, Chicago	DS-3	45 Mbps	12,589	220
StarLight, Chicago	Indianapolis	OC-192	10 Gbps	399	8
Indianapolis	Kansas City	OC-192	10 Gbps	742	10
Kansas City	Denver	OC-192	10 Gbps	876	10
Denver	Sunnyvale	OC-192	10 Gbps	1515	25
Sunnyvale	IPI/USC	OC-192	10 Gbps	489	8
Overall					
PolyU	IPI/USC			16,613	292

As mentioned in Section 2, the performance of international Internet-2 connection highly depends on the topology of its routing path. Let us consider the connection between the PolyU and the IPI/USC as an example to illustrate the limitations on throughput and response time by network topology. Table 1 lists every individual connection in the routing path and their detailed information including technology, bandwidth, distance, and round-trip response time between nodes. Note that the node 'PCCW, Hong Kong' represents a router and the HARNET ATM switch.

The following sub-sections will further discuss three important issues degrading the network performance: network looping in the US, bottleneck of connection and asymmetric hardware configurations. Only the first two issues exist when considering international Internet-2 connection to Hong Kong.

4.1. Network looping in the US: inducing high latency

As illustrated in Fig. 4, the cross-Pacific DS-3 connection, Abilene and CalREN-2 form a network looping in the US continent. It is mainly caused by choosing StarLight as the global interconnection point of Internet-2. If the alternative interconnection points, such as *Los Angeles* shown in Fig. 3, were used, the looping would not exist. The total distance of the routing path between the PolyU and the IPI/USC is 16613 km. If the interconnection point *Los Angeles* were chosen, the total distance would be less than 11,680 km. It implies that the network looping produces unnecessary increase in the routing path and thus the length of carriers by about 4933 km!

The looping raises not only the cost of cross-Pacific DS-3 connection but also the round-trip response time. According to the fact that the traveling speed of any signal is limited by speed of light in a vacuum, the least round-trip response time is proportional to the distance of route [3]. Thus the longer the routing path we have, the longer the round-trip response time we get. From Table 1, the total round-trip response time is 292 ms. If the interconnection point *Los Angeles* were chosen, the routing path could exclude the connection between *Chicago* and *Sunnyvale* saving 53 ms round-trip response time and also the connection between *Los Angeles* and *Hong Kong* could take the place of that between *Chicago* and *Hong Kong*, thus shortening the length of DS-3 connection between US and Hong Kong and reducing the existing round-trip response time. Though there exists the trade-off between the cost and performance of the tremendously long cross-Pacific connection, the choice of interconnection point determines the performance of international Internet-2 to a very large extent.

4.2. Bottleneck of international Internet-2: limiting maximum available bandwidth

Among all the connections along the routing path (Table 1), the first two connections, the cross-Pacific link and the HARNET ATM network, both having 45 Mbps bandwidth only, form the bottleneck of the international connection, the throughput of which is bounded by 45 Mbps. The Information Technology Services of the PolyU finds that the total usage of international Internet-2 is still far below the available bandwidth as many international Internet-2 applications cannot fully utilize the bandwidth. Indeed, many end-users do not recognize yet that the throughput of their applications is limited by the buffer size of the operating systems they are using.

The TCP/IP communication protocol provides the sending and receiving computers with a number of parameters for data transmission, in which TCP window buffer size accounts for the amount of unacknowledged data in the flight between the sender and the receiver in the computer network. Round-trip response time is defined as the weighted average time for one complete cycle of a send and acknowledgement receipt for a single data segment between the sender and the receiver in the computer network. The greatest throughput is equal to the buffer size divided by the round-trip response time [3]. The default TCP window buffer size of most operating systems is 64 kB, i.e. 512 kbits. It is found in Section 4.1 that the round-trip response time of the connection between the PolyU and the IPI/USC is 292 ms. If the default TCP/IP settings are used in both sending and receiving, the greatest throughput in this case will be 1.75 Mbps only and the utilization of available bandwidth is less than 4%. Though there exists a bottleneck, the underutilization of bandwidth of international Internet-2 will be the burning question we need to answer. A clinical protocol for tuning the TCP/IP parameters, such as this TCP window buffer size, will be derived in Section 5 to answer this question.

4.3. Computer performances at sending and receiving ends

Besides the latency and bandwidth of the carriers, the network throughput is also determined by the maximum memory capacity of the slowest computer amongst the sender and the receiver [15]. If the capacity of the sending or receiving computer is low, the throughput will be confined to the performance of the computer with lowest capacity. Thus the hardware configurations of both the sending and receiving computers are required to be similar and at high-end technology.

Fig. 6 illustrates an example of teleradiology through international Internet-2. The mechanisms of image transmission at both the sending and receiving computers are represented by the layers of the Open Systems Interconnect (OSI) model [9,15]. The highest layer is the application/process layer, which provides the most front-end services to the users, e.g. DICOM 'send' and 'receive'. The second layer is the transport layer where the image data is decomposed into segments for sending and the received segments are combined to the image data. The number of segments in the flight is defined by the size of TCP window buffer configured at the sending computer. The Internet layer adds the IP header to the sending datagram and removes the IP header at the receiving end. The IP header contains the destination computer address, which is in IPv4 for the ordinary Internet and can be in IPv6 for the Internet-2 [5]. The lowest layer is the network access layer physically sending and receiving the data frame through the network carriers, say international Internet-2 in this example. Since these layers control how the images are transmitted from end to end, the performance of every layer can also affect the throughput of the international Internet-2 connection.

5. Repeatable and consistent clinical protocol

It is shown in Section 4 that the existing international Internet-2 applications, in an ordinary way, can only make use of at the most 4% of the usable bandwidth. The same

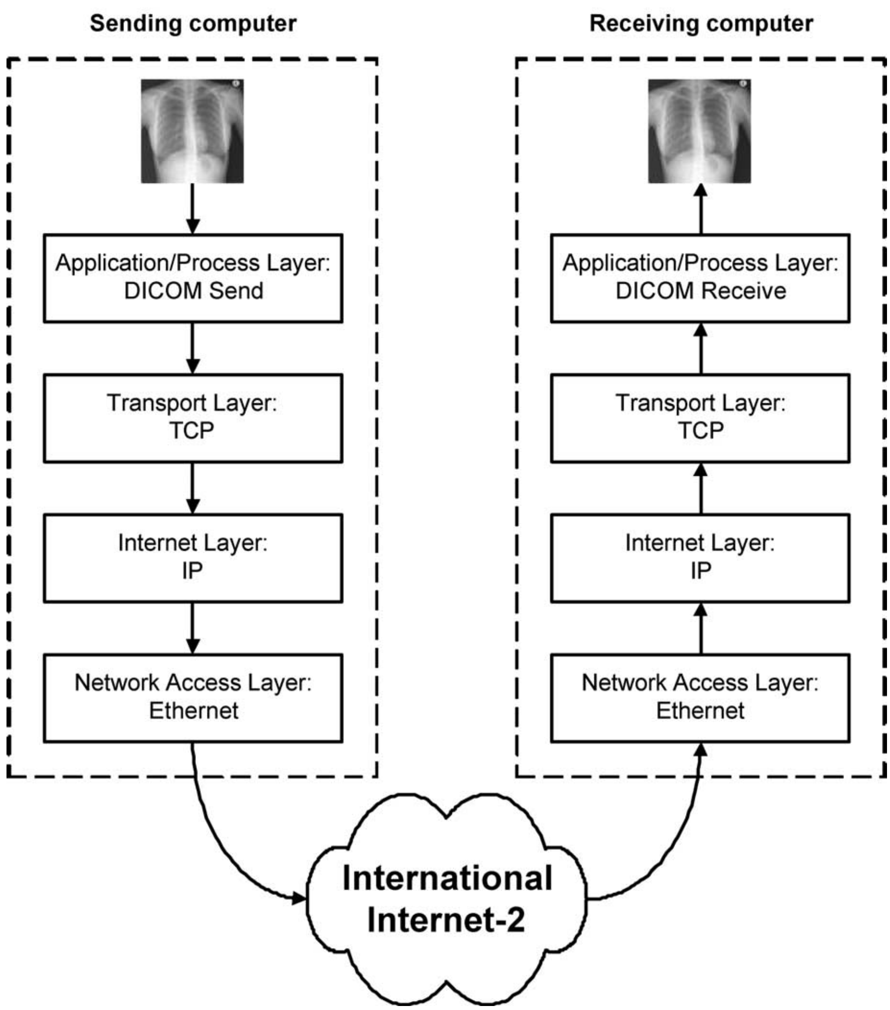


Fig. 6. The OSI models of the sending and receiving computers for teleradiology and their connection through international Internet-2.

problem also happens in many high performance networks, including the gigabits optical fiber network of Internet-2 in the US [14]. However, this bandwidth utilization problem is not easy to recognize in applications that do not require very high network throughput. Such applications include video-conferencing, in which we can have video streaming of acceptable picture quality with 384 kbps throughput. In contrast, this problem is obvious for the medical imaging applications, such as tele-imaging consultation, in which rapid large-volume image transmission and synchronized image manipulation are performed.

Supported by the National Science Foundation (NSF), the Web100 project is a joint effort between the Pittsburgh Supercomputing Center (PSC), the National Center for Atmospheric Research (NCAR), and the National Center for Supercomputing Applications (NCSA), providing the tuning software and monitoring and system diagnostic tools necessary for the computer end-users to automatically and transparently achieve high bandwidth data rates over the high performance research networks, such as Internet-2 [12].

However, the installation and implementation of the Web100 software and tools requires expert attention from the network engineers, making usage not as practical in the clinical environment. To cope with this issue, the PACS team and Information Technology Services (ITS) of the PolyU and the IPI/USC have jointly developed a repeatable and consistent clinical protocol for implementing the Web100 tuning method, which will be described in the following sub-sections.

5.1. Web100 tuning method

It is common for applications, hosts, researchers, and other computer network users not being able to take full advantage of the high performance network growth in bandwidth and accessibility. Web100 project is to enable ordinary network users to attain full network data rates without the frequent assistance from network experts, and to develop the software components necessary for a high performance network host software environment. Web100 software consists of a TCP Kernel Instrument set (TCP-KIS), Derived Instrument Set (DIS) and application code. TPC-KIS represents instruments coded directly in to the operating system kernel. DIS is the information collected based on the KIS parameters. The application code includes tools and applications that use the information provided by the KIS and DIS. These three components are aimed at enhancing the TCP/IP parameters with better kernel instrumentation and automatic controls.

5.2. Implementation in clinical environment

Web100 software optimizes automatically the network throughput by applying patch to the operating system kernels at both the sending and receiving computers. The deployment of the patch is not so straightforward and still needs the involvement of network engineers. Even experts cannot be sure that the patch is successfully applied to the kernels. It will take very long time to deploy a new and reliable computer with Web100 installed. Since the patch is installed to the kernel, which is the core of the operating system but not simply an individual application only, this large amount of manual settings will induce more human errors and make the tuning method less feasible in the clinical environment where the reliability of computer systems is required to be very high. Moreover, the deployment and setting of Web100 should be the same at both connecting peers. If the connecting peers are far apart from each other, the network engineers will be unlikely to deploy both computers by his/her own. Larger discrepancy will exist in settings of both connecting peers if the patches are applied by different network engineers. As mentioned in Section 4.3, the throughput will be confined by the computer with lower performance.

To cope with these problems, a successfully patched and stable Linux system is completely migrated to the setup disks. The setup disks are required to be compatible to a wide range of computer hardware configurations, so that the tuning protocol embedded with Linux kernel can be easily duplicated to other computers in the network. Furthermore, the automatic tuning protocol is seamlessly embedded with the kernel of the computer and thus the end-users, such as the radiologists and the physicians, will not recognize any changes in the system but the improvement in the performance. In case the operating system is crashed with other newly installed applications, it is also very easy to deploy this embedded tuning protocol again. When both the connecting peers have comparably high performance hardware configurations, we can have symmetric software and hardware configurations to optimize the capacity of the network connection. Because of its high latency induced by the network looping and low bandwidth limited by the bottleneck, international Internet-2 access from Hong Kong requires this repeatable and consistent protocol to raise the capacity of the network connection in the clinical environment. To verify the feasibility of this protocol, an experiment was performed to obtain the network performances when this protocol is deployed at both the PolyU and the IPI/USC.

6. Experimental methodology

In this paper, the FTP file transmission and the DICOM image transmission are considered in the experiments.

6.1. Hardware and software specifications

At the PolyU, computer hardware configuration is:

Model: Dell LATITUDE C840
Processor: Intel® Pentium® 4 Mobile CPU 1.6 GHz

Memory: 512 MB RAM
Hard disk: Hitachi DK23FB-20, 20 GB
Network adapter: 3Com 3C920 Integrated Fast Ethernet Controller 100 Mbps

At the IPI/USC, computer hardware configuration is:

Model: Dell Optilex GX 270.
Processor: Intel® Pentium® 4, 2.8 GHz
Memory: 512 MB RAM
Hard disk: 40 GB hardware
Network adapter: 1 Gbps network Card

At both sites, software configuration is:

Operating system: Linux 2.4.24 RedHat 8 i386
Desktop Environment: GNOME
FTP Server: 'anonftp' and 'vsftpd'
DICOM components: DICOM 'send' and 'receive'
Tuning software: Web100 release 2.3.4 for Linux 2.4.* kernels
GUI tools: Web100 Userland alpha 1.3

6.2. TCP/IP tuning parameters

Due to the transaction overhead at the routers and the application layers of sender and receiver, we assume that there is 12% IP and transaction overhead for the international Internet-2 connection. Thus,

Usable bandwidth = 45 Mbps × 0.88 = 39.6 Mbps = 4.95 MB/s

By the definition of bandwidth delay product [15,17], we need data sent in window size given by,

Maximum window size = 4.95 MB/s × 292 ms = 1.4 MB

to fill up the link. It is the theoretical TCP send and receive socket buffer sizes that fill up the link in any one time. If it is too small, the TCP congestion windows will never open up fully. If it is too large, the sender can overrun the receiver, and the TCP congestion window will shut down. The Web100 software will automatically tune all the TCP/IP parameters including this TCP window buffer size at both international peers. Note the maximum window size is an indicator of the maximum size of data segment on-the-fly. If a series of CT images, each in the size of 512 kB, the DICOM image transmission will not take advantage from the 1.4 MB window size since the transmission of images is one-by-one.

6.3. Measuring procedures

6.3.1. FTP file transmission

The baseline measurements were taken using Web100 patched Linux at both the PolyU and the IPI/USC. The FTP file transmissions are performed in the experiment where files of five different sizes are considered: 0.5, 5, 10, 20 and 160 MByte. The Linux systems at both the PolyU and the IPI/USC initiated FTP 'get' and 'put' commands one-by-one as given by the following procedures:

(1) PolyU computer initiated 'put' command. Files were sent from PolyU to IPI.
(2) PolyU computer initiated 'get' command. Files were sent from IPI to PolyU.
(3) IPI computer initiated 'put' command. Files were sent from IPI to PolyU.
(4) IPI computer initiated 'get' command. Files were sent from PolyU to IPI.

6.3.2. DICOM image transmission

Two sets of medical images are considered.

(i) CR examination, 2 DICOM images, 15.3 MByte
(ii) CT examination, 29 DICOM images, 14.6 MByte

Both computers at the PolyU and the IPI/USC, with the Web100 patched Linux, have the DICOM 'send' and 'receive' components, facilitating the DICOM communication. The images were transmitted between these two computers through the DICOM communication protocol. The Linux systems at both the PolyU and the IPI/USC initiated DICOM 'send' commands one-by-one as given by the following procedures:

(1) PolyU computer initiated 'send' command. DICOM images were sent from PolyU to IPI.
(2) IPI computer initiated 'send' command. DICOM images were sent from IPI to PolyU.

6.3.3. Comparison for different experimental settings

Consider four different combinations of system configurations:

(1) Standard Linux computers without tuning protocol were used at PolyU and IPI.
(2) Linux computer with automatic tuning protocol was used at PolyU and standard Linux computer was used at IPI.
(3) Standard Linux computer was used at PolyU and Linux computer with automatic tuning protocol was used at IPI.
(4) Linux computer with automatic tuning protocol were used at PolyU and IPI.

For each setting, a CR examination consisting of two images was sent from PolyU to IPI using FTP 'put' command and DICOM 'send' command, and the transmission time for each command were recorded. The size of a CR image is 7.65 MB and the total size of this examination is 15.3 MB. The transmission rates of different settings will be compared.

Table 2
PolyU to IPI ('put' command by PolyU computer)

File size (MB)	Average throughput (Mbit/s)
0.5	1.1
5	6.7
10	10.2
20	14.8
160	20.1
Mean	10.57

Table 4
IPI to PolyU ('put' command by IPI computer)

File size (MB)	Average throughput (Mbit/s)
0.5	1.1
5	6.8
10	10.7
20	14.8
160	22.4
Mean	11.17

7. Experimental results and discussion

7.1. FTP file transmission

Tables 2–5 show the average throughputs against the file sizes for each procedure given in Section 6. It is found that the maximum throughput we got in the experiment is 22.4 Mbit/s. It is the rate of transmitting a 160 MB file from IPI to PolyU by the FTP 'put' command. The tuning protocol leads to almost 50% utilization of available bandwidth of the international Internet-2 connection, which is above 12 times of the upper limit of throughput without tuning, 1.75 Mbit/s, given in Section 4.2.

Fig. 7 shows that the throughput increases with the file size for any combination of transmitting direction and command. It is because the IP and transaction overheads account for a fixed time taken in the total transmission time no matter how large the file is, while the time required for on-the-fly sending the file over the network increases with the file size. Thus, the percentage of these overheads in the total transmission time will become comparatively smaller and then the throughput will become larger, when transmitting larger file.

When comparing the throughput means of Tables 2–5 with respect to the direction of file transmission, the throughput for file transmission from PolyU to IPI and that from IPI to PolyU are similar to each other. There is no any equipment, such as firewall, which can make asymmetric throughput results. Moreover, it is found that the throughput mean using 'get' command is smaller than that using 'put' command since the 'get' command requires more transactions than the 'put' command does.

The throughput is very small when transmitting files of 0.5 Mbyte. It is because the file size is smaller than the theoretical maximum window size and thus the tuning protocol will not take any effect on the throughput in transmitting files of small size.

Table 3
IPI to PolyU ('get' command by PolyU computer)

File size (MB)	Average throughput (Mbit/s)
0.5	0.7
5	5.4
10	8.9
20	11.5
160	17.7
Mean	8.83

7.2. DICOM image transmission

Tables 6 and 7 show the transmission rates for CR and CT examinations from PolyU to IPI and that from IPI to PolyU. It is found that it takes not more than 41 s to transmit a CR examination of 15.3 MB in size and the transmission rate is about 3 Mbps. The transmission rate is higher than that using ordinary Internet but it is not comparable to the result for the FTP file transmission. It is because the DICOM 'send' and 'receive' components also have their own buffer sizes for DICOM image transmission and for the time being these buffer size values are not synchronized with the Web100 tuning. Since the DICOM 'send' and 'receive' components are working independently with the Web100 tuning, the performance of the DICOM transmission will not be improved too much. For the CT examination, this phenomenon becomes more obvious. It is found that the transmission rates are 0.78 and 0.88 Mbps for PolyU-to-IPI and IPI-to-PolyU, respectively. This result is very close to that using FTP transmission for 0.5 MB file, having the same size as a CT image, where the file size is too small when comparing with the tuned window buffer size and thus the tuning did not affect too much on the transmission rate.

7.3. Comparison for different experimental settings

Table 8 shows the summary of the transmission rates for different experimental settings. It is found that highest transmission rates, 9.79 Mbps for FTP and 3.06 Mbps for DICOM, were obtained at the setting that both PolyU and IPI were using the automatic tuning protocol. The rest of the settings yielded transmission rates for FTP ranging from 1.63 to 2 Mbps and that for DICOM ranging from 1.36 to 1.38 Mbps. The FTP transmission rate was increased to about 5–6 times, while the DICOM transmission rate was increased to about 2.2 times. The improvement of

Table 5
PolyU to IPI ('get' command by IPI computer)

File size (MB)	Average throughput (Mbit/s)
0.5	0.8
5	5.6
10	8.9
20	13.0
160	20.3
Mean	9.71

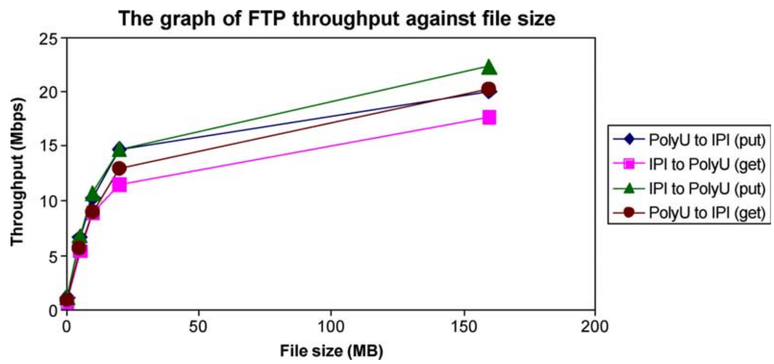


Fig. 7. The FTP throughput against file size for different transmission direction and FTP command.

the DICOM transmission rates was less than that of the FTP transmission because the DICOM 'send' component has not been linked up with the tuning protocol and the window buffer size of the application/process layer (the DICOM 'send' component) is not synchronized with that of the transport layer (TCP with the automatic tuning protocol).

8. Summary

Because of its high throughput and short response time, Internet-2 is an ideal computer network for medical imaging applications. However, the global access of Internet-2 cannot retain the high network performance as what we get within the US continent. It is because there is a bottleneck at the international link between Internet-2 in the US and the international network peer. For the international connection to HARNET in Hong Kong, the available bandwidth is limited by 45 Mbps. Furthermore, the choice of interconnection point, StarLight, leads to a network looping in the US and thus increases round-trip response time by 53 ms at least. Considering the connection between the PolyU and the IPI/USC, the total round-trip response time is 292 ms. If the default TCP/IP settings are used, the upper limit of throughput under such a long round-trip response time is 1.75 Mbps only. An automatic tuning protocol under Linux platform is derived and presented in this paper. The protocol is aimed at tuning the window size seamlessly and automatically to increase the network throughput. It will be beneficial to the applications such as teleradiology and teleconsultation in clinical environment because this method can yield fast round-trip response and high network throughput without any user intervention. The experimental result shows that the tuning protocol can increase the maximum throughput to 22.4 Mbps, which is above 50% of the available bandwidth. There is a very large improvement when comparing with the theoretical throughput without tuning, 1.75 Mbps, and the experimental throughput without tuning, about 1.4 Mbps, respectively. For the DICOM image transmission, the transmission rate for a 15.3 MB CR exam is about 3 Mbps, which is still very slow when comparing with the throughput using non-DICOM FTP file transmission. Lower performance was obtained when CT exam was considered. It is found that the DICOM transmission rates are 0.78 and 0.88 Mbps for PolyU-to-IPI and IPI-to-PolyU, respectively. The performance was low because the image size of CT is too small when comparing with the tuned window buffer size and thus the tuning takes less effect on the transmission rate. Among different experimental settings, the highest transmission rates, 9.79 Mbps for FTP and 3.06 Mbps for DICOM, were obtained by the setting that the automatic tuning protocol is applied at both PolyU and IPI/USC. However, the improvement of the DICOM transmission is much smaller than that of the FTP transmission because the existing

Table 6
PolyU to IPI (DICOM)

Exam.	No. of images	Total size (MB)	Time taken (s)	Transmission rate (Mbps)
CR	2	15.3	40	3.06
CT	29	14.6	150	0.78

Table 7
IPI to PolyU (DICOM)

Exam.	No. of images	Total size (MB)	Time taken (s)	Transmission rate (Mbps)
CR	2	15.3	41	2.99
CT	29	14.6	133	0.88

Table 8
Comparison of transmission rates for different combinations of system configurations

	IPI/USC	
PolyU	Standard Linux	Linux with automatic tuning protocol
Standard Linux	FTP: 1.63 Mbps DICOM: 1.36 Mbps	FTP: 2.00 Mbps DICOM: 1.36 Mbps
Linux with automatic tuning protocol	FTP: 1.91 Mbps DICOM: 1.38 Mbps	FTP: 9.79 Mbps DICOM: 3.06 Mbps

CR examination consisting of two images and having total size of 15.3 MB was considered.

DICOM components are using fixed window buffer sizes and are unable to synchronize these parameters with the automatic tuning protocol. It is suggested to link up the DICOM components with the automatic tuning in terms of the TCP/IP parameters, such as window buffer sizes, to yield the optimal performance subject to the current situation of the international Internet-2 connection.

Acknowledgements

The authors wish to thank University of Hawaii, which loaned the VTC devices to the PACS team of the PolyU for this research. ITS/PolyU also provides substantial technical support in applying Web100 patches to the Linux kernel, making the proposed tuning protocol possible.

References

[1] Chan LWC, Cao F, Zhou M, Hau SK. Connectivity issues and performance monitoring of international Internet-2 in tele-imaging consultation, Education exhibit, RSNA; 2003.

[2] Huang HK, Lou SL. Telemammography: a technical overview. In: Haus AG, Yaffe MJ, editors. RSNA categorical course 1999, Oak Brook, IL, 1999. p. 273–81.

[3] Cao F, Huang HK, Zhou M, Mogel G, Chan L, Gill M. Internet2 end-to-end performance for medical imaging applications. Submitted to J High Speed Netw.

[4] Yu F, Hwang K, Gill M, Huang HK. Some connectivity and security issues of NGI in medical imaging applications. J High Speed Netw 2000;9:3–13.

[5] Internet2 consortium. UCAID, http://www.internet2.edu.

[6] Abilene network, http://abilene.internet2.edu/.

[7] Digital imaging and communications in medicine (DICOM). National Electrical Manufacturers' Association. Rosslyn, VA: NEMA; 1996.

[8] http://networks.internet2.edu/rons.html

[9] Huang HK. PACS and imaging informatics: basic principles and applications. New York: Wiley; 2004.

[10] California research and education network. CalREN2, http://www.calren2.net.

[11] Mogel G, Cao F, Huang HK, Zhou M, et al. Internet2 performance for medical imaging applications. RSNA 2002 Fall meeting, http://apps.internet2.edu/rsna2002-demos.html.

[12] Web100 project, http://www.web100.org/.

[13] StarLight, http://www.startap.net/starlight/.

[14] Huntoon W, Ferguson J. Internet2 presentation—May 2002, in I2 Members meeting May 2002, http://www.web100.org/docs/Web100_I2MM.ppt.

[15] Shanmugam R, Padmini R, Nivedita S. Special edition using TCP/IP. 2nd ed. Indianapolis, IN: Que Pub; 2002.

[16] Chan LWC, Li R, Leung K, Hau SK, Wong ECH, Cao F. International Internet2 connectivity requirements for tele-imaging consultation. PACS and integrated medical information systems: design and evaluation (M105), SPIE 2003 conference.

[17] Tierney BL. TCP tuning guide for distributed application on wide area networks. Networking/Computing 2001;February.

[18] Zhou Z, Law MY, Huang HK, Cao F, Liu BJ, Zhang J, et al. Educational RIS/PACS simulator. SPIE Med Imaging 2003;5033: 139–47.

[19] Zhang J, Stahl JN, Huang HK. Real-time teleconsultation with high-resolution and large-volume medical images for collaborative healthcare. IEEE Trans Inform Technol Biomed 2000;4(2).

[20] Huang HK. Teleradiology today, telehealth, business briefing: next generation healthcare, p. 1–7.

[21] Stahl JN, Zhang J, Zellner C, Pomerantsev EV, Chou TM, Huang HK. Teleconferencing with dynamic medical images. IEEE Trans Inform Technol Biomed 2000;4(2).

[22] Huang HK. PACS, image management, and imaging informatics: a tutorial. Hong Kong Radiographers J 2000;6(1).

[23] Remote treatment planning for radiation therapy. Internet2/NLM infoRAD exhibit, RSNA 2003, http://apps.internet2.edu/rsna2003-demos.html#RemoteTreatmentPlanning.

Lawrence Chan is the research fellow and PACS manager of the Optometry and Radiography Department at the Hong Kong Polytechnic University. He was awarded by the University of Hong Kong his BEng and PhD in Mechanical Engineering in 1996 and 2001, respectively. In the postgraduate research, he has been working in the field of artificial intelligent control for more than 4 years. In 2000, he joined the Jockey Club Research and Information Centre at the University of Hong Kong as database administrator for almost 2 years. His current research interests include Internet-2 application in medical imaging, PACS-RIS-CMS system integration and Radiation Therapy Information System.

Z. Zhou, a PhD candidate of Biomedical Engineering Department at University of Southern California, is a research assistant at Image Processing and Informatics (IPI) Laboratory, Radiology, USC now. He has 6 years experience in developments and researches on PACS and Medical Imaging Informatics. His current interest is in medical image security, grid computing for medical imaging and other medical imaging related applications.

F.H. Tang is an Assistant Professor of the Department of Optometry and Radiography at the Hong Kong Polytechnic University. He received his MPhil degree in Biomechanics and PhD degree in Computer Science in 1992 and 2000, respectively. Dr Tang has been working as qualified Diagnostic Radiographer until 1994 and thereafter he joined the University. Since then, he has been leading the development of Medical Informatics in the BSc Radiography program. Besides his academic activities, Dr Tang initiated the development of the Picture Archiving and Communication System (PACS) with collaboration of Professor HK Huang in the University in 1999. His current research interest includes image processing, PACS, image content retrieval, image registration and E-learning. He is also a member of the editorial board of Hong Kong Radiographers Journal and a Board Examiner of the Hong Kong Radiographers Board of the Hong Kong SAR government.

Computerized Medical Imaging and Graphics

Computerized Medical Imaging and Graphics xx (xxxx) 1–11

www.elsevier.com/locate/compmedimag

Designing high-quality interactive multimedia learning modules

Camillan Huang

Education and Multimedia Applications, Stanford University School of Medicine, SUMMIT @ Wallenberg Hall, 450 Serra Mall, Building 160, Stanford, CA 94305-2055, USA

Received 6 February 2004; revised 25 August 2004; accepted 30 September 2004

Abstract

Modern research has broadened scientific knowledge and revealed the interdisciplinary nature of the sciences. For today's students, this advance translates to learning a more diverse range of concepts, usually in less time, and without supporting resources. Students can benefit from technology-enhanced learning supplements that unify concepts and are delivered on-demand over the Internet. Such supplements, like imaging informatics databases, serve as innovative references for biomedical information, but could improve their interaction interfaces to support learning. With information from these digital datasets, multimedia learning tools can be designed to transform learning into an active process where students can visualize relationships over time, interact with dynamic content, and immediately test their knowledge. This approach bridges knowledge gaps, fosters conceptual understanding, and builds problem-solving and critical thinking skills—all essential components to informatics training for science and medicine. Additional benefits include cost-free access and ease of dissemination over the Internet or CD-ROM. However, current methods for the design of multimedia learning modules are not standardized and lack strong instructional design.

Pressure from administrators at the top and students from the bottom are pushing faculty to use modern technology to address the learning needs and expectations of contemporary students. Yet, faculty lack adequate support and training to adopt this new approach. So how can faculty learn to create educational multimedia materials for their students? This paper provides guidelines on best practices in educational multimedia design, derived from the Virtual Labs Project at Stanford University. The development of a multimedia module consists of five phases: (1) understand the learning problem and the users needs; (2) design the content to harness the enabling technologies; (3) build multimedia materials with web style standards and human factors principles; (4) user testing; (5) Evaluate and improve design.

Keywords: Educational media; Interactive teaching for biology and medicine; Instructional web design; Physiology modules; Multimedia module design

1. Introduction and background

Technology is outpacing us. In the 20th century, technology has become a staple in our everyday lives and has catalyzed innovations in healthcare and biomedical research. This progression has enriched our lives and has brought about the need for imaging informatics databases. However, it has also become difficult for today's student to learn and assimilate the growing amount of content found in these databases (without misconceptions). Technology creates a second problem by changing the expectations of incoming students. Today's students are so surrounded by technology (computers, chat, email, and the web) that it is natural for them to expect coursework to incorporate similar standards. How can we, as teachers, incorporate technology into our teaching so that we can train our students for tomorrow? In turn, how can incorporating technology into our teaching benefit teachers?

2. Teaching: past, present, and future

Traditionally, print textbooks have been the standard reference and learning tool for students. In general, such textbooks are thorough (they cover a large number of topics in detail), well-organized, and incorporate the basics of the life sciences. However, textbooks are static, are not easily customized to different students and classes, and fail to adequately highlight the intersections between different disciplines. Textbooks also cannot provide students with information on the newest scientific breakthroughs since revisions are a major effort and they are costly to students.

0895-6111/$ - see front matter © 2004 Published by Elsevier Ltd.
doi:10.1016/j.compmedimag.2004.09.017

Informatics can address the need that textbooks cannot provide. Informatics can be used to provide students with up-to-date information that can be disseminated freely through the Internet. A wealth of information datasets can be found in imaging informatics, like computerized tomography (CTs), magnetic resonance imaging (MRIs), functional MRIs, and positron emission tomography (PET) that can be centralized and distributed. Information in these databases may also include simulations and systems modeling. These datasets and patient histories are today's electronic medical encyclopedia and allow interdisciplinary users to integrate, visualize, and cross-reference different studies performed on areas in our body. However, this information often lacks a usable instructional interface necessary to transform the information from a reference tool into a powerful training tool to help students learn.

Educational media, in contrast, can be powerful to train students and is considered the next generation of learning materials. Designed to complement traditional educational methods and information datasets, educational media can supplement core material with animations, interactivity, and visual design. Unlike traditional textbooks, educational media is dynamic, easily customizable, and can be designed with an interdisciplinary approach. It can also be used to teach content in a way traditional teaching materials cannot—for example, animations of mechanisms and processes can help students visualize how biological systems work together. Once taught, interactive media can be used to ensure that students have learned key concepts and understand the basics. In addition, students are more likely to adopt these materials since they are influenced by a world of computers, media, and the Internet.

For teachers, multimedia education can help explain difficult concepts more clearly than a textbook or PowerPoint lecture. When students are able to manipulate experimental factors to see cause-and-effect relationships, they move beyond rote memorization and passive learning to truly understand the material. Educational media can also be designed to correct common misconceptions by targeting difficult or frequently misunderstood concepts.

So what is educational media? There is a trend to move content into the digital medium PowerPoint (Microsoft, Redmond, WA)—but it is still passive and has only limited animation and video support. In addition, many 'interactive modules' contain no more interaction than a click to advance to the next page. True educational media should incorporate dynamic animations, interactivity, and visual design to stimulate, challenge, and test students. The design of a good instructional interface requires an integrated design approach that incorporates best practices from in education, human computer interaction, and instructional technology.

As an example, the Virtual Labs Project, funded by the Howard Hughes Medical Institute, is an initiative to augment the core undergraduate courses in biology at Stanford University. This is achieved by delivering interactive multimedia over the Internet to teach big picture and difficult concepts in physiologic systems (cardiovascular, gastrointestinal, respiratory, renal, vision, cranial nerves, and other mini modules). The concepts in these modules create a foundation not only for the life sciences and medicine, but also for an increasing number of interdisciplines, such as biomedicine, medical informatics, and bioengineering.

For example, if a medical student understands how the kidney produces and concentrates urine, he or she will be better able to understand how diuretic drugs work. A researcher, who learns about the molecular aspects of the concentrating mechanism, could apply this knowledge to understand how sodium influences body fluid maintenance. A bioengineer could apply the same knowledge to design a more efficient kidney dialysis system.

From 2000 to the present, Virtual Labs has tested their interactive multimedia modules on thousands of students in Stanford classrooms and those of our collaborators. And with over 500 pages of development experience, Virtual Labs has developed best practices to aid future module development. These best practices can be applied to any scientific concept.

2.1. How do we create new educational media resources?

Key to the success of the VL modules, and educational media materials, lies in the presentation of information that integrates an appropriate media technique (imaging dataset, technical illustration, animation, interactivity) with best practices in learning (Table 1). In addition, VL emphasizes a user-centric design, that is, we make design decisions based on human factors and match them to the user's needs and expectations. This paper will incorporate these design strategies into a detailed protocol to help guide module development by faculty and a multimedia development team.

When designed correctly, a multimedia module can visually stimulate a student and transform learning into an active, engaging process. A good design [24,25] will allow students to (1) visualize difficult and naturally dynamic concepts, (2) promote active learning, problem-solving, and critical thinking with interactive simulations and virtual environments, (3) interact with the content with

Table 1
The table below lists the pedagogical and human–computer interaction design principles behind Virtual Labs

Design principle	Reference
How people learn	[1–3]
Instructional design and evaluation	[4–10]
Interaction design usability, and human factors	[11–15]
Presentation of information: visual perception, style guides, design strategies	[16–21]
Motivational strategies	[22,23]

self-quizzes, and (4) access content anytime, anywhere, at any pace.

3. Interactive media design: from concept to reality

This section describes the design process for educational media production. It begins by providing special considerations for user-centric design, then discusses the required planning, finishing with an introduction to the five phases of development.

3.1. Design for your users' needs and expectations

It is important to consider the user's perspective and learning method when designing an educational media module. During the entire process, keep the user's voice in your head. Specifically, ask yourself what the students need to support their learning and fulfill their expectations. The following is a list of questions that should be answered by and for the user.

In parentheses are the phases (for details on Phases 1–5, see *Process*) during which these questions will be most critical during development.

- Show me what I will learn. (Phase 1–2)Clearly state the learning goals and learning outcomes of each page.
- Why should I care to learn this? (Phase 2)Give a real-world example of the concept to make it relevant to the student. Without giving a significant context to the student, the student may not understand the implications or know how to apply the principles of the concept.
- What can I do on the page? (Phase 2–3)Design the interactions on the page to engage the student and support learning. Consider teaching the concept by using interactions and visualizations that cannot be accomplished in a classroom or by a piece of paper.
- Tell me how I'm doing! (Phase 2, 3, 4)Design assessments to ensure that the students are learning the concepts. Include a way for students to self-gauge how well they have mastered the learning material.

3.2. Planning

While considerations for the user are core to the development process, planning is also essential to effectively and efficiently develop a multimedia module. Each phase requires planning, teamwork, and flexibility. A careful initial plan will save time, costs, and team frustration during the downstream production. During the planning phase, the content expert/teacher presents the educational challenge to all team members. All team members—scientists, educators, technology specialists—draw upon their domain of expertise and brainstorm collectively to design a module that integrates what the users need with what they should learn (Fig. 1). A prototype is developed from the preliminary plans and then tested on typical users. Based on the user's feedback, the module returns to production and is redesigned for improvement. The original plan may often evolve due to resource restrictions, programming considerations, and user feedback. This flux is common since the development process evolves constantly and the team needs to be flexible to adapt. In the end, adaptability will keep the project on target with deadlines and with your user's needs. The following sections describe in detail how to complete each phase.

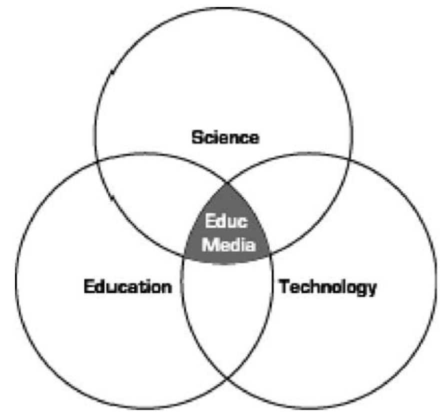


Fig. 1. *Interdisciplinary design.* Good educational media design results from integrating best practices in teaching and learning from the disciplines of science, education, and technology. The team consists of content experts from the sciences, education experts, and technology specialists. Together the team can create materials that integrate the best practices in teaching and learning from all disciplines.

3.3. Process

The process of creating multimedia learning modules consists of five phases: (1) Understand, (2) Design, (3) Build, (4) Test, (5) Improve (Fig. 2). Phase 1 and 2 is primarily led by the content expert, while the multimedia development in Phase 3 is primarily performed by the development team. Phase 4 and 5 includes all team members and the users. Below is a brief description of the five phases.

- Phase 1. Understand—understand who will use the module *(target group)* and address how the module will help the users learn *(educational challenge, needs assessment).*
- Phase 2. Design—design the module for your user *(learning design)*, from a user's point of view *(user-centric).*
- Phase 3. Build—build interactivity and multimedia components using best practices *(media development and coding).*
- Phase 4. Test—see how well your users respond to the module *(user testing, usability heuristics).*

Fig. 2. *Summary of module development process*. The figure below outlines the general project flow and the main phases in module development. Ideally, all team members—content experts, development team, and users—should be involved in the entire process. Shaded areas on the team members indicate areas where it is necessary for them to participate. A general framework [24] and evaluation methods [25] were discussed in previous published works.

- Phase 5. improve—Evaluate how well students learned and how well the module performed *(evaluation of learning outcomes).*

The five phases above is expanded in detailed below. Each phase is broken down into several components and each component lists questions to help guide your thought process. As you go through each step, it will be helpful to write out the answers to each question. The answer to your questions should be shared with your team to provide groundwork and rationale for the project. These answers also serve as the reference point for the multimedia development and for yourself as you get the results from the users you test on.

4. Phase 1: understand the problem and needs

Phase 1 focuses on the educational challenges of the student. These challenges are used to determine the design rationale of the module (for example, see Fig. 3). Draw upon your personal teaching experience and apply your classroom knowledge towards designing the module. The first issue to address is identifying what you want your students to learn (learning goal) and then how you plan to integrate what your students need and expectations.

1. Understand needs and goals.
   - Why do you want to build this multimedia module?
   - What do you want the outcome to be?
   - Who will your users be?
2. Determine learning goals.
   - What do you want the students to gain after using the module—in terms of knowledge, understanding, or skills?
   - How do you plan to have your students use the module and how would this module fit into your teaching curriculum?

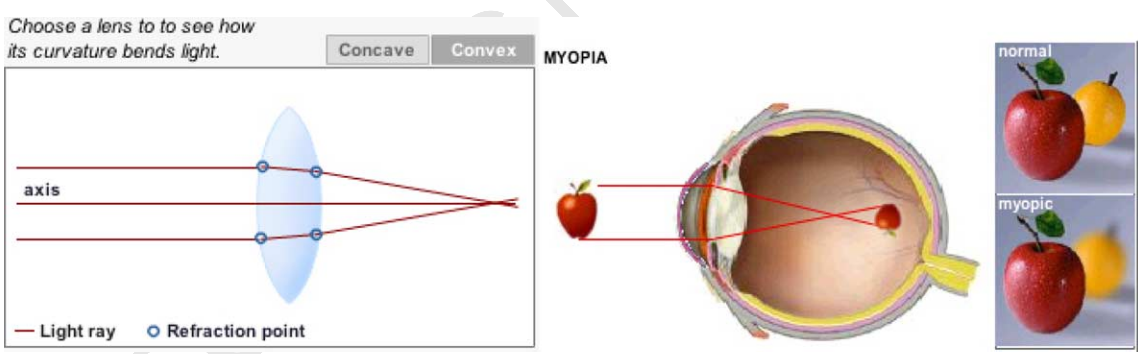


Fig. 3. *Design Rationale*. In Fall 2002, the Alliance for Lifelong Learning offered an online distance learning course called Art and Vision which explored the process of vision: how an object's color, texture, and shape is interpreted by the nervous system and the impact of visual diseases for artists. The students were post college graduate learners who were located around the world. Student materials included the textbook 'The Eye of the Artist' by Dr Michael Marmor, course guidebook, CD-ROM on vision from Virtual Labs, and access to a supporting website with videos, discussion forums, and weekly chat sessions. The module was customized to follow the curriculum in the book. The goal of the VL module supplement was to provide an integrated, interactive approach for students to learn the physiologic process of vision and apply these principles to understand how vision (or a lack of) affects art and vice versa. The figure below, left, illustrates the physics behind lenses and the one, right, shows how nearsightedness results from a change in the focusing power of the cornea. Both images show an animation of how light rays entering the eye are refracted. In a subsequent simulation lab (not shown), a student is given a myopic or hyperopic eye and the student prescribes the correct eye glasses prescription after dynamically adjusting the focal length of the lens to correct for the disorder. Meanwhile, a corresponding image from a first-person view also changes dynamically to reflect the changes in the focal length. Such content presented as a single static image would not engage a user as much as a dynamic image that the user could control and interact with. This type of interaction allows the user to make predictions about the action he or she is about to perform while receiving immediate feedback on his or her actions.

5. Phase 2: design the content for the module

Phase 2 walks through the thought process for identifying a concept and presenting it for the students in an engaging way that helps them learn. An example is shown in Fig. 4.

1. Pick a concept.
   - What concepts do your students repeatedly have a difficult time understanding?
   - Are these concepts essential, important, or common? Why are the essential concepts important for the student to learn?
   - What are the common misconceptions for each of these concepts?
   - How do you know if students are having a misconception and how have you corrected for it before?
   - What are the 'correct' models for each misconception?
2. Decide how to present the concept.
   - How could you present the material using technology enhancements so it makes learning easier?
   - Use the web to identify what resources exist. Identify websites that specifically address your educational challenge and websites that present material in a similar format.
     - What did you like about the websites you saw?
     - How would you improve it, given unlimited time and resources?

     Web research is essential to be exposed to other representations of the content. This process may find an educational tool that has already been created!
   - How will the student be challenged by what is presented in the module?
3. Design activities to engage and motivate the student.
   - What would your typical student find interesting about the topic?
   - How can you present the material so your target student is interested in it?
   - What is a real-world example of the concept? It is often helpful to use everyday examples so students can relate to the context.
4. Devise assessments to evaluate how well students learned.
   - How will your student have benefited from using these materials over traditional materials?
   - What do you hope your student will gain after using the module?
   - What is the test to see if your student understood this concept?
   - How will you know if the module was effective?

6. Phase 3: build the multimedia assets and interactivity

Phase 3 describes the process from assembling a multimedia development team to storyboarding ideas and prototyping the module.

6.1. The interdisciplinary development team

A solid team with the appropriate skills and teamwork is key to the success of a multimedia module. This team consists of experts in education, science teaching, and technology who collaborate to design the blueprint for the module. This interdisciplinary team should consist of:

- A project manager
- A content expert with teaching experience
- A multimedia development team
  - Graphic designers
  - Educational media developers/ learning designers
  - Programmers
- An evaluation expert
- The users!

6.2. Project manager

The project manager is the liaison between the stakeholder of the project, who is usually the funding agency or content expert, and the team members. The manager allocates tasks for the team and sets milestones, deliverables, and deadlines for the project, all within the proposed budget and resources. The manager should have a strong background in the sciences, technology, and education to understand how to allocate the time and resources for the tasks and provide the backbone for the interdisciplinary team.

6.3. Content expert and teacher

The content expert is often the expert in the scientific field and has extensive teaching experience. The content expert designs the student learning outcomes and learning goals for the module. This expert also helps to design user assessments to validate that the student has understood the module.

6.4. Graphic designer

The graphic designer must have artistic abilities and is trained in visual communication, visual design, and have extensive knowledge with Adobe Photoshop and Illustrator. The designer transforms the abstract learning content into a visual language and layout that communicates the learning outcomes and learning goals.

6.5. Educational media developer/learning designer

The educational media developer is trained in learning with technology design (human–computer interaction, user-centered design principles, user studies, iterative design, rapid prototyping, and evaluation). The developer designs the user interactions and validates the learning effectiveness of the module.

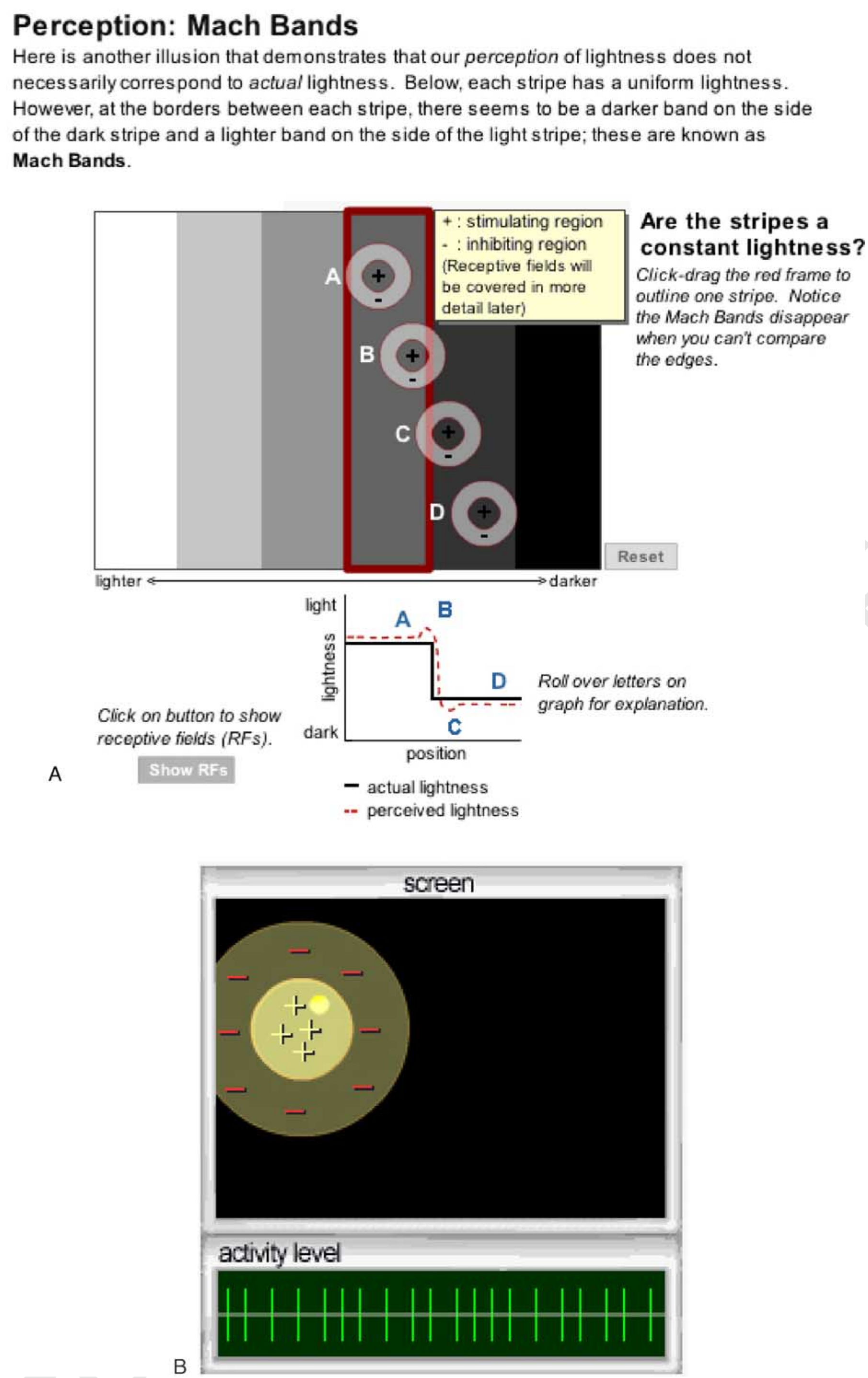


Fig. 4. *Content design*. One example topic, receptive fields, from the vision VL module is presented. We picked this topic because students struggle to understand how the physiologic concept of receptive fields creates the perception of contrast. A receptive field is defined as the set of stimulus or characteristics (light in this example) that a neuron optimally responds to. This concept is fundamental in understanding how the visual system constructs contrast. The visual system accomplishes this by constructing ‘receptive fields’ to respond to, filter, and process the incoming light stimulus. This transformation is difficult since it requires knowledge of retinal anatomy, movement of light through the retina, and the interaction of neurons. A. To represent how contrast is created, we incorporated a common illusion to illustrate the example. When the user comes to this page, he or she is instructed to look at the borders between the stripes and observe the appearance of a small gradient, called Mach Bands, at the intersection of each border. The neighboring stripe helps to enhance the crispness of the edges by making the one edge darker on one side of the edge and lighter on the other. The user can remove the illusion by moving the thick rectangular outline to remove the influence of the neighboring stripe. Without allowing a user to control what he or she sees, the user may be apt to absorb the information passively instead of actively understanding the results of his or her actions. B. Using a virtual lab with animations, a student can learn about receptive fields. This virtual lab was designed to allow the user to perform a simulated experiment of the original scientific discovery of receptive fields by Hubel and Wiesel [28]. This principle is often represented as a series of static figures where a spot of light is shown somewhere on each concentric circle and a graph shows the resulting

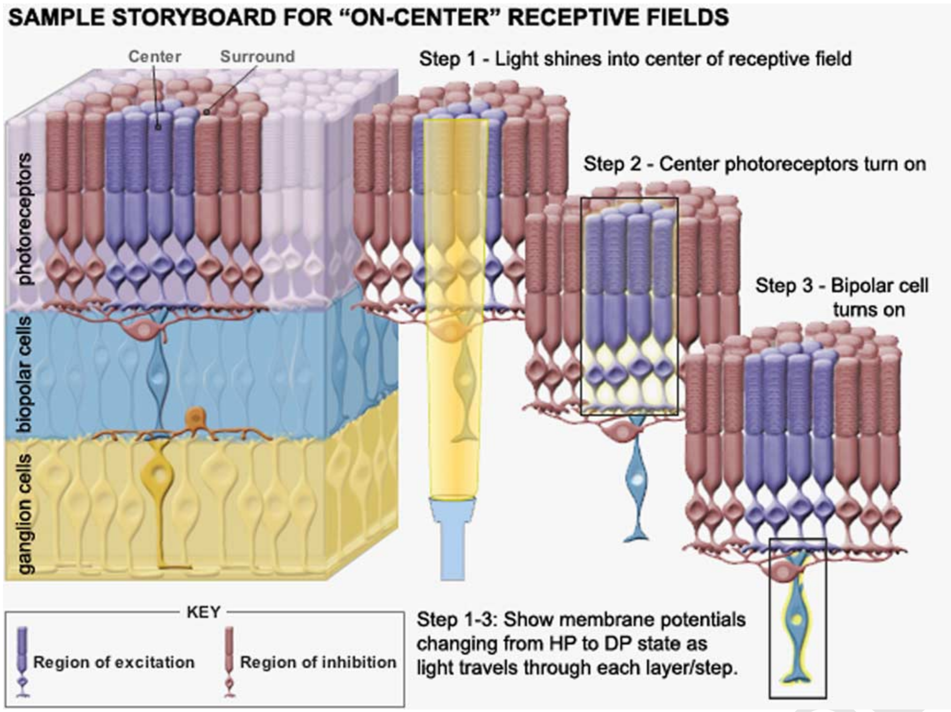


Fig. 5. *Storyboard*. A storyboard should include sketches (finished drawings are included here) to describe the details of the process in a step-wise manner. This storyboard has three key steps in time and the black box showed which areas are activated at which time. A corresponding graph showing membrane potentials would be shown next to this animation (not represented in this storyboard). A textbook may represent this animation almost as indicated below, but a dynamic animation of this principle shows the user a progression of the process over time. Animations can help decrease misconceptions since students often struggle to stitch steps together accurately.

6.6. Programmer/technical lead

The programmer provides technological experience and authors the interactive content for the multimedia. The programmer should be able to develop rapid digital prototypes.

6.7. Evaluation experts

Evaluators provide expertise on formative and summative assessments, measures for educational achievement, conduct field tests on users, design how to implement the module into the curriculum, and report to stakeholders on user feedback.

6.8. Users

A user is defined as any student today who will be using the learning module, usually one who is taking a class relating to the content in the module.

6.9. Storyboard ideas

A storyboard is the visual blueprint of the module; it describes the details of the module pages and choreographs the teamwork. It should describe how the content is laid out, list the media elements (images, animations, video, sound), and outline the user interactions (Fig. 5). A well-thought out and detailed storyboard is essential for the production team to translate your thoughts into a working module.

6.10. Develop assets

There are five main components that make up a multimedia module: content, media, interface layout, programming, and project administration.

— Content
  ○ Write material using world wide web style and content guidelines [15,16].

activity level when the spot is first off, then on. In this virtual lab, the user drags an recording electrode to the retina (not shown) and then explores the black screen with a spot of light to find the areas of high activity (indicated with a '+') or low activity (indicated with a '—'). The students can map the activity levels they discovered to see how this receptive field contains a center circle of high activity and an outer field of lower activity (indicating an On-centered cell). Then to test their understanding, students answer questions relating to their result. If the student moves the recording electrode to a new location, the activation zones, location on screen, size and shape of the receptive field change. After completing this virtual experiment, students learned about the different types of receptive fields as well as the dynamic nature of the research and exploration process. This learning process may not occur if the user memorizes the series of static figures.

    - Organize content into meaningful chunks of information. For example, bullet points can be used similar to PowerPoint.
    - Write concisely and minimize jargon to ensure the message is clear [17].
- Media
    - Determine how you want to use visuals to communicate the content *(visual language and visual communication)* and set the mood of the page *(look and feel)*.
    - Determine color palettes for the pages and modules to be created.
    - Draw images to show a visual representation of the text content.
    - Structure animations that capture the dynamics of complex processes.
- Interface layout
    - Determine what you want your users to experience while going through the module *(user experience)*:
        - Determine the technological features to enhance the content: graphics, visualizations, animations, simulations, virtual labs, virtual patients, patient cases, games, etc.
        - Determine what kinds of interactions you want the user to perform on each page, e.g., what the buttons are to click on, how the user will navigate.
        - Assure that the users are on track with what you originally intended and that you are on track in terms of what your users wanted (*Usability* [26]).
    - Establish the design of the layout using guidelines, best practices, usability compliances [15,16]
    - Consider how to present the different kinds of information you plan to have on each page and how each kind of information works as a unified whole (information design [18])
        - Visualization strategies
- Code the module. Assign the task of writing source code to a programmer The programmer will take the storyboards and make the module interactive.
- Perform project administration. Each phase should be documented and archived after they have been completed. Each phase will have files containing content, reference material, images in raw format, source codes, administrative, and management files.

6.11. Prototyping

A rapid prototype is a very rough model (paper or digital) that can be quickly constructed and tested on the users. This testing occurs early in the instructional design process (method described in Rettig [9]). The module design is continually revised until there is confidence in the design. Once a prototype is established, subsequent modules can be developed with assurance in the design.

7. Phase 4: test and evaluate—test modules on your target audience

Phase 4 defines the general areas addressed during a user test. Instruments for measure include focus groups, surveys, interviews, log data, pre-/post-tests, and exam scores. Each area has standard evaluation metrics [5,7,8,27]. We triangulated data from many instruments to assess higher order learning, since that is often more difficult to quantitatively measure.

7.1. User tests

- Usability and interface:
    - Is the module usable?
    - Are there any technical problems that interfere with the learning outcome?
- Content:
    - Is the content easy to understand?
    - Does it convey the complexity of the information?
- Learning effectiveness and attitudes:
    - How did the module help the student learn?

Perform surveys in the following areas: user satisfaction, usage, assessment, and instructor feedback data. Use a five point Liekert rating scale for the indicated areas to measure attitudinal responses: strongly disagree, disagree, undecided, agree, strongly agree.

- User satisfaction data: How well did the students like the module?

This information can be gathered with self-report surveys or by interviews. The following list areas to ask your students about, followed by examples:

- Demographics of your user:
  What is your knowledge of topic X?
  How often do you use a computer?
  What do you use a computer for?
- Technical and usability (Liekert scale)
  I had no technical difficulties with the module.
  Navigation through the module was clear.
  I was able to find the information that I was looking for.
- Usefulness of the module (Liekert scale)
  The content was laid out in a clear and concise fashion.
  Feature ‘A’ in the module was useful in helping me understand ‘X’ concept.
  List features as separate questions: graphics, text, animation, simulation, virtual lab, quizzes, games.
- Usefulness in learning. (Liekert scale)
  The module helped me with conceptual understanding.

The level of content was appropriate.
I liked the way the module was integrated with the class.

- Attitudinal affect (Liekert scale)
  I enjoyed learning ‘X’ with the module.
  Using module ‘X’ motivated me to learn.
- Open ended questions
  Please list 3 difficult concepts in the course.
  What was the most memorable item from the module?
  Suggestions for improvement?

- Usage data: How much did the students use the module?

If the module is distributed over the Internet, you can collect useful information about how much the user used the module. In compliance with the Institutional Review Board’s human subjects protocols, a consent form is required to collect data from your users. A backend tracking system can invisibly collect log file data from the user. This data contains information on which page, what items the user clicked on, and amount of time the user spent on that page. The following is a list of suggested data to collect for each user:

- User ID
- IP address of user
- Last session
- Number of sessions
- Time spent on entire module
- By individual pages: page visited, time spent on each page, items clicked

- Assessment data: How well did the students learn?

This information is collected using testing measures. Design questions that specifically test conceptual understanding and applications of the concepts rather than factual questions.

- Instructor feedback: How well did the students perform with the module?

This information can be gathered with self-report surveys or by interviews (Textbox 1).

- How do you feel your students learned the concept with the module?
- What would you improve for the next module?

8. Phase 5: improve—iterative redesign

Phase 4 will show what needs to be improved in the module. In an iterative design cycle [6], redesigns after user testing are common (back to Phase 3) and decisions for specific redesign areas are based on time, resources, impact, and importance.

Textbox 1
Testimonial. A faculty member used a Virtual Lab Module for their class on renal physiology and responded positively. The design rationale for the module follows.Biology is a very visual subject and it deals with change through time. The new technologies provide powerful tools for dealing with these types of teaching challenges. My favorite example is teaching how urine is processed in the kidney which has always been a very difficult thing to teach and for students to grasp. Once we instituted a virtual laboratory on this subject where students could visualize the processes involved and manipulate them, the kidney was no longer the huge teaching challenge it had been for me in the past. I couldn’t believe it when the teaching assistants reported to me that the students had no questions on the topic!

9. Prospective in interactive media design for imaging informatics

Technology has opened the doorway in informatics to support databases that contain a wealth of information from many disciplines. However, these information systems often have complex interfaces and require training in order for the user to effectively and efficiently gain access to the information. In order to access any information on these databases, a user needs to learn how the interface works and query the correct search to access the material. One way to address this need is to create a training module on how to use these information systems. With the methodology described above, these information systems can be transformed into usable information and the training modules can be disseminated to any user. The training modules can be further enhanced to support the needs of different types of users on the system. For example, a clinician might want to search the organ systems and see how a particular drug may interact with a system physiologically. A researcher may want to search for the specific type of receptors that a particular drug may bind with across different physiologic systems. The design of the interface would be different since the needs of both users are different.

Other educational modules can be developed to train users in specific areas. These training modules can bridge knowledge and expertise and become the interdisciplinary learning material for tomorrow’s student. One such module could be an interactive module on congenital hand anomalies to train clinicians how to analyze different radiographs for making diagnoses and treatments. This module could also be modified to address questions that parents may have about their child's disease and offer coping mechanisms.

The medical and research knowledge bases found in informatics can be harnessed to create effective multimedia learning materials. Already enriched with high-quality graphics and dynamic simulations, only an interface with best practices in instructional design principles and human computer interaction are needed to transform this information from merely an information archive into a rich educational resource.

10. Discussion

Students entering college today comprise a more ethnically, culturally, and linguistically diverse population than at any time in the past 50 years. Yet among this heterogeneous population, a common thread exists: technology. As educators, we should exploit this commonality to improve our teaching, our students' experiences, and the quality of education.

Educational media is a way of harnessing technology and building the future of learning. Unlike PowerPoint or traditional textbooks, educational media incorporates high resolution animations and videos that bring dynamic processes to life. In addition, educational media can incorporate interactivity that stimulates and challenges the learner. As a result, students will find science learning fun and engaging. Educational media can train many students: biology, medicine, bioengineers, and medical informatics, with one common core learning module and then integrate knowledge from the different organ systems databases and relate them to medical conditions, symptoms, and treatment. With efficient and engaging learning tools, we can continue to attract and retain a talented and increasingly diverse pool of science majors.

This paper is an attempt to describe and document methods for building new and creative educational media content. It provides a framework for designing and implementing new content into classrooms. While the details of each project will be tailored to suit each team's needs, the basic phases of understand, design, build, test, and improve will remain relatively unchanged. This framework, far from being untested, was developed by the Virtual Labs project during the course of creating over 500 pages of material to teach physiology. It is from this experience that new designers will hopefully avoid pitfalls and produce quality educational media.

The importance of creating new teaching tools cannot be understated. Educational media is poised to help train new scientists, clinicians, information technologists, and health care providers by creating modules that teach relevant concepts and are sharable and distributable to any student, anywhere, anytime. In order to provide the diversity and depth of materials needed, we must empower teams around the world with the necessary skills to design new educational media. We hope that this paper will serve as an introduction to the methods behind creating new educational media resources and we are taking initiatives to train teams locally and abroad. Such cross-cultural exchanges are taking place with our community schools and our global partners in Sweden and India, with plans to collaborate with China.

Together, we can create the next generation of educational media tools which will train the next generation of scientists, clinicians, and health care workers.

11. Summary

The design of high-quality interactive multimedia for learning requires creators to incorporate best practices in education, instructional technology, and human computer interaction to create a useful and effective online learning environment for students. This paper describes how multimedia can enhance learning and offers best practice guidelines for building innovative multimedia materials. These development strategies were derived from six years of experience with the Virtual Labs Project at Stanford University. These strategies can be applied to new educational media development in imaging informatics and be disseminated with informatics databases. The development of a multimedia module consists of five phases: (1) Understand the learning problem and the users needs; (2) Design the content to harness the enabling technologies; (3) Build multimedia materials with web style standards and human factors principles; (4) User testing; (5) Evaluate and improve design.

References

[1] Alessi SM, Trollip SR. Multimedia for learning: methods and development.: Allyn and Bacon; 2001.

[2] Bransford J, Brown A, Cocking R, editors. How people learn: mind, brain, experience and school. Washington, DC: National Academy Press; 2000. p. 39–66.

[3] Kozma RB. Learning with media. Rev Edu Res 1991;61(2):179–211.

[4] Clark RC, Mayer RE. e-Learning and the science of instruction. New York: Wiley; 2003 p. 322.

[5] Creswell J. In: Laughton C, editor. Research design: qualitative, quantitative, and mixed methods approaches. Thousand Oaks: Sage Publications; 2003. p. 246.

[6] Dick W, Carey L. The systematic design of instruction. New York: Harper Collins Publishing; 1996.

[7] Harvey J, editor. Evaluation cookbook.

[8] Kellogg WK. Kellogg foundation evaluation handbook 1998. Battle Creek, MI.

[9] Rettig M. Prototyping for tiny fingers. Commun ACM 1994;37(4): 21–7.

[10] Flashlight Tool Kit. American Association of Higher Education's Teaching.

[11] Cooper A. About face: the essentials of user interface design. Foster City: IDG Books Worldwide; 1995 p. 580.
[12] Nielsen J. Usability engineering. Indianapolis: Morgan Kaufmann Publishers; 1994.
[13] Nielsen J. Designing web usability: the practice of simplicity.: New Riders Publishing; 1999.
[14] Shneiderman B. Designing the user interface: strategies for effective human–computer interaction. Reading: Addison-Wesley; 1997.
[15] Research-based Web Design and Usability Guidelines.
[16] Lynch PJ, Horton S. Web style guide. New Haven and London: Yale University Press; 2001 p. 223.
[17] Strunk W, White EB. The elements of style.: Pearson PTR; 1995.
[18] Tufte E. The visual display of quantitative information. Cheshire: Graphics Press; 1983.
[19] Tufte E. Envisioning information. Cheshire: Graphics Press; 1990 p. 126.
[20] Tufte E. Visual explanations. Cheshire: Graphics Press; 1997 p. 156.
[21] Wickens CD, Holand JG. Engineering psychology and human performance: attention in perception and display space. 3rd ed. Upper addle River, NJ: Prentice Hall; 2000 p. 350.
[22] Fogg BJ. Persuasive technology: using computers to change what we think and do. San Francisco: Morgan Kaufmann; 2002.
[23] Lepper M, Malone T. Intrinsic motivation and instructional effectiveness in computer-based education. In: Snow R, Farr M, editors. Aptitude, learning and instruction. Cognitive and affective process analyses, vol. 3. Hillsdale, NJ: Lawrence Erlbaum Associates; 1987. p. 256–86.
[24] Huang C. Changing learning with new interactive and media-rich learning environments: virtual labs case study report. JCMIG 2003; 27(2–3):157–64 [Special Issue].
[25] Huang C, Huang H. Interactive instruction of cellular physiology for remote learning. Cell Mol Biol 2003;49(8):9.
[26] Research-based web design and usability guidelines. Volume 2004.
[27] Rogers G. Materials development and the ATE program.: Western Michigan University; 2001.
[28] Hubel DH, Wiesel TN. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. J Physiol 1962;160: 106–54.

Camillan Huang is the Director of the Virtual Labs Project at Stanford University's Medical Media and Information Technologies (SUMMIT) in the School of Medicine and a lecturer in the department of Computer Science at Stanford University. Cammy received her degrees from the University of California, Berkeley: PhD in 1999 in the department of Molecular and Cell Biology and B.A. with majors in Molecular and Cell Biology (emphasis, neurobiology) and Classical Civilizations. Her training includes neuroscience research (neuroimaging and neuroanatomy), traditional and digital arts, animation, project management, educational media development, human–computer interaction, and instructional design. Her current research and teaching interests include use of educational technology for biology and instructional media design for students learning science. She has taught in undergraduate biology classrooms at Stanford University and University of California at Berkeley, and in interdisciplinary courses: Art and Vision at the Alliance for Lifelong Learning and Interactive Media Design for Kids Learning Biology at Stanford University. She is also currently an academic advisor and collaborates on many outreach projects including H.E.L.P. for Kids to design computer-based teaching media for kids (grades 4–12) learning health education.

DTD 5
ARTICLE IN PRESS

ELSEVIER

Computerized Medical Imaging and Graphics xx (2005) 1–3

Computerized Medical Imaging and Graphics

www.elsevier.com/locate/compmedimag

Editorial

Medical imaging informatics research and development trends—an editorial

Medical informatics, which studies healthcare information, has gradually evolved and established itself as a rigorous scientific discipline during the past 15 years [1–7]. There are now at least 10 major training programs supported by the National Library of Medicine (NLM) in the US with many more throughout the world, teaching students, the concepts and methods of health information gathering, data structure, information extraction, retrieval and distribution, and knowledge representation [8–14].

Medical Imaging Informatics is a subset of medical informatics that studies image/data information acquisition; processing; manipulation; storage; transmission; security; management; distribution; visualization; image-aided detection, diagnosis, surgery, and therapy; as well as knowledge discovery from large-scale biomedical image/data sets. Although, it is based on many existing concepts, theories, terminology, and methodology derived from medical informatics, Imaging Informatics deals with different types of data including multi-dimensional medical images, graphics, waveforms, and text. Accordingly, imaging informatics requires new concepts and new tool sets to handle these types of data [15]. Although there exist training programs in medical imaging and telemedicine [16–18], focused training dedicated to imaging informatics is limited because of its novelty [19].

Recent advances in medical imaging technology, e.g. Picture Archive and Communication systems (PACS), image-guided surgery and therapy, computer-aided diagnosis (CAD), electronic Patient Record (ePR) with image distribution have propelled imaging informatics as a discipline to manage and synthesize knowledge from medical images for effective and efficient patient care as well as outcomes [20].

This Special Issue in Medical Imaging Informatics consists of 13 papers spanning the Imaging Informatics field. The first paper is an editorial on 'Medical Imaging Informatics Research and Development Trends'. The subsequent 12 papers are divided into five groups as follows.

1. Group 1: imaging informatics technology

Two important imaging informatics technologies are image storage and distribution, and image communication. Two papers are included in this group.

Liu et al.—Utilizing data grid architecture for the backup and recovery of clinical image data
Chan et al.—International Internet 2 performance and automatic tuning protocol for medical imaging applications.

Grid computing is the most exciting incarnation of contemporary computing technologies including parallel, peer-to-peer, and client–server models. It has been used mostly in physical and engineering applications. Liu et al. present a model using the Data Grid for the backup and recovery of clinical image data—one of the most difficult problems facing large-scale clinical image databases today.

Internet 2 is the high-speed communication backbones established by the education and research community during the past 5 years through the support of the US National Science Foundation (NSF). Internet 2 applications have been mostly confined to the US. Chan et al. describe their experience of connecting the Hong Kong Polytechnic University networks to the Image Processing and Informatics (IPI) Laboratory, University of Southern California (USC) through Internet 2 with medical image communication as the application.

2. Group 2: electronic patient record (ePR) with image distribution

Electronic patient record (ePR) with image distribution is a current R&D topic in imaging informatics. Traditional customized small-scale ePR systems without images have been developed for over 10 years, however, ePR systems with image distribution is relatively novel derived from clinical needs and the successful implementation of many Picture Archiving and Communication systems

0895-6111/$ - see front matter © 2004 Published by Elsevier Ltd.
doi:10.1016/j.compmedimag.2004.09.003

(PAC systems) around the world. Three papers are included in this group:

Zhang et al.—Web-based ePR for collaborative medical applications
Law—A model of DICOM-based ePR in radiation therapy
Cheung et al.—Integrating images into the ePR of the hospital authority of Hong Kong.

Zhang et al. describe a web-based ePR system for collaborative consultation with medical image/data in a teleradiology and collaborative consultation application during the Severe Acute Respiratory Syndrome (SARS) period, May 2003, in China. This paper represents an integration of four current information technologies: Web, ePR, DICOM (Digital Imaging and Communication in Medicine) standard, and collaborative consultation.

DICOM-based medical image standard has been successfully implemented in diagnostic radiology during the past 10 years. However, despite the introduction of DICOM radiation therapy (RT) objects since 1999, the RT community has not yet to take advantage of the power and infrastructure of imaging informatics. Law presents an ePR model of DICOM-based radiation therapy information system, where seven RT DICOM objects are included, as a first attempt at such a development to utilize these concepts.

In the past, ePR has been developed as a small-scale customized electronic patient based system. Cheung et al.'s paper describes the planning and preliminary results of an enterprise level ePR with image distribution in Hong Kong. The ePR system involves 43 hospitals (93% of the Hong Kong market) with a total of 29,000 beds. The ePR system is based on the existing Clinical Management System (CMS) developed in-house at the Hong Kong Hospital Authority which currently contains 6.4 million patients' records.

3. Group 3: image-aided detection and diagnosis

One reason for developing the medical imaging infrastructure is to take advantage of tools within the infrastructure for large-scale longitudinal and horizontal clinical service and research, as well as systematic education and training. Among these tools are image processing, visualization, image matching, content-based retrieval, data mining, and computer-assisted detection and diagnosis. Four papers are included in this group:

Lehmann et al.—Automatic categorization of medical images for content-based retrieval and data mining
Pietka et al.—Informatics infrastructure of CAD system
Long et al.—Image Informatics at a National Research Center
H. Huang et al.—Imaging matching as a diagnostic support tool for brain diseases in children.

Automatic categorization is the first step for image content-based retrieval and data mining. Lehmann et al. present a method to automatically categorize medical images into 81 classes using 10,000 clinical images with the highest accuracy compared with other existing categorization systems.

Pietka et al. describe an informatics infrastructure for large-scale computer-aided diagnosis (CAD) system for bone age assessment of children based on a digital atlas containing over 1000 normal hand digital radiographs of four different ethnic origins and genders.

Among all Institutes, Centers, and Libraries under the umbrella of NIH (National Institutes of Health), the National Library of Medicine (NLM) is probably the most active in intramural informatics and imaging informatics research. Long et al. describe the activities with collaboration around the world at the Lister Hill National Center for Biomedical Communications, NLM.

Imaging matching is an important research area in imaging informatics. H. Huang et al. present a novel diagnostic support tool based on image matching using an image database containing 2500 pediatric MR brain images.

4. Group 4: surgical simulation and interactive multimedia learning

Imaging informatics is a very broad field using many tools and technologies, which may or may not be developed originally for medical imaging applications. Two papers are included in this group. Although, they may not fall exactly under the realm of imaging informatics, the outcomes of using imaging and technologies to achieve their goals are very similar to that of imaging informatics. In addition, some of the tools and technologies used have great potential of being adopted for imaging informatics applications and expanding the realm of Imaging Informatics.

Montgomery et al.—User interface paradigms for patient-specific surgical planning
C. Huang—Designing high-quality interactive multimedia learning modules.

In surgery, the first innovations of using image-aided tools are 3D display and image-based surgical planning. The current research and development trend is in surgical simulation. Although contents in Montgomery et al.'s paper did not explicitly mention imaging informatics, readers can no doubt identify many terms and technologies in the paper that overlap with those used in imaging informatics.

In interactive multimedia learning, images, graphic, drawings, video clips are among those being used as inputs to the learning tools. C. Huang's paper describes a methodology currently being used for building learning modules. In the paper, the author also points out the similarity of technologies used in imaging informatics and in interactive multimedia learning.

5. Group 5: HIPAA compliance

Imaging informatics is dealing with clinical images that are acquired, stored, transmitted through public networks, and displayed. During these processes, the image integrity could have been compromised at any stage. HIPAA (Health Insurance Portability and Accountability Act) is a mandate for medical image user compliance. However, HIPAA only tells the user to comply, but does not provide protocols to be followed. Zhou et al.'s paper 'HIPAA Compliant Auditing System for Medical Images' describes a framework for HIPAA compliance using a workflow-auditing paradigm.

In summary, the 12 papers included in this Special Issue in Medical Imaging Informatics covers five groups of papers: state-of-the-art technologies ePR with images; image-aided detection and diagnosis; surgical simulation and interactive multimedia learning; and HIPAA compliance. Together, they provide a glimpse of current research and development (R&D) trends in this field. It is hoped that this issue will stimulate further R&D in imaging informatics to benefit the steadily fast growing medical imaging community.

Reference

[1] Johns ML. The development of a graduate program in health information management. Methods Inf Med 1994;33:278–81.
[2] Graves JR, Amos LK, Huether S, Lange LL, Thompson CB. Description of a graduate program in clinical nursing informatics. Comput Nurs 1995;13(2):60–70.
[3] Shortliffe EH. Medical informatics training at Stanford University School of Medicine Yearbook of medical informatics 1995 p. 6.
[4] Shortliffe EH, Patel VL, Cimino JJ, Barnett GO, Greenes RA. A study of collaboration among medical informatics research laboratories. Artif Intell Med 1998;12:97–123.
[5] Musen MA. Stanford medical informatics: uncommon research, common goals. MD Comput 1999;16(1):47–8.
[6] Gell G, Errath M, Simonic KM. Research at the Department of Medical Informatics, Statistics and Documentation of the University of Graz. Yearbook of medical informatics 2000: research and education. vol. 11 2000 p. 1114–9.
[7] Leven FJ, Haux R. Twenty-five years of medical informatics education at Heildelberg/Heilbronn: discussion of a specialized curriculum for medical informatics. Yearbook of medical informatics 2000: research and education. vol. 11 2000 p. 120–7.
[8] Greenes RA, Shortliffe EH. Medical informatics: an emerging academic discipline and institutional priority. JAMA 1990;263(8): 1114–20.
[9] Heathfield HA, Wyatt J. Medical informatics: hiding our light under a bushel, or the Emperor's new clothes? Methods Inf Med 1993;32(2): 181–2.
[10] Van der Maas A, Johannes ten Hoopen A, Ter Hofstede A. Progress with formalization in medical informatics? J Am Med Inform 2001; 8(2):126–30.
[11] Warner HR. Medical informatics: a real discipline? J Am Med Inform Assoc 1995;2(4):207–14.
[12] Staggers N, Gassert CA, Skiba DJ. Health professionals' views of informatics education: findings from the AMIA spring conference. J Am Med Inform 2000;7(6):550–8.
[13] Hasman A, Talmon JL. Education and research at the Department of Medical Informatics Masstricht. Yearbook of medical informatics 2000: research and education. vol. 11 2000 p. 100–6.
[14] Warner HR. Medical informatics: a real discipline? J Am Med Inform Assoc 1995;2(4):207–14.
[15] Huang HK, Wong STC, Pietka E. Medical image informatics infrastructure design and applications. Med Inform 1997;22(4): 279–89.
[16] McCray AT. In: Medical informatics research and training at the Lister Hill National Center for Biomedical Communications. In: Yearbook of medical informatics 2000: research and education, vol. 11, 2000. p. 95–9 p. 95–9.
[17] Fuller S, Kalet I, Tarczy-Hornoch P. Biomedical and health informatics research and education at the University of Washington. Yearbook of Medical Informatics 2000: Research and Education. vol. 11 2000 p. 107–13.
[18] Patton GA, Gardner RM. Medical informatics education: the University of Utah experience. J Am Med Inform 1999;6(6):457–65.
[19] Kangarloo H, Huang HK. NIH/NLM, IT15LM07356 'Training program for imaging based medical informatics'; 2002.
[20] Huang HK. PACS and imaging informatics: principles and applications. Hoboken, NJ: Wiley; 2004 p. 704.

H.K. Huang*
Division of Imaging Informatics, Department of Radiology,
The Keck School of Medicine,
Information Sciences Institute,
University of Southern California, 4676 Admiralty Way,
Suite 601, Marina del Ray, CA 90292, USA
E-mail address: hkhuang@aol.com

Received 28 August 2004
Revised 28 August 2004
Accepted 30 September 2004

*Tel.: +1 310 448 9435; fax: +1 310 448 9441.

ELSEVIER

Computerized Medical Imaging and Graphics xx (xxxx) 1–8

Computerized Medical Imaging and Graphics

www.elsevier.com/locate/compmedimag

Image-matching as a medical diagnostic support tool (DST) for brain diseases in children

H.K. Huang[a,*], J.F. Nielsen[b], Marvin D. Nelson[b], Lifeng Liu[c]

[a] *Image Processing and Informatics Lab, Information Science Institute/University of Southern California, 4676 Admiralty Way, Suite 601, Marina del Rey, CA 90292, USA*

[b] *Department of Radiology, Childrens Hospital Los Angeles/University of Southern California, 4650 Sunset Blvd, MS 81, Los Angeles, CA 90027, USA*

[c] *MD Online, Inc., 99 Hayden Ave, Lexington, MA 02421, USA*

Received 19 April 2004; revised 19 August 2004; accepted 30 September 2004

Abstract

Imaging-matching is an important research area in imaging informatics. We have developed and evaluated a novel diagnostic support tool (DST) based on medical image matching using MR brain images. The approach consists of two steps, database generation and image matching. The database contains pre-diagnosed MR brain images. As the images are added to the database, they are registered to the 3D Talairach coordinate system. In addition, regions of interests (ROI) are generated, and image-processing techniques are used to extract relevant image parameters related to the brain and diseases from the ROIs and from the entire MR image. The second step is to retrieve relevant information from the database by performing image matching. In this step, the physician first submits a query image. The DST computes the similarity between the query image and each of the images in the database, and then presents the most similar images to the user. Since the database contains pre-diagnosed images, the retrieved cases tend to contain relevant diagnostic information.

To evaluate the usefulness of the DST in a clinical setting, pediatric brain diseases were used. The database contains 2500 pediatric patients between ages 0 and 18 with brain Magnetic Resonance (MR) images of known brain lesions. A testbed was established at the Childrens Hospital Los Angeles (CHLA) for acquiring MR images from the PACS server of patients with known lesions. These images were matched against those in the DST pediatric brain MR database. An expert pediatric neuroradiologist evaluated the matched results. We found that in most cases, the image-matching method was able to quickly retrieve images with relevant diagnostic content. The evaluation method and results are given.

Keywords: Image-matching; Neuroanatomy; Medical image databases; Diagnostic support

1. Concept of image matching

The use of computers in disease detection and diagnosis has gained increased popularity in recent years. Several commercial Computer Aided Detection (CAD) and Computer Aided Diagnosis (CADx) systems have by now received FDA approval for detection and diagnosis of lung nodules and breast tumors [1–3]. These systems are used to analyze a single image at a time, using computer algorithms that are trained to detect and diagnose specific diseases. Their performance is sufficient to employ these systems as ‘second readers’ that assist the radiologist in making a decision involving chest CT images or mammograms. However, besides lung diseases and breast tumors, the use of computers in diagnostic support systems has been limited, and radiologists still rely almost exclusively on personal experience and memory in their decision-making process. In this paper we describe a recently developed computer-based diagnostic support system based on image matching that works differently from existing CAD/CADx systems [4–7]. Rather than analyzing a single image at a time, image matching makes use of a large database of pre-diagnosed images.

* Corresponding author. Tel.: +1 310 448 9435; fax: +1 310 448 9441.
E-mail address: hkhuang@aol.com (H.K. Huang).

0895-6111/$ - see front matter © 2004 Published by Elsevier Ltd.
doi:10.1016/j.compmedimag.2004.09.008

Image Match is a branch of computer-aided diagnosis (CAD) that is a system permitting a user with access to:

(a) Large database of certain category of medical images with extracted parameters,
(b) High-speed broadband networks like the Internet 2 connection to submit an image as a query, and
(c) Receive in return a set of images from the database of those already diagnosed images that are most similar and relevant. The degrees of similarity and relevance are in terms of the extract parameters.

An Image Matching system utilizes sophisticated image processing algorithms, combined with robust database technology and state-of-the-art web-enabled hardware and software, and packages these capabilities into a tightly integrated workflow environment in the framework of medical image informatics infrastructure [8].

The idea to use image matching to retrieve clinically relevant information from image databases is not new. For more than two decades, numerous groups have researched methods for Image Retrieval in large image databases. An excellent review of general methods for Image Retrieval from any image database (not just medical image databases) is given in Ref. [9]. Two of the best-known general methods are IBM's QBIC system [10] and the Virage Image Retrieval system [11]. Previous implementations typically attempt to provide an automatic, gross description of the entire image, and then retrieve images from a database based on this gross description. However, this is of limited use in medical imaging, since the relevant information contained in a diagnostic image is often quite subtle, and can be highly localized. Furthermore, in medical images, it is often impractical or even impossible to automatically segment the pathology-bearing region (PBR). It has therefore been recognized that Image Retrieval in medical image databases should directly involve a human expert, leading to a 'human-in-the-loop' approach [12].

2. Methodology

2.1. Data collection of the image matching database

Let us use brain MR image database as an example to illustrate the method of data collection In order for an image set in the database to be meaningful, several types of data in addition to the image set are required; these include

- Relevant patient information
- Image data set in PACS DICOM format
- Corresponding radiological reports
- Histopathology reports, if available, and
- Any other relevant reports.

When a new case is collected into the database, various pre-processing steps are performed. The images are *registered* to a common global coordinate system as defined by all other images in the database. This *alignment* step ensures the continued consistency of the database as additional images are input. The pathology in each database image is then localized and highlighted with a region of interest (ROI) in *a segmentation step* in order to focus the matching algorithm on the visual cues corresponding to the given pathology. After the registration, alignment, and segmentation steps, a select set of *features* is extracted from the imagery using image processing. The features are chosen to be robust to slight misalignments and invariant to various imaging artifacts while capturing the salient information contained within the imagery and highlighting the indications of pathology. Features can include the location, size, shape, texture, denseness, and others. These features are input into the database appended to the image set, and will be used as keys to the Image Match algorithm(s) [13,14]. Fig. 1 shows the procedure.

2.2. Image registration and segmentation

The concept of Image Matching is measurement of the similarity of the submitted images with images already in

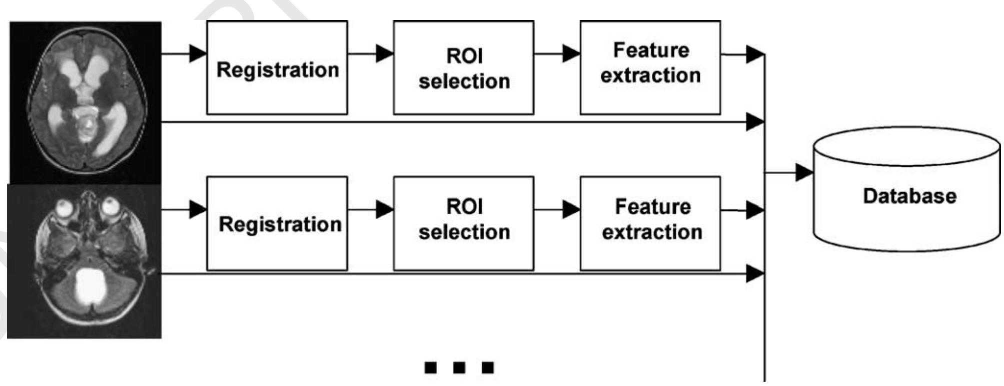


Fig. 1. Populating the database with already diagnosed images. Each new image is first registered to a global coordinate system. The seed pixel is then selected, and the region of interest (ROI) is generated. Finally, image processing is used to extract several select image parameters, which will be used to perform the image matching. The database is populated with these parameters, along with the original (unprocessed) images.

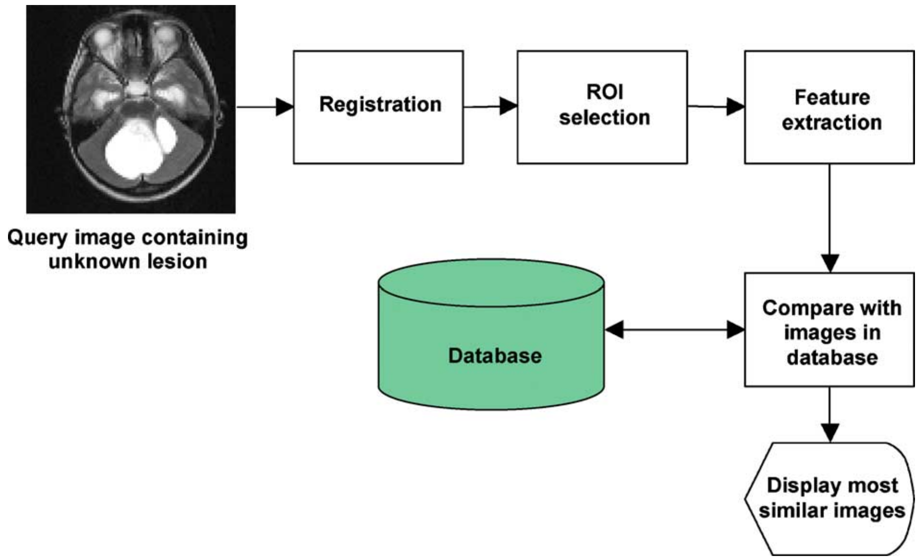


Fig. 2. Matching an image ('query image') with an unknown lesion to images stored in the database. The query image is first registered to the same global coordinate system as that used to populate the database. Seed pixel selection can be done either manually by the user or automatically. The same set of image parameters computed for each of the images in the database is then computed for the query image. These parameters are used to compute the similarity between the query image and each of the images in the database.

the database Similarity can be determined by those parameters extracted from each image set in the database. These parameters are characteristics of different applications and imaging modality used. To obtain these parameters requires advanced image processing algorithms for data extraction. To measure the similarity between the submitted image with images in the database requires the access to this large image database and rapid measurement of these predetermined parameters, and return similar images as queried by the user.

Two key organizational factors for the images in the database are registration and segmentation.

- Registration is the key aspect of the organization of the database, and it ensures that only corresponding regions of images are compared during a match, so as to save the computation time and reduce the possibility of false matches. A registered database of medical images provides tremendous information about statistical properties of regions of the anatomy in combination with various pathologies.
- Segmentation, or labeling of the images in the database, provides a means of focusing the attention of the algorithms on a particular region of interest. The attention mechanism eliminates outside distracters that can contribute to a false match.

When measuring the similarity between two images, it is of utmost importance that the images themselves be at least somewhat aligned, to ensure that certain anatomical structures from one image are compared with the corresponding structures in the other. However, any two medical images, in general, will not be in alignment, because they may be acquired at different times, on different subjects, or in different scanners. Accurate comparison of grossly misaligned medical images is almost impossible. Thus, sophisticated registration algorithms are incorporated to roughly align all images into a common global coordinate system, in combination with comparison measures that are robust to the slight, residual misalignment that remains after the registration is performed.

When a query is submitted, the images must also be aligned to the database global coordinate system before matching is performed. Without such alignment, subtle pathologies apparent in the query image may not be matched to the correct region of a database image showing signs of a similar pathology. The same set of features previously extracted from each of the reference images in the database is extracted from the query image in a similar manner. The features are then matched to every relevant image in the database, the match scores are compared and sorted, and the best matches are reported to the query, along with additional information, such as the diagnoses and findings, associated with each matching case [4]. Fig. 2 depicts how to submit a query image for comparison with images in the database.

3. An example of MRI brain image matching with images in the database

A well-organized medical image database can offer much useful related medical information, such as high probability distribution of one particular disease related to the anatomy location, age, gender, nation and so on. Figs. 3 and 4 show the similarity in locations of two brains diseases:

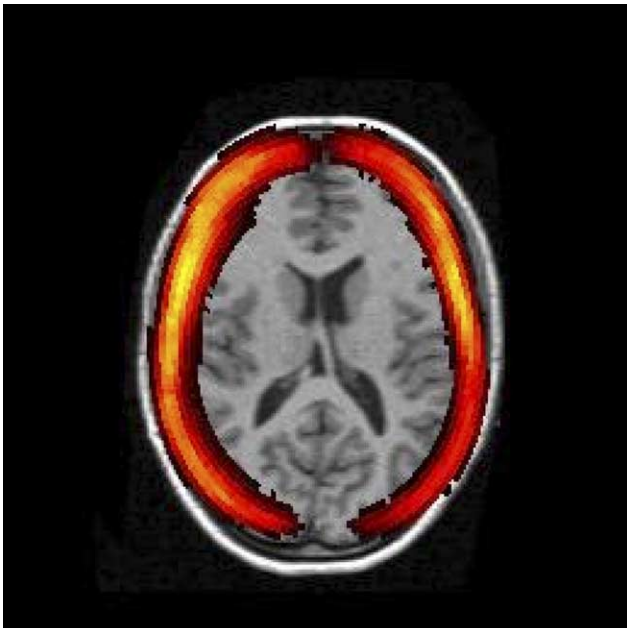


Fig. 3. Distribution of Subdural Hematoma in a database of 107 cases Bright yellow (B and W: two lighter gray lateral rings): Highest probability [4].

Subdural Hematoma and Meningioma in an existing MRI brain database. For each disease, a low probability threshold is first applied on the distribution map (combination of parameters of similarity) of similar disease in the database, the anatomical locations of the distribution map (in color) is then overlaid on the middle slice of an MR T1 image (in gray scale). In these figures, the color scale from dark-red to red to white-yellow of the distribution map means higher occurrence chance of the disease.

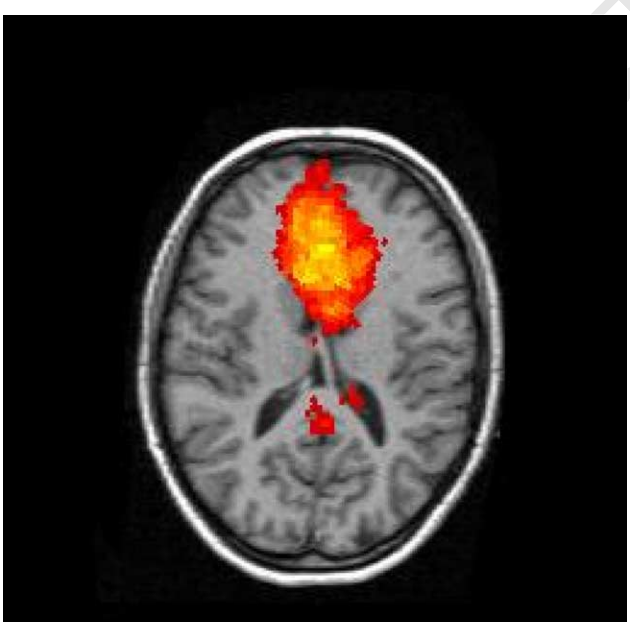


Fig. 4. Distribution of Meningioma in a database of 380 cases Bright yellow (B and W: midline lighter gray area): highest probability [4].

In the case of Subdural Hematoma (Fig. 3), 107 cases were used. After the Image Matching, it is obvious that this tumor is distributed as two symmetric half rings near the skull. In Meningioma (Fig. 4), 380 cases were used. The tumor is mainly distributed in the middle of frontal lobe.

4. Image-matching as a diagnostic support tool (DST) for brain diseases in children

Let us consider another example of submitting a pediatric MRI with brain disease to match images in a pediatric MR image database [15].

4.1. Methods

In this example, we describe an image-matching method as a diagnostic support tool (DST) for brain diseases in children in our laboratory. The first step is to assemble the database of already diagnosed brain MR images. Fig. 1 illustrates the steps for constructing the database. The Talairach brain atlas was used to define the global coordinate system [16]. As new, already diagnosed images were added to the database, the slice position of each image was determined by first segmenting the various anatomical structures of the brain, such as white matter, gray matter, and basal-ganglia. An anterior-posterior asymmetric, elliptical multiloop polar tile coordinate system was then fit to each image [6]. Finally, a mathematical model relating slice position to image tile features was used to roughly estimate the slice position for each image with respect to the global coordinate system. These registration steps ensure that a given pixel position in one image corresponds to positions in each of the other images in the database that correspond to the same anatomical feature. In this way, a database of registered, already diagnosed images containing data for 2500 children with a variety of brain diseases was assembled. The database contains a complete MR study for each patient, along with a patient radiology report summarizing the diagnosis.

Once the images have been registered to the global coordinate system, a seed pixel within the pathology-bearing region (PBR) in each image was selected manually by an expert clinician. From the seed pixel, a region of interest (ROI) for each image was then generated automatically by the database software. Finally, feature extraction techniques were used to extract a select set of pixel-based parameters from each image in the database. These parameters include geometric and texture features, as well as Fourier coefficients. The ROI was given extra 'weight' when computing this parameter set. This set of parameters is used by the image-matching algorithm to compute the similarity between an image with an unknown pathology ('query image') and each of the already diagnosed images in the database. Similarity is defined as the Euclidean distance between two sets of parameters. Note that the similarity

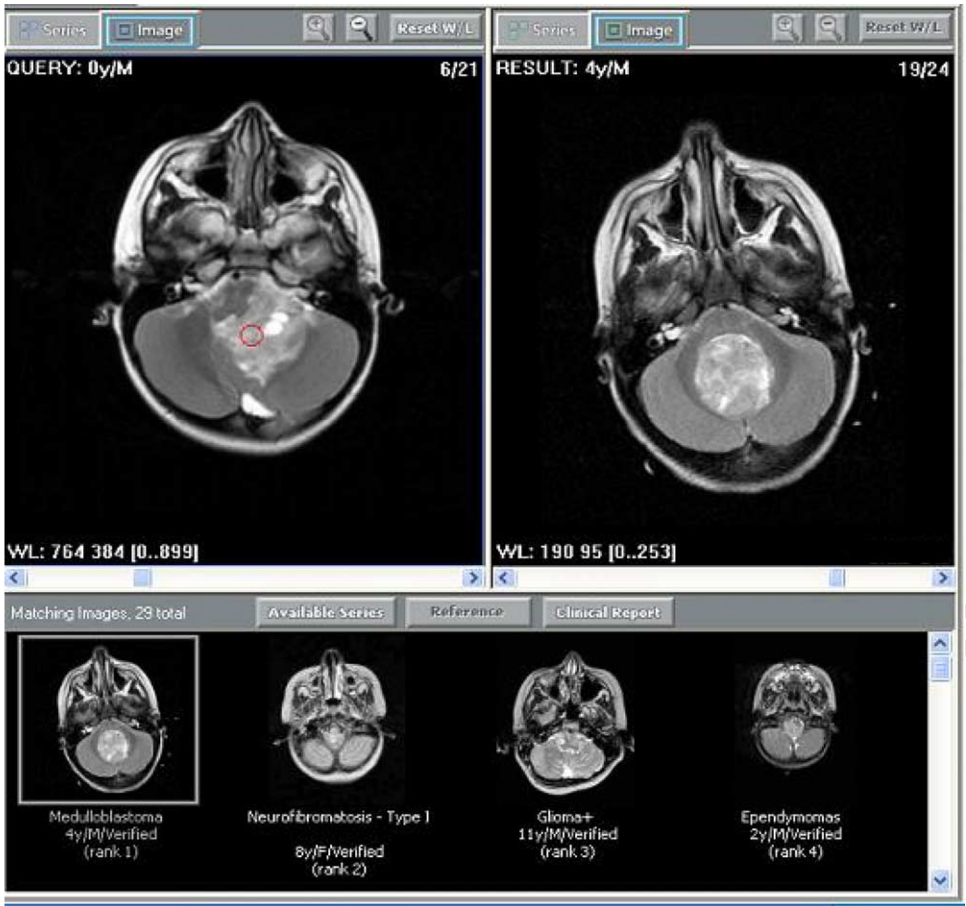


Fig. 5. Graphical user interface (GUI) for matching a query image (top left) with already diagnosed images in the database. The query image in the figure is a brain MR scan of a child with an ependymoma. For the query shown, the user selected a seed pixel, and the program generates the region of interest (ROI) indicated by the open circle. The bottom panel shows the four images from the database that best match the query image. A close-up of the image that best matches the query image (rank 1) is also shown (top right) [5].

measurements used are robust to the slight residual misalignment that remains even after the images have been registered.

4.2. Submitting an image for matching

Fig. 2 illustrates the image-matching process. This example is a comparison of an image with unknown pathology in a brain MR image (the figure shows an unknown lesion) with the already diagnosed images in the database. The initial steps of this process are similar to those taken when forming the database with a new image (see Fig. 1). The query image is first registered to the Talairach atlas. The user then marks a seed pixel in the image. From the seed pixel, the software then generates a geometric shape that represents the ROI. Next, the same set of image parameters previously extracted from each of the already diagnosed images in the database is computed. This parameter set is then used to compute the similarity between the query image and each of the images in the database. The best matches are then retrieved and presented to the user.

Fig. 5 shows part of the graphical user interface (GUI) used to perform image matching in this study. The user submits a brain MR image as a query image. The user then manually selects a seed pixel, and the software generates an ROI (open red circle). The program then searches the already diagnosed image database for images that contain a pathology best matching the ROI of the query image. The retrieved results are shown in the bottom panel in the figure.

4.3. Evaluation

To evaluate the performance of this image matching method as a diagnostic support tool, an expert pediatric neuroradiologist selected brain MR studies from 7 pediatric patients at Childrens Hospital Los Angeles (CHLA). These cases represent a variety of brain lesions with radiographic appearance ranging from localized lesions to more distributed white- and gray-matter diseases. For each patient, the neuroradiologist first consulted the patient pathology report describing the known diagnosis. The neuroradiologist then selected the image from the full set of images of the patient that best represented the pathology, as well as a seed

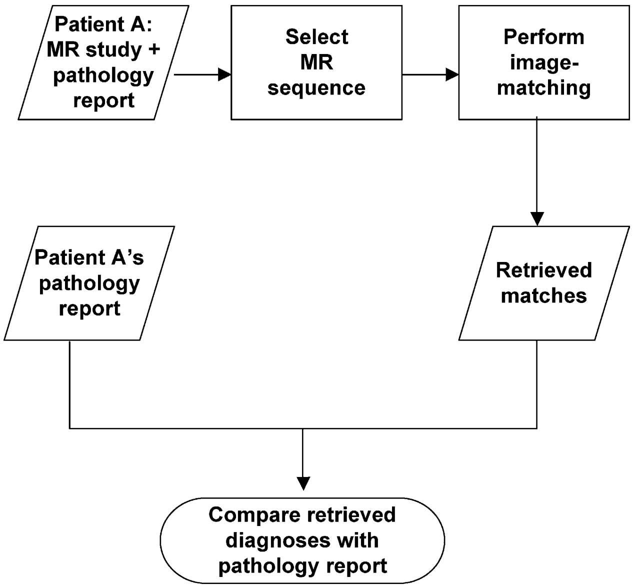


Fig. 6. Overview of the evaluation process. The image that best represents the disease is first selected from the MR image set being studied. This image becomes the input (query image) into the image-matching process (as described in Fig. 5). The retrieved diagnoses are compared with the known diagnosis from the patient's pathology report.

pixel within the PBR. This image served as the 'unknown' query image. Image matching was then performed in the way described in Figs. 2 and 5. In this evaluation study, the database server was located at a remote site (MD Online, Lexington, MA). The interaction with this server was done using a PC computer at CHLA with software for accessing the server over the Internet. The images in the server database that best matched the query image were retrieved from the already diagnosed database, and the diagnostic results corresponding to each image in the database were recorded. In addition, each retrieved case was given a rank according to the degree of similarity with the query image. No information other than the MR images was used to perform the image matching, that is, no additional patient information (such as DICOM header information) was used. After the search results were received, the retrieved diagnoses corresponding to the best 12 matches were then compared with the known diagnosis from each patient's pathology report. Note that this evaluation is a strict test of the diagnostic content of the retrieved results, and does not involve a subjective measurement by the user of the apparent similarity between the query image and the retrieved images. The test procedure is summarized in Fig. 6.

The results of using image matching as a support tool for diagnosing pediatric brain diseases are shown in Table 1. In each case, the total search time was less than 2 s. We see from Table 1 that in five of the seven query cases the correct diagnosis was indeed retrieved from the database. Furthermore, among these 5 cases, between 10 and 60% of the search results either produced the known, true diagnosis, or

Table 1
Summary of test results

Patient	MR sequence type	Disease	# Matches evaluated	% Correct
A	T2	Ependymoma	12	25
B	T2	Neurofibromatosis	10	60
C	T2	Diffuse cerebral malformation	12	0
D	T2	Arachnoid cyst	12	10
E	T1	Heterotopic gray matter	12	50
F	PD	Arterial venous malformation	8	0
G	T2	Heterotopic gray matter	12	50

The total search time in each case was less than 2 s. The table lists the results for each of the 7 cases tested (A–G). For each patient, the table lists the MR sequence used; the disease; the number of cases that were evaluated by the expert neuroradiologist; and the percentage (to nearest 5%) of the retrieved results that contained a relevant diagnosis for each disease. For each patient, the top 12 retrieved matches were evaluated. For patients B and F, only 10 and 8 matches were returned by the database software, respectively. For example, for Patient A, 3 of the top 12 matches (or 25%) contained either (i) the correct (known) diagnosis, or (ii) a diagnosis that was judged to be a possible alternate (differential) diagnosis [15].

diagnoses that were judged by the expert neuroradiologist to represent possible alternative (differential) diagnoses for each disease. In other words, up to 60% of the match results contain relevant and potentially useful diagnostic information. Furthermore, it is reassuring to note that the accuracy of the retrieved results for patients E and G, who both suffer from heterotopic gray matter, is comparable, even when tested on different MR sequences (T1 vs. T2). For patients C and F, who exhibited diffuse cerebral malformation and arterial venous malformation, respectively, the retrieved diagnoses were not considered to represent possible alternative diagnoses for the disease in question.

It is noted that this match method is for providing a probability of the occurrence of neuro diseases based on the features extracted as a second opinion for the radiologist to consider. The successful rate is measured on the percentage of correct matching.

5. Summary

Image matching is a new tool of diagnostic support system. This method allows the user to quickly search a large image database designed to provide diagnostic support for given diseases. In light of this evaluation of the method for diagnosis of pediatric brain diseases, it can be concluded that image matching can be effective in assisting the physician in surveying possible differential diagnoses for a range of diseases. Further more, these results suggest that the accuracy of the search results may vary for different diseases. It should be emphasized, however, that the poor performance for cases C and F in Table 1 does not necessarily mean that

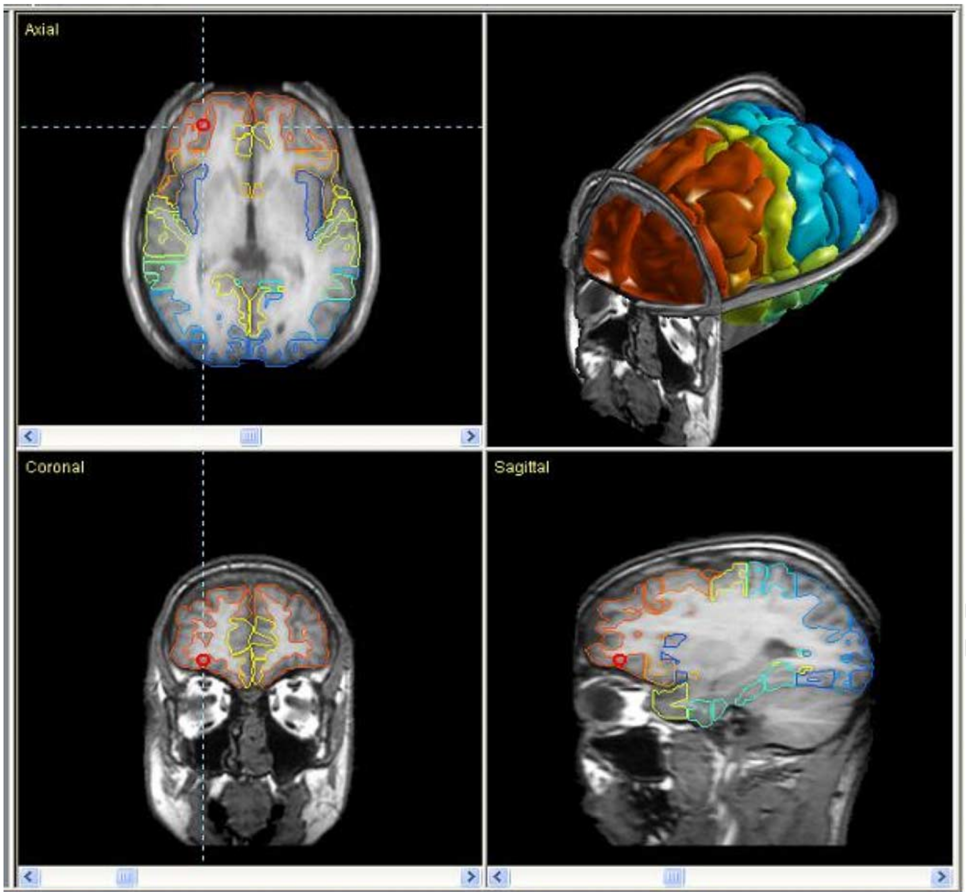


Fig. 7. A GUI for selecting an ROI in a 3D brain atlas. The cross-hair in the axial view indicates the user-selected point. A 3D view is shown in the top right panel, along with the 2D axial, coronal, and sagittal planes containing the user-selected point [5].

the retrieved images in these two cases appear different (as viewed by a human observer) from the query images, since the image matching only retrieved results on the basis of their diagnostic content, and not on appearance. We believe the image-matching performance for such malformations will improve as more cases are added to the database of already diagnosed 3images.

Image matching as a diagnostic support tool is most useful for diseases that are difficult to diagnose and for which many differential diagnoses are possible, as for diseases in other regions like the liver and lung. Also, this image-matching method could be a useful tool for general radiologists, and for radiologists working in smaller hospitals with few colleagues to consult with.

In addition to performing image matching, there is another interesting and useful way to use the large database technology. As described above, all already diagnosed images in the database are registered to the Talairach brain atlas. This means that it is possible for the user to select an ROI in a three-dimensional (3D) model of the brain, and then retrieve the cases from the database for which the pathology-bearing region(s) overlap4 with the user-selected ROI. This, in essence, is the image content indexing, another important research topic in imaging informatics. Fig. 7 shows a GUI that can be used for selecting a single point in a 3D brain volume. Once the ROI is selected, the user can search the already diagnosed database and retrieve the cases that contain a pathology relating to the selected ROI (not shown). As such, this becomes a powerful tool for brain disease education for students and residents, as well as a research tool for exploring the statistical properties of brain diseases.

The image matching method is a general technique for information retrieval from medical image databases. Diagnostic support is only one of the several possible applications for this technology. For example, an intelligent search method could enable ‘data mining’ or ‘knowledge discovery’ tools based on large medical image database. This, in turn, holds the promise of tapping into an enormous reservoir of ‘hidden’ knowledge in the large medical databases that already exist today.

Acknowledgements

This research was partially supported by NIH Grants R01EB00298, R01LM07606, and US Army Medical R&D contract DAMD 17-99-P-3732.

References

[1] Giger ML, Karssemeijer N, Armato SG. Computer-aided diagnosis in medical imaging. IEEE Trans Med Imag 2001;20:1205–8.
[2] Van Ginneken B, ter Haar Romeny BM, Viergever MA. Computer-aided diagnosis in chest radiography: a survey. IEEE Trans Med Imag 2001;20:1228–41.
[3] Giger ML, Huo Z, Kupinski MA, Vyborny CJ. Computer-aided diagnosis in mammography. In: Sonka M, Fitxpatric MJ, editors. Handbook of medical imaging. Medical image processing and analysis, vol. 2. SPIE; 1999. p. 249–72.
[4] Leventon ME, Zhang M, Liu L, Huang HK. CAD: Beyond breast care. Adv Imag Oncol Administr 2002;12:69–72.
[5] Nielsen JF, Nelson MD, Cao F, Liu L, Huang HK. Image-matching as a diagnostic support tool. Proc SPIE Med Imag 2003;5033:19.
[6] Zhang M, Liu L, Leventon ME. A global coordinate system based on large training sets for brain image registration. Radiology 2002;225: P762.
[7] Nielsen JF, Nelson MD, Cao F, Liu L, Huang HK. Pediatric MR image database for teaching neuro anatomy and diseases. RSNA; 2002 InfoRAD exhibit.
[8] Huang HK. PACS and imaging informatics: principles and applications. Hoboken, NJ: Wiley; 2004 p. 704.
[9] Rui Y, Huang TS, Chang SF. Image retrieval: current techniques, promising directions, and open issues. J Vis Commun Image Representation 1999;10:39–62.
[10] Flickner M, Sawhney H, Niblack W, Ashley J, Huang Q, Dom B, et al. Query by image and video content: the QBIC system. IEEE Comput Sept 1995;23–32.
[11] Bach J, Fuller C, Gupta A, Hampapur A, Horowitz B, Humphrey R, Jain R, Shu C. The virage image search engine: an open framework for image management. Proceedings of SPIE storage and retrieval for image and video databases; 1996.
[12] Shyu CR, Brodley CE, Kak AC, Kosaka A, Aisen AM, Broderick LS. ASSERT: a physician-in-the-loop content-based retrieval system for HRCT image databases. Comput Vis Image Understand 1999;75: 111–32.
[13] MD Online, Inc., Lexington, MA 02421, www.mdol.com.
[14] M.E. Leventon, Statistical models for medical image analysis, MIT PhD Thesis; 2000.
[15] Nelson M, Cao F, Liu L, Huang HK. A pediatric mr image database for teaching neuro anatomy and diseases. Radiology 2002;225:P757.
[16] Talairach J, Tournoux P. Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system—an approach to cerebral imaging. New York: Thieme Medical Publishers; 1988.

H.K. (Bernie) Huang, DSc, FRCR(Hon.), Professor of Radiology and Biomedical Engineering; Director, Division of Imaging Informatics, Department of Radiology, and Director MS Program in Medical Imaging and Imaging Informatics, Department of Biomedical Engineering, University of Southern California, USA; Chair Professor of Medical Informatics, The Hong Kong Polytechnic University; and Honorary Professor, Shanghai Institute of Technical Physics, The Chinese Academy of Sciences.Dr Huang's current research interests are in tele-imaging and telemedicine, fault-tolerant PACS server, PACS ASP model, Internet 2, PACS-based CAD and surgery, imaging informatics, image recovery during disaster, image integrity, data grid, grid computing, and HIPAA compliance. He has co-authored and authored seven books, published over 200 peer-reviewed articles, and received several patents. His latest book: 'PACS and Imaging Informatics' was published by John Wiley & Sons in 2004.Dr Huang was inducted into the Royal College of Radiologists, London as an Honorary Fellow, for his contribution in PACS research and development in November, 1992; the American Institute of Medical and Biological Engineering as a Founding Fellow, for his contribution in medical imaging in March 1993; the EuroPACS Society as an Honorary Member for his contribution in PACS in October, 1996; and was elected as the Honorary President, 2003 International CARS Congress, London. Dr Huang has been Visiting Professor in many leading universities around the world.

DTD 5 ARTICLE IN PRESS

ELSEVIER

Computerized Medical Imaging and Graphics xx (2004) 1–12

Computerized Medical Imaging and Graphics

www.elsevier.com/locate/compmedimag

A model of DICOM-based electronic patient record in radiation therapy

Maria Y.Y. Law*

Department of Optometry and Radiography, The Hong Kong Polytechnic university, Hung Hom, Hong Kong
Shanghai Institute of Technical Physics, Chinese Academy of Sciences, China

Received 1 June 2004; revised 30 June 2004; accepted 30 September 2004

Abstract

Electronic patient records for radiation therapy (RT) consists of text, images and graphics. To enable the exchange of RT patient information between systems and institutions, a common standard is called for and the DICOM standard extension to radiation therapy is an appropriate means for standardization. This paper describes a model of DICOM-based electronic patient record system for information exchange and sharing. The system used a DICOM-based RT archive server as a common platform for archival of all RT related information including images and the web technology for distribution and viewing of the patient electronic record.

Keywords: DICOM-RT; DICOM; PACS; Electronic patient record; Radiation therapy

1. Introduction

Electronic patient records (EPR) started to evolve almost a decade ago mainly in conjunction with the hospital information systems for administration and access to laboratory results [1,2]. Such records are textual based and it was not until the past 3 years or so that images from radiological procedures or electrocardiograms were incorporated into electronic patient records to make them 'complete'. Completeness of patient data will only be realized when medical information of all specialties can be integrated into a patient's electronic record. From such understanding, radiation therapy patient's record could never be complete if the radiation therapy data, which include images, treatment plans and records, were not integrated into the electronic patient record system.

Radiation therapy (RT) uses radiation for treatment of diseases, most of which are malignant. Before delivering the dose of radiation, careful treatment planning needs to be done. This ensures that the target tumor volume is accurately irradiated while the neighboring normal tissue is spared as much as possible. Such treatment planning results in isodose treatment plans superimposed on CT images illustrating the radiation dose distribution in the irradiated volume, dose volume histograms, treatment parameters and treatment records. It can be seen that in radiation therapy, not only radiological images are involved, but graphics and textual information are also generated. Fig. 1 is a brief description of the work and data involved in radiation therapy. To integrate all such RT information so that it can be communicable with other RT systems, a standard is called for.

The Digital Communication in Medicine (DICOM) standard is the cornerstone to the successful implementation of Picture Archiving and Communication System (PACS) in radiology. By using the DICOM format and communication protocol, images acquired from equipment of different vendors can readily communicate with one another [3]. Following the implementation of DICOM format, seven DICOM radiotherapy (RT) objects in DICOM format have been ratified by the DICOM Committee for transmission and storage of radiotherapy images and related information [4,5]. The seven DICOM-RT objects include RT Image, RT Plan, RT Structure Set, RT Dose, RT Beams Treatment Record, RT Brachy Treatment Record and RT Summary Record (Fig. 2) [6,7]. Using RT Structure Set as example, Table 1 illustrates the attributes in a DICOM-RT object. The DICOM-RT information model can be used as the data structure for the electronic patient record. However, for

* Tel.: +852 2766 6106; fax: +852 2764 6051.
E-mail address: ormaria@polyu.edu.hk.

0895-6111/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compmedimag.2004.09.012

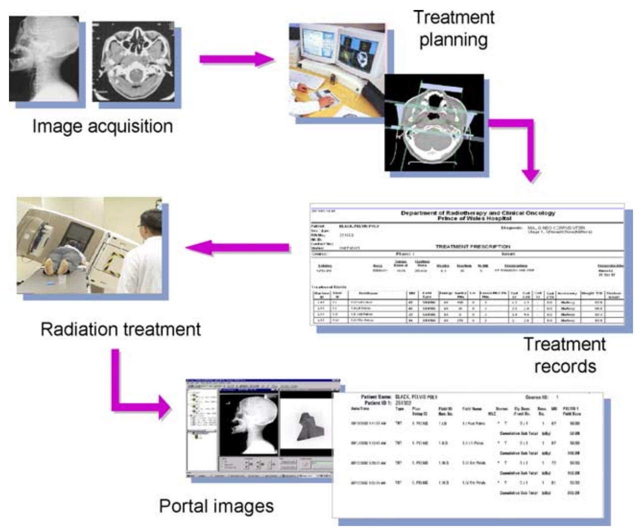


Fig. 1. Radiation therapy data and workflow.

various reasons, the utilization of the RT objects has not been implemented in daily clinical operation. This has impeded the exchange of RT information between systems and between departments. This paper aims to describe a model of integrated electronic patient record system based on the DICOM-RT standard for radiation therapy patients.

2. Methods

2.1. Data modeling

To develop an electronic patient record, a conceptual data model of the clinical department is required and this in turn determines how patient data are physically represented in the database [8,9] To develop the conceptual data model, the radiation therapy workflow was first reviewed from which the data required were defined [6,10,11]. At the same time, the views of the users (radiation oncologists and radiation therapists) were obtained. From these views, the conceptual data model became increasingly refined with more details. Table 2 shows the design of the conceptual data model of an RT electronic patient record. The model was based on the daily operation mode in a radiation therapy department and the users' requirements. The data in the model are parsed from the DICOM images and DICOM-RT objects. When an object is inserted, a link is created at the corresponding location in the table. For example, when a simulator image was performed on 2 August 2002, an 'x' appeared against the corresponding category RT Image and along the column of 2 August 2002. This 'x' indicates that a procedure had been done and also the nature of the procedure. Likewise, when a treatment is done, an RT Beams Record and an updated summary record are created and routed to an archive server. A cross will be created on the date that the treatment is done in this conceptual data model.

In Table 2, the patient demographics in each patient record are parsed from the patient module and the study module of the DICOM standard. As to the RT procedures (the most left hand column below the dates), 'RT Structure Set', 'RT Plan' and 'RT Dose' are grouped as Treatment Plan because in clinical, they are rarely viewed separately on their own. Occurrence of any one or more of these three objects for the related plan is placed in the same box of the corresponding date. 'RT Beams Record' indicates that the record is prepared (or the preparation for treatment is ready) and 'No. of treatment' counts the fractions of radiation that have been delivered. 'Brachy Record' provides similar information as 'RT Beams Record' but will be implemented at a later stage. 'Treatment status' and 'Treatment comment' are attributes in the DICOM-RT Record. These two are often used when a radiation oncologist reviews a patient

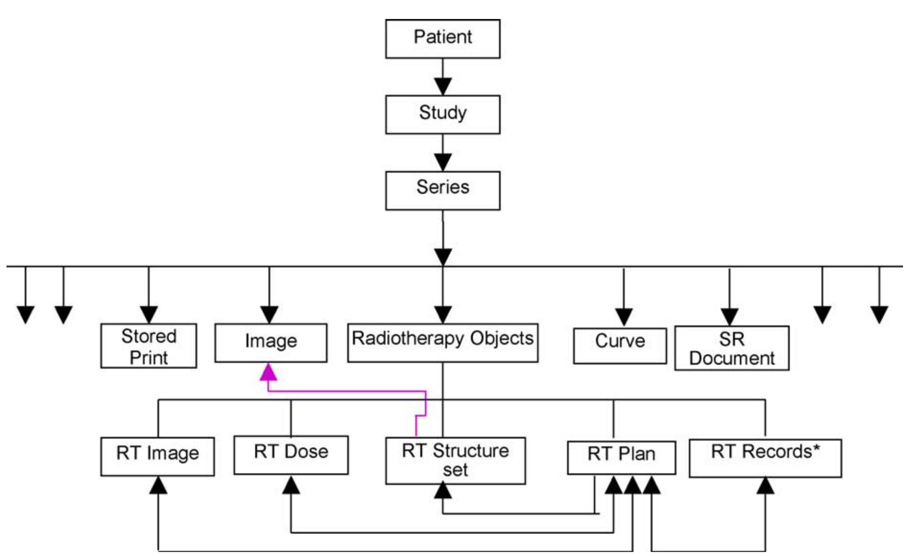


Fig. 2. Abbreviated DICOM Model of the real world. *The three records in DICOM-RT are grouped under RT Records.

Table 1
Example of attributes of in a DICOM-RT object

RT _STRUCTURESET
*STRUCTNo
FKSeriesNo
FKStudyNo
FKPatientNo
File Attribute Object
RT StructureSet Attribute Object
RT Contour Attribute Object
RT ROI Observations Attribute Object
RT Approval Attribute
SOP Common Attribute Object

The asterisk shows the primary key and the italics show the attributes specific to the object.

weekly during a course of radiation therapy. 'Continue RT' (as shown in Table 2) indicates to the radiation therapist that the patient should be continued with the treatment. If any medicine is given during the course or any special precautions highlighted, they can be entered in the 'Treatment comment' field. These parameters nicely summarized what has been done on a patient. Clicking on the highlighted box (or icon) should show the detail of the object, e.g. a set of CT image or the summary of the record as on certain date.

2.2. Data flow

Since DICOM-RT objects are also based on the DICOM standard, the data model of the PACS Server and the PACS data flow can be adopted in the DICOM-based RT electronic patient record [9]. Following the clinical data flow and based on the generic PACS data flow, Fig. 3 is a design of the RT data flow through the computer components in the DICOM-based RT electronic patient record system [12]. The RT data input is from various RT equipment, imaging modalities or the PACS. In this system prototype, an RT Modality Simulator is used in

Table 2
A conceptual data model with time line showing the occurrence of DICOM images and DICOM objects of a radiation treatment for breast cancer

Patient name	Chan Mei Mei																
Patient ID	A123456																
Sex	Female																
Age	45																
Telephone	987654																
Address	10 Peter street, Hong Kong																
Course	1																
Region	Breast Axilla																
	1 Aug 02	2 Aug 02	3 Aug 02	4 Aug 02	5 Aug 02	6 Aug 02	7 Aug 02	8 Aug 02	9 Aug 02	10 Aug 02	11 Aug 02	12 Aug 02	13 Aug 02	14 Aug 02	15 Aug 02	16 Aug 02	17 Aug 02
DICOM image																	
CT/CT sim	x																
MR																	
RT Image																	
Simulation image		x															
DRR																	
Portal image							x										
Treatment plan (RT Structure Set, RT Plan, RT Dose)					x												
RT Beams Record					x												
No. of treatment							1	2	3			4	5	6	7	8	
Brachy Record																	
Treatment Status									Continue RT							Continue RT	
Treatment comment																	
Treatment Summary Record	x	x			x		x	x	x			x	x	x	x	x	

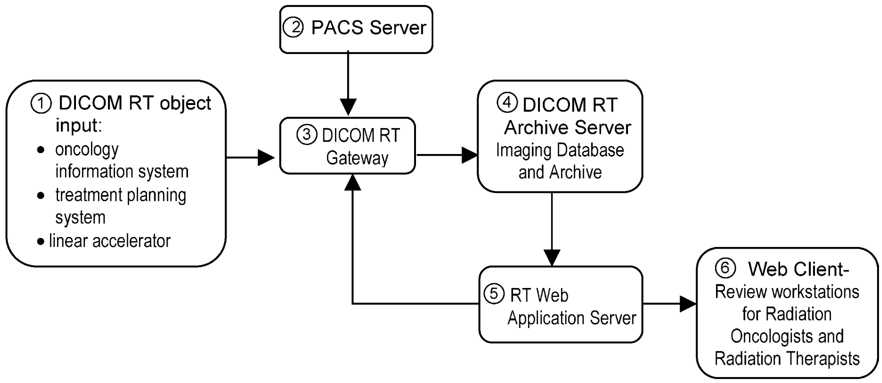


Fig. 3. RT data flow.

the laboratory for input of RT objects obtained from clinical departments. Using the RT Modality Simulator can avoid disturbance to clinical service. Thus, in this system prototype, the computer components consist of the RT Modality Simulator, the DICOM-RT Gateway, the DICOM-RT Archive Server, RT Web Application Server and the Web Client. Their functions are explained in the following paragraphs.

2.2.1. DICOM-RT Modality Simulator

This is a laboratory workstation that simulates RT equipment in inputting DICOM-RT objects. At the current stage of development, many RT modalities are not DICOM compliant yet. It was difficult to collect sufficient DICOM-RT objects for system evaluation. Also the translated DICOM-RT objects needed testing and validation. A radiotherapy modality simulator was built in the laboratory to simulate the RT object input in clinical settings. All RT objects (including those translated from non-DICOM standards) and DICOM images could be exported from this RT Modality Simulator to the RT Archive Server.

2.2.2. PACS

From the PACS, images used in radiation therapy images are forwarded either to the Treatment Planning System or Virtual Simulator for radiation treatment planning or to the RT Archive Server for storage to be reviewed later.

2.2.3. RT DICOM Gateway

The Gateway receives all DICOM objects, acknowledges receipt of the objects and extracts information from the objects and put them into the RT data model as required in the RT Archive Server.

2.2.4. RT Archive Server

The RT Archive Server is a PACS-based server in that the data model of the PACS Server was adopted for this research with some modifications to accept all DICOM-RT objects. For the RT Archive Server, the SUN Ultra 2 computer with the Oracle database operating in the UNIX Solaris 8 environment was used. The database schema is shown in Fig. 4. It shoulders all the functions of storage, management, controlling and being a database server. On receiving DICOM-RT objects and images from the DICOM-RT Gateway, the RT Archive Server extracts only the essential aspects of the object entities and auto-routes all the data to the Web Application Server.

2.2.5. RT Web Application Server

The RT Web Application Server focuses on the management and processing of the patient planning and treatment information that are decoded from the RT objects and be used by the RT Web client. Microsoft Access 2000 database was used as the database for the Application Server. The database schema is described in Section 2.3. For distribution of RT information, a web server, the Window Internet Information Server (WIIS) was used and the data were sent using the hypertext transfer protocol (HTTP).

2.2.6. RT Web clients

Two types of clients were identified. They are the radiation oncologists and the radiation therapists. User interface was designed for the clients.

2.3. Database schema

The database schema refers to how data are physically represented. It is concerned with data structures, file organizations and mechanism for the operation of the system and data storage. From the data flow diagram, it can be seen that in this EPR system, two databases are designed, one for the RT Archive Server and the other for the RT Web Application Server. The former is for management and storage of DICOM objects (including DICOM-RT objects) and the latter is for the data parsed from the DICOM-RT objects to be viewed by users at the web client.

2.3.1. Database schema of the RT Archive Server

The server system in this project was designed as a three-tier client/server architecture. In a three-tier architecture, a middle tier is added between the user interface client environment and the database management server environment. For this system, the three-tiers are: (1) the RT Archive Server which provides

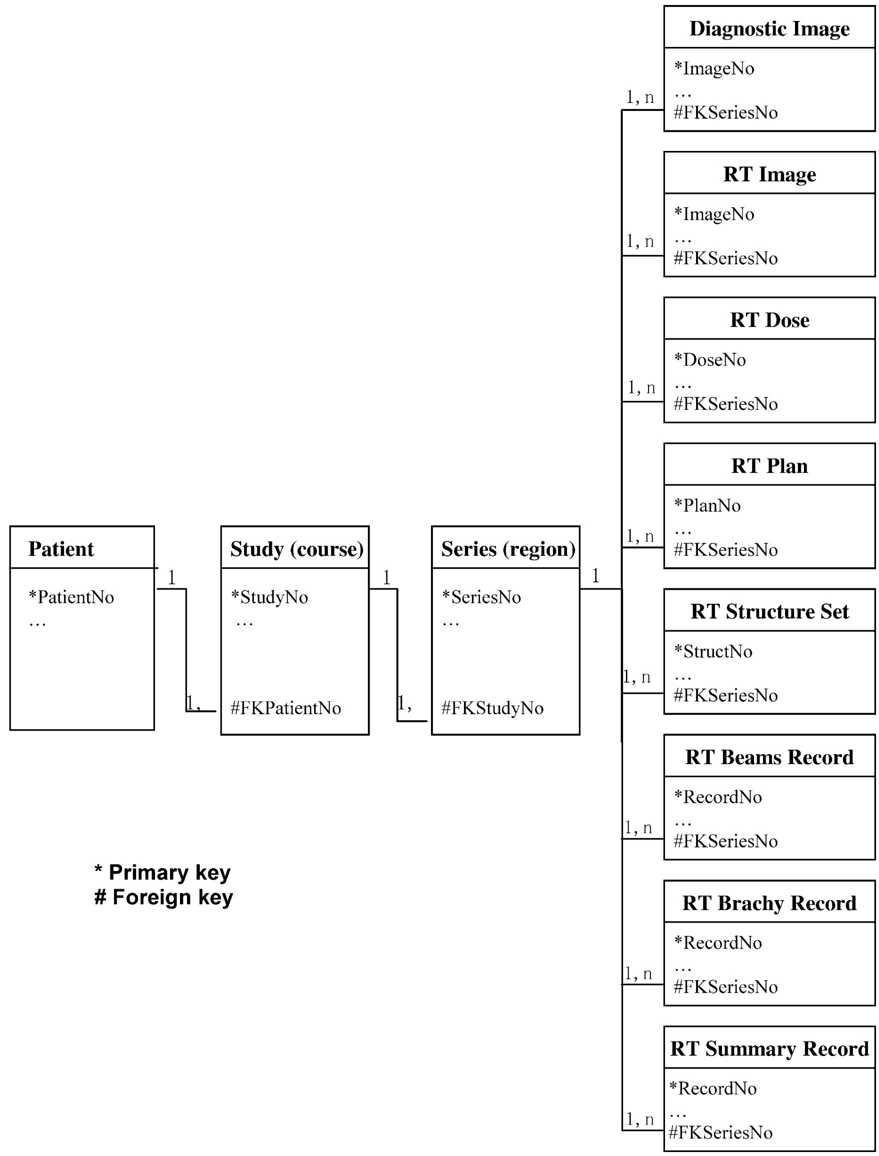


Fig. 4. Database schema of RT Archive Server.

such functions as managing, archiving, and transferring of the DICOM images and DICOM-RT objects; (2) The RT Web Application Server focusing on processing of the patient planning and treatment data; (3) The RT Web Client that presents the patient's record (Fig. 3).

DICOM is an object-oriented standard. The external aspects (operation) of an object are separated from the internal details of the data, which are hidden from the outside world (information hiding) [9]. This organization allows identification of the object or any operations on the object first and delays the implementation of details. Also changes to the internal details at a later stage will not affect the applications that use it, provided that the external aspects remain unchanged. In this way, the database server only needs to identify what an object is and what it does. The internal details or the data structure of an object can be implemented or processed later in the application server. This has been the way that the PACS Server was designed; hence it has a simple data model for the operations of its objects. Such a design was also adopted for this RT Archive Server which only manages the essential aspects of the RT objects, leaving the implementation of details to the RT Web Application Server. It thus has only the basic database schema based on the first few levels of the DICOM hierarchical structure (Fig. 2), i.e. 'patient', 'study', 'series' and its RT objects and diagnostic images (Fig. 4). Oracle8i release 8.1.7 Database Management System was used in the RT Archive Server.

2.3.2. Data schema of RT Web Application Server

For implementing the details of the DICOM-RT objects using the web-based application server, a data model different and more elaborated than that of the RT Server is required. The basic data structures are given in the

DICOM standard documents. How the data would be used depends on the actual application at the client workstations, which in turn determines how much should be included in the data model. From the user requirements collected earlier, the physical data model for the web application server was designed and implemented following the DICOM standard (Fig. 5). Microsoft Access 2000 was used for the database.

2.4. User interface design

The user interface presents the patients' information to the users and is an important part of the EPR. The design of the interface was based on the survey of the user requirements mentioned in Section 2.1. Radiation therapists and oncologists were involved in the design process. Fig. 6 shows the hierarchical structure of the user interface windows and the functions served by each window are explained in Fig. 7. The user interface was created in the form of Active Server Page (ASP), which is a feature of WIIS. The interactive functions of the web graphic user interfaces (GUI) were implemented by Visual Basic (VB) scripts embedded in the ASP.

2.5. Data collection

Different types of RT files (DICOM and non-DICOM) were collected from RT vendors and clinical departments. DICOM files include CT and MR images, digitized simulator and portal images, RT Plan, RT Structure Set and RT Dose. Non-DICOM files include treatment planning files and treatment record in textual format, and portal image in tiff/bitmap format. The names of patients were anonymized. The non-DICOM files were translated to DICOM format. After testing the files for successful transmission through the laboratory computer components, the DICOM files were grouped into folders to form 10 virtual patients so that their electronic record could be displayed in the web client.

3. Results

The EPR System was implemented and the user interface windows for the web client workstations were created with interactive functions embedded. The virtual patients' information was successfully transmitted, stored and viewed at the RT Web Client. Using the user interface windows, the

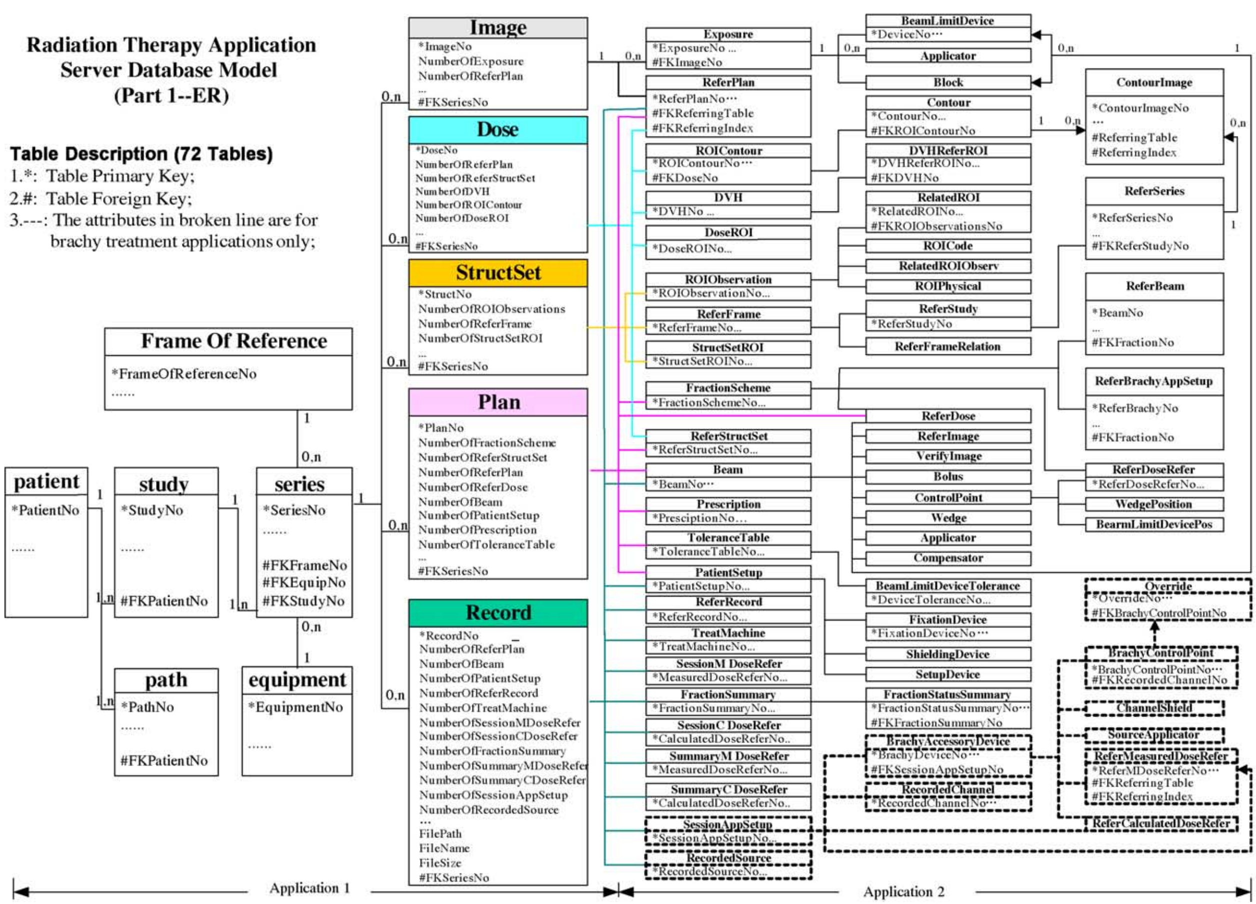


Fig. 5. Database schema of the RT Web application server. Courtesy of Zhang Xiaoyan.

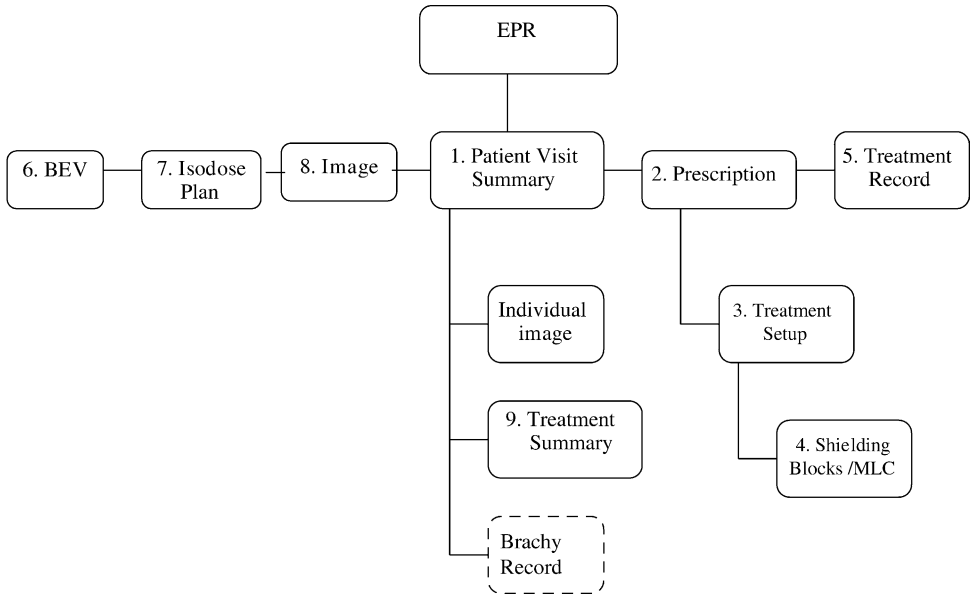


Fig. 6. Web site hierarchy for the Radiotherapy Electronic Patient Record System.

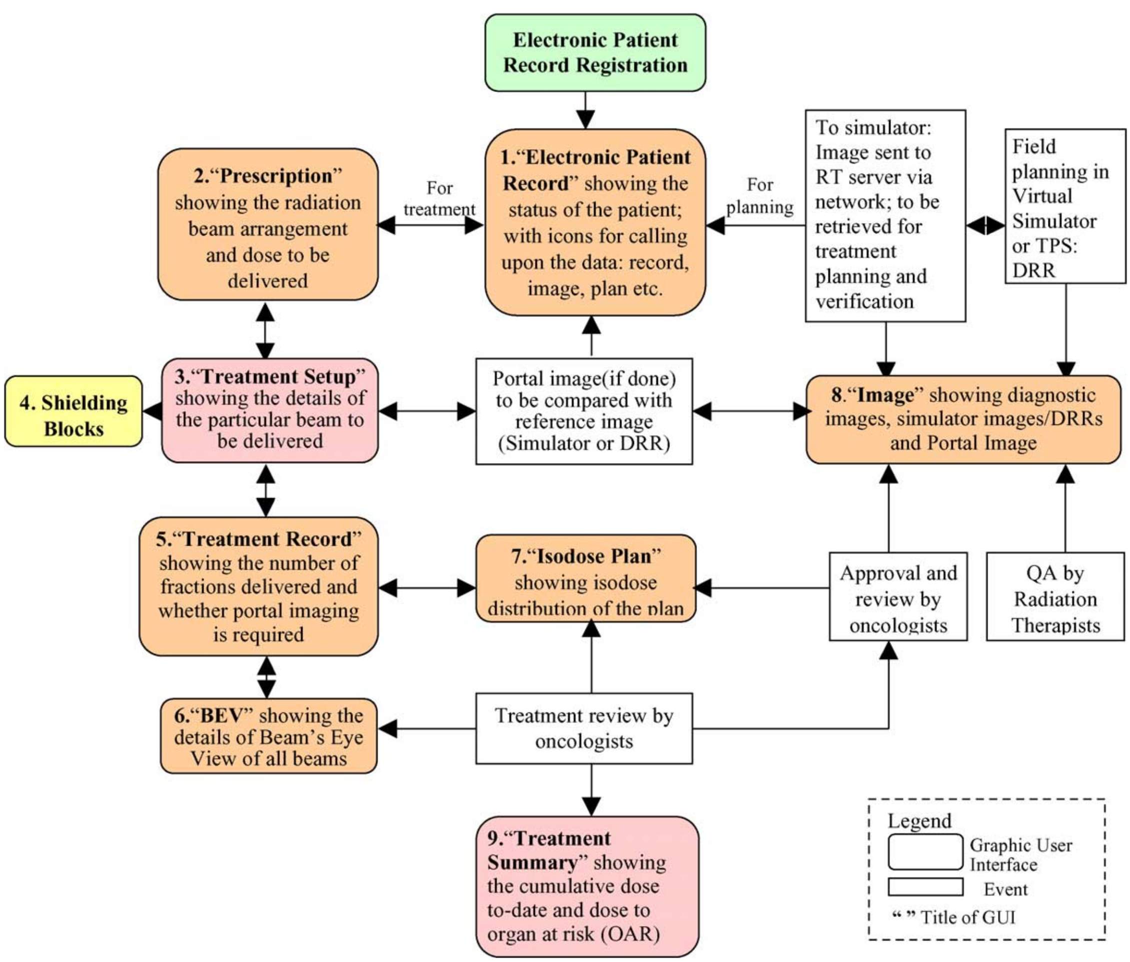


Fig. 7. Flow of Graphic User Interface based on clinical workflow.

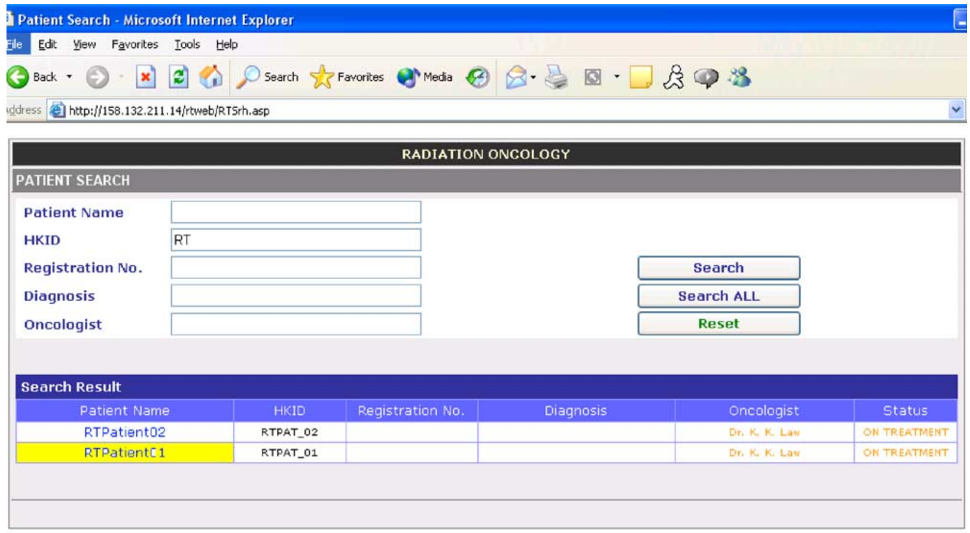


Fig. 8. Patient list.

following paragraphs illustrate the EPR of one of the patients stored in the archive server.

3.1. The RT electronic patient record

A patient, whose name is RTPatient01, with a Hong Kong Identity card number (HKID) of RTPat_01, has been planned for a course of radiotherapy. He has finished all the treatment planning procedures and is receiving the radiation treatment. He comes back for treatment. At the reception of the treatment unit, the receptionist or the radiation therapist will call up the Patient Query page and type in the patient's HKID number or just 'RT' against the HKID to search for the patient's detail. The list of patient beginning with HKID number of 'RT' is shown (Fig. 8). In the database, only two patients with such a beginning and so only two patients' names are shown and RTPatient02 is the other patient. The same page and search procedure can be used when a patient re-visits the department for planning procedures or goes to see a radiation oncologist for review or to receive radiation treatment.

A click on the patient's name, RTPatient01 pops up the patient's visit summary (Fig. 9) with all the procedures done. The radiation therapist or the radiation oncologist can then at a glance see to the status of the patient. In this case, in the treatment status row, the latest comment from the radiation oncologist is 'Cont.

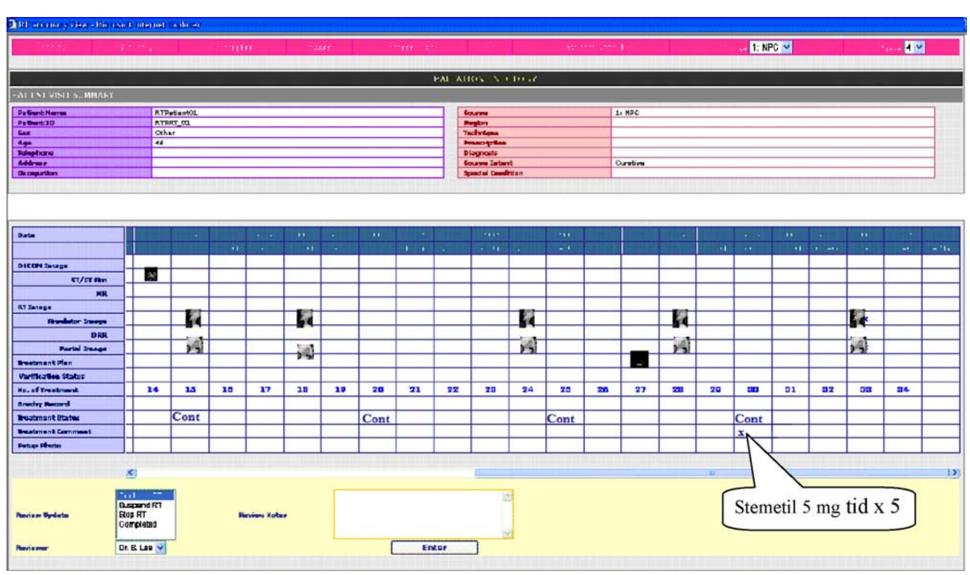


Fig. 9. Patient visit summary.

RT', meaning to continue with the treatment. If the visit is to consult the oncologist, the same Visit Summary can be used for the oncologist to add his/her comments about the patient. Pointing at the treatment comment shows the balloon containing the oncologist's comments.

To set up the patient for treatment, the radiation therapist needs to refer to the prescription of the patient. In the toolbar, at the top of the window is a list of function for users to switch between pages for searching the information about the patient as needed. One of the function button is 'prescription'. A click on the 'prescrip-prescription' pops up the prescription page with all the essential detail about the treatment prescription (Fig. 10) including the radiation fields to be treated.

Clicking on any of the button in the Field ID leads to the set-up details for that radiation field. In this case, Field 24 is to be treated, the Field ID '24' is clicked and that leads to the Set-up page where greater details about the treatment set-up are provided (Figs. 11 and 12). On the Treatment Set-up page, the set-up details are shown. A click on the Block position button shows the shielding block as in Figs. 11 and 12.

Similarly, a click to the MLC Plan button (right hand side) will call up the MLC shape. From the recommendation of a clinical radiation therapist, a button is added to link to any photo that is taken related to the set-up.

When the radiation dose for a field is delivered, a click on the 'Treated' button will update the field in the Treatment Record (Fig. 13). When all the fields prescribed for the day are treated, the cumulative dose in the Treatment Record will be added and the Summary Record will also be updated (Fig. 14).

Some buttons at the tool bar on the top of each page provides a link to the stated information. For example, 'Summary' leads to the Treatment Summary Record window (Fig. 14), Images (Fig. 15) leads to all the images of the patient for the particular course, the 'Isodose Plan' button shows the distribution of radiation dose around the marked target volume and other nearby anatomical structures on the cross-sectional CT images (Fig. 16). The BEV is a reserved button for showing the beam's eye view. The reconstruction of the Beam's Eye View is to be added in the future research.

4. Discussion

4.1. Communication between isolated information systems and archival of information

In a department, often there are different isolated information systems with only small scope and for single-purpose applications. They usually come with the purchase of individual applications, like one system for brachytherapy, another system for stereotactic radiotherapy/radiosurgery (SRT/SRS). They often stand alone, having little interfacing with other systems. A BrainLab workstation for SRT/SRS has its own storage for the plans performed at its workstation. For conventional radiotherapy, the treatment plans are stored in the conventional Treatment Planning System (TPS) while the treatment

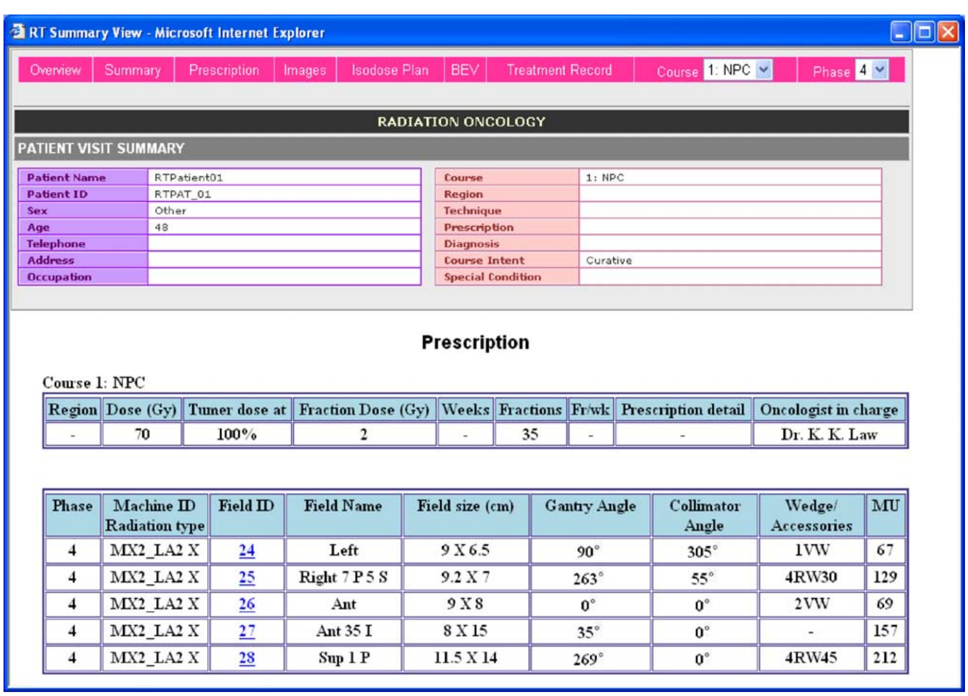
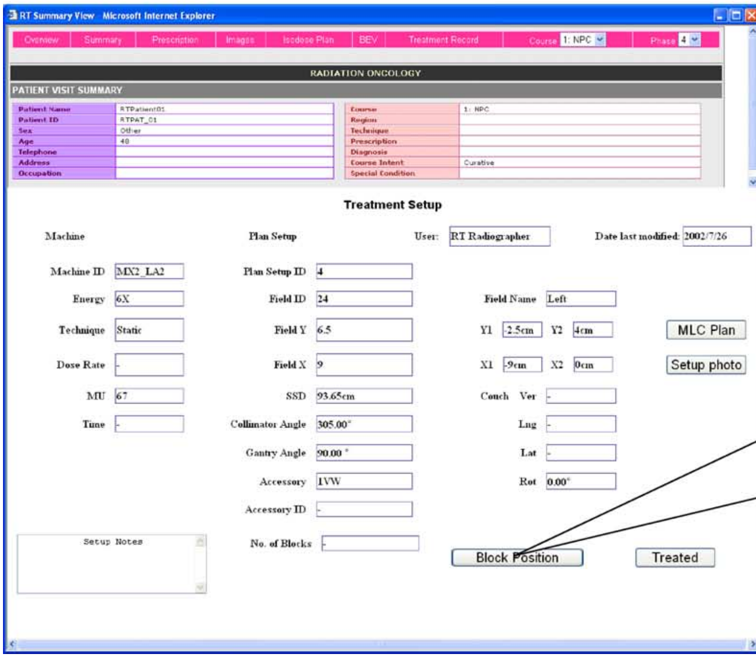
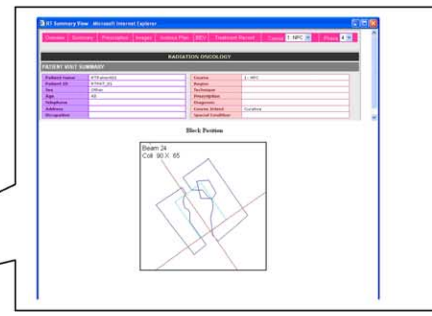


Fig. 10. Prescription page.

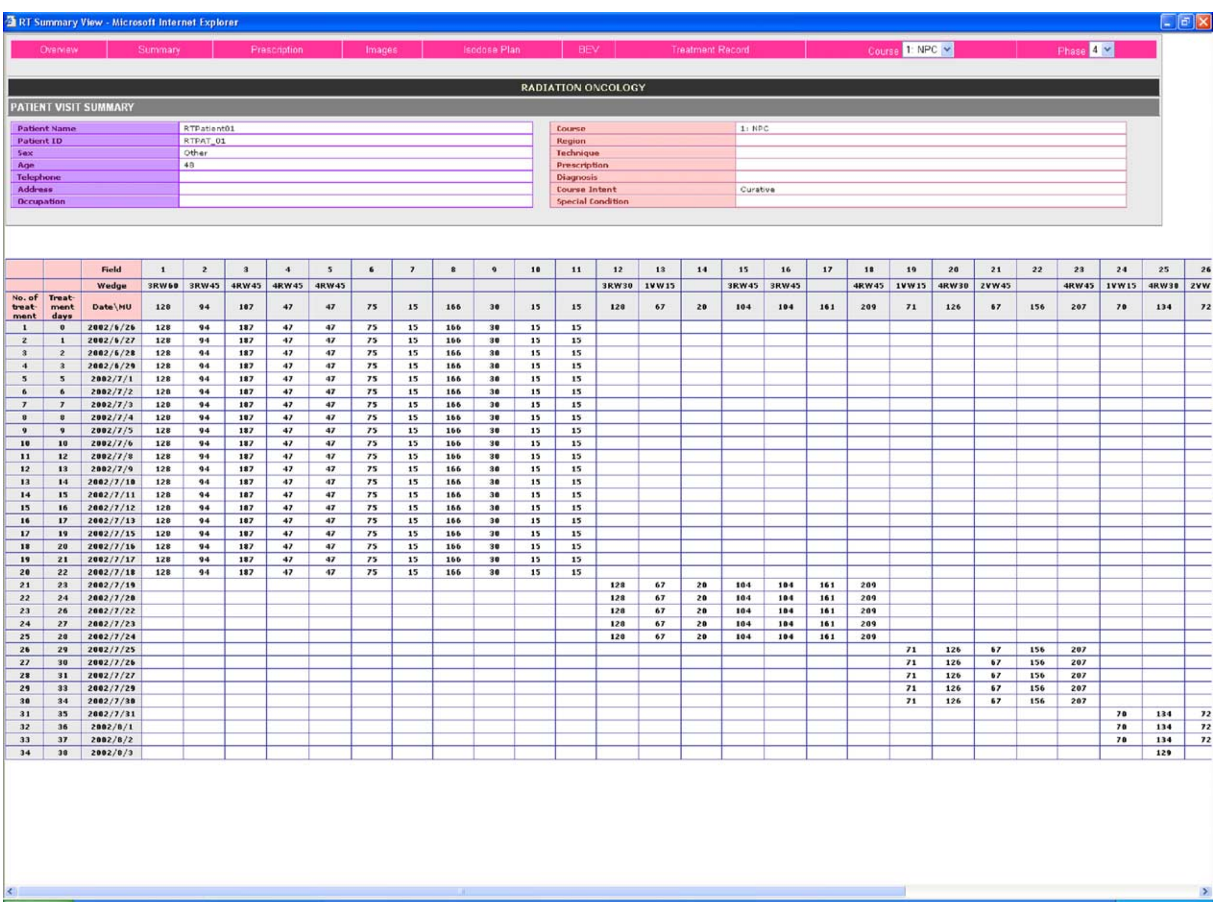
(11)



(12)



Figs. 11 and 12. (11) Treatment set-up page. (12) Shielding block position.



		Field	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
		Wedge	3RW60	3RW45	4RW45	4RW45	4RW45							3RW30	1VW15		3RW45	3RW45		4RW45	1VW15	4RW30	2VW45		4RW45	1VW15	4RW30	2VW
No. of treatment	Treatment days	Date\MU	120	94	107	47	47	75	15	166	30	15	15	120	67	20	104	104	161	209	71	126	67	156	207	70	134	72
1	0	2002/6/26	120	94	107	47	47	75	15	166	30	15	15															
2	1	2002/6/27	120	94	107	47	47	75	15	166	30	15	15															
3	2	2002/6/28	120	94	107	47	47	75	15	166	30	15	15															
4	3	2002/6/29	120	94	107	47	47	75	15	166	30	15	15															
5	5	2002/7/1	120	94	107	47	47	75	15	166	30	15	15															
6	6	2002/7/2	120	94	107	47	47	75	15	166	30	15	15															
7	7	2002/7/3	120	94	107	47	47	75	15	166	30	15	15															
8	8	2002/7/4	120	94	107	47	47	75	15	166	30	15	15															
9	9	2002/7/5	120	94	107	47	47	75	15	166	30	15	15															
10	10	2002/7/6	120	94	107	47	47	75	15	166	30	15	15															
11	12	2002/7/8	120	94	107	47	47	75	15	166	30	15	15															
12	13	2002/7/9	120	94	107	47	47	75	15	166	30	15	15															
13	14	2002/7/10	120	94	107	47	47	75	15	166	30	15	15															
14	15	2002/7/11	120	94	107	47	47	75	15	166	30	15	15															
15	16	2002/7/12	120	94	107	47	47	75	15	166	30	15	15															
16	17	2002/7/13	120	94	107	47	47	75	15	166	30	15	15															
17	19	2002/7/15	120	94	107	47	47	75	15	166	30	15	15															
18	20	2002/7/16	120	94	107	47	47	75	15	166	30	15	15															
19	21	2002/7/17	120	94	107	47	47	75	15	166	30	15	15															
20	22	2002/7/18	120	94	107	47	47	75	15	166	30	15	15															
21	23	2002/7/19												120	67	20	104	104	161	209								
22	24	2002/7/20												120	67	20	104	104	161	209								
23	26	2002/7/22												120	67	20	104	104	161	209								
24	27	2002/7/23												120	67	20	104	104	161	209								
25	28	2002/7/24												120	67	20	104	104	161	209								
26	29	2002/7/25																			71	126	67	156	207			
27	30	2002/7/26																			71	126	67	156	207			
28	31	2002/7/27																			71	126	67	156	207			
29	33	2002/7/29																			71	126	67	156	207			
30	34	2002/7/30																			71	126	67	156	207			
31	35	2002/7/31																								70	134	72
32	36	2002/8/1																								70	134	72
33	37	2002/8/2																								70	134	72
34	38	2002/8/3																									129	

Fig. 13. Treatment Record.

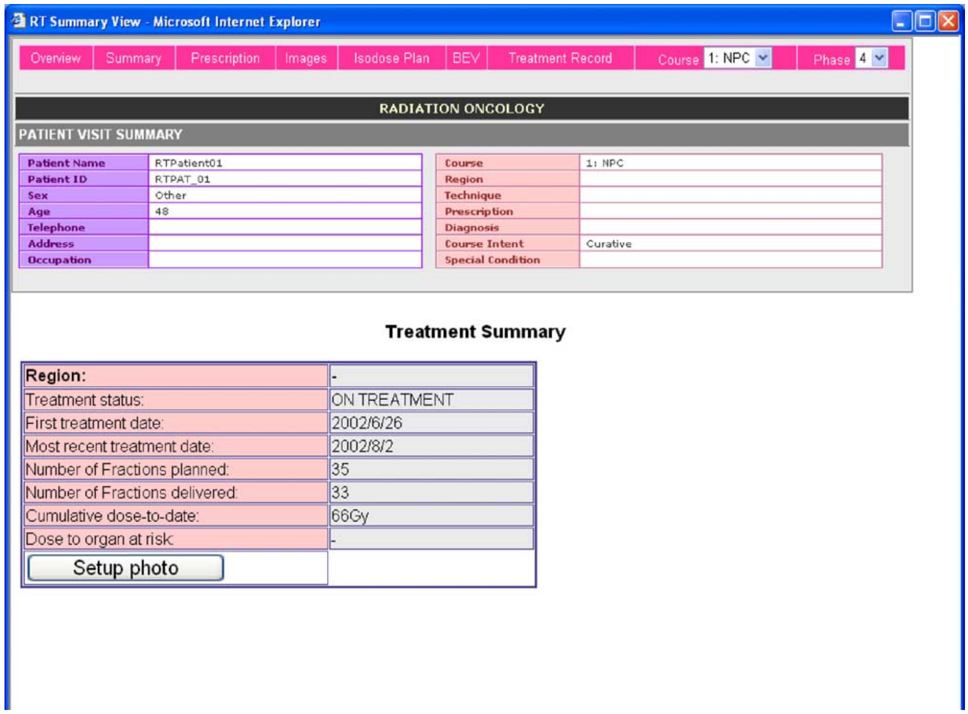


Fig. 14. Treatment Summary Record.

records in another information system. A patient whose treatment involves all three workstations will have his treatment information in three different places. Currently, such treatment information is normally 'linked' by a paper record or folder of the patient. This has not taken into account the hardcopy films of images that are stored separately in the film library. This is very common in many radiotherapy departments. Were the paper record lost, then the patient treatment information would be 'disintegrated'. Using the DICOM and DICOM-RT standards, this EPR system integrates patient information from different systems and a live summary of the patient's treatment record can be displayed when required. This would help archive patient information from different systems and save efforts and time in searching for records and films as well as safeguarding such loss.

4.2. Information sharing

Most hospital information systems and sub-systems have been organization-oriented or system-oriented rather than patient-oriented. This means that to query medical information of a patient, one may need to go through several systems. Also patient data collected in these systems are generally not widely available for immediate integration due to the differences in formats between workstations and systems. This becomes worse if consultation is required across institutions. The lack of an integrated database causes discontinuities in care and often results in redundant questioning or, worse, clinical decisions based on incomplete data. It also limits the ability to conduct clinical and research queries, including the creation of patient cohorts for prospective or retrospective studies [13]. The DICOM-based RT Archive Server in this system provides a platform for information sharing.

Fig. 15. Image window. From left to right: CT image, simulator image, portal image.

Fig. 16. Isodose plan window.

4.3. A model of comprehensive electronic patient record

With the maturity of DICOM standard, PACS and IHE (Integrating the Healthcare Enterprise), researchers are now working towards incorporating medical images, e.g. radiology images, endoscopy images, microscopy images into the electronic patient records [1]. However, the radiation therapy plans and records have not been taken care of since there does not exist yet a system with a common standard integrating text, image and graphics. This is because other than being image-intensive, radiation therapy is highly technical and its use of radiation also involves radiobiological factors. All these parameters have to be recorded for future reference in the management of cancer patients treated by radiation therapy. Hence, other than textual information, all related treatment plans and images need to go into the patient's record. The DICOM-RT standards are now set. It is a matter of implementation and refinement before radiotherapy information can be, like images, linked to the electronic patient record to make it complete. The integrated prototype system in this research is a starting point to this initiative of completing the electronic patient record.

5. Summary

All radiation therapy vendors are moving towards implementing their information systems from which, a kind of electronic patient record is generated, be it complete or not. Nevertheless, such records are still in vendor specific formats that cannot be readily read in other systems. A major impeding factor is that DICOM-RT records are still not implemented in most cases, let alone DICOM-RT Dose in some. Except for DICOM-RT brachy record, this research rendered all RT information to the DICOM standard and in so doing, provides a model for integrating RT information into a patient-oriented electronic record. Other than providing a comprehensive electronic record for radiation therapy patients, this EPR system helps in the archival of RT information from different systems, exchange of information between institutions and collaborative research between clinical departments.

References

[1] Kuzmak PM, Dayhoff RE. The use of digital imaging and communications in medicine (DICOM) in the integration of imaging into the electronic patient record at the Department of Veterans Affairs. J Digit Imaging 2000;13(2 Suppl. 1):133–7.

[2] Ratib O, Swiernik M, McCoy JM. From PACS to integrated EMR. Comput Med Imaging Graph 2003;27(2–3):207–15.

[3] Law MYY, Zhou Z. New direction in PACS training. Comput Med Imaging Graph 2003;27(2–3):147–56.

[4] Digital imaging and communications in medicine (DICOM). Radiother Objects 1997;Suppl. 11.

[5] Digital imaging and communications in medicine (DICOM). Radiother Treat Rec Radiother Media Extensions 1999;Suppl. 29.

[6] Law MYY, Huang HK. Concept of a PACS and imaging informatics-based server for radiation therapy. Comput Med Imaging Graph 2003; 27(1):1–9.

[7] DICOM Standard 2003, http://medical.nema.org/dicom/2003.html

[8] Johns ML. Information management for health professions. 2nd ed. Australia: Delmar Thomson Lerning; 2002.

[9] Connolly T, Begg C. Database systems—a practical approach to design, implementation, and management. 2nd ed. England: Addison Wesley; 1998.

[10] Law MYY, Huang HK, Zhang X, Zhang J. DICOM and imaging informatics-based radiation therapy server. Proceedings of the SPIE on CD-ROM: medical imaging, San Diego, USA 2002 p. 160–167.

[11] Law MYY, Huang HK, Zhang X, Zhang J. The data model of a PACS-based DICOM radiation therapy server. Proceedings of the SPIE on CD-ROM: medical imaging, San Diego, USA 2003 p. 118–129.

[12] Law MYY, Huang HK, Chan CW, Zhang X, Zhang J. A DICOM-based radiotherapy information system. Proceedings of the SPIE on CD-ROM: medical imaging, San Diego, USA 2004 p. 118–129.

[13] Palta JR, Frouhar VA, Dempsey JF. Web-based submission, archive, and review of radiotherapy data for clinical quality assurance: a new paradigm. Int J Radiat Oncol Biol Phys 2003;57(5):1427–36.

Maria Y.Y. Law PhD, is the Assistant Professor of Radiation therapy in the Department of Optometry and Radiography, the Hong Kong Polytechnic University (PolyU) and the Associate of the Shanghai Institute of Technical Physics, Chinese Academy of Sciences. While specializing in radiation therapy, Maria is also involved in the development of the PACS and related research in the PolyU. She is also appointed as the Honorary Consultant of the PACS project of the Princess Margaret Hospital in Hong Kong. Her current research interest is in radiation therapy informatics. A radiation therapist by profession, Maria is deeply involved in activities of the local and international professional societies. She is the editor of the Hong Kong Radiographers Journal and the Chairman of the Local Organizing Committee of the 2005 World Congress of the International Society of Radiographers and Radiological Technologists

Computerized Medical Imaging and Graphics xx (xxxx) 1–8
Computerized Medical Imaging and Graphics
www.elsevier.com/locate/compmedimag

Utilizing data grid architecture for the backup and recovery of clinical image data

Brent J. Liu*, M.Z. Zhou, J. Documet

Image Processing and Informatics Laboratory, Department of Radiology, Keck School of Medicine, USC, Los Angeles, CA 90033, USA

Received 30 March 2004; revised 18 August 2004; accepted 30 September 2004

Abstract

Grid Computing represents the latest and most exciting technology to evolve from the familiar realm of parallel, peer-to-peer and client-server models. However, there has been limited investigation into the impact of this emerging technology in medical imaging and informatics. In particular, PACS technology, an established clinical image repository system, while having matured significantly during the past ten years, still remains weak in the area of clinical image data backup. Current solutions are expensive or time consuming and the technology is far from foolproof. Many large-scale PACS archive systems still encounter downtime for hours or days, which has the critical effect of crippling daily clinical operations. In this paper, a review of current backup solutions will be presented along with a brief introduction to grid technology. Finally, research and development utilizing the grid architecture for the recovery of clinical image data, in particular, PACS image data, will be presented.

The focus of this paper is centered on applying a grid computing architecture to a DICOM environment since DICOM has become the standard for clinical image data and PACS utilizes this standard. A federation of PACS can be created allowing a failed PACS archive to recover its image data from others in the federation in a seamless fashion. The design reflects the five-layer architecture of grid computing: Fabric, Resource, Connectivity, Collective, and Application Layers. The testbed Data Grid is composed of one research laboratory and two clinical sites. The Globus 3.0 Toolkit (Co-developed by the Argonne National Laboratory and Information Sciences Institute, USC) for developing the core and user level middleware is utilized to achieve grid connectivity.

The successful implementation and evaluation of utilizing data grid architecture for clinical PACS data backup and recovery will provide an understanding of the methodology for using Data Grid in clinical image data backup for PACS, as well as establishment of benchmarks for performance from future grid technology improvements. In addition, the testbed can serve as a road map for expanded research into large enterprise and federation level data grids to guarantee CA (Continuous Availability, 99.999% up time) in a variety of medical data archiving, retrieval, and distribution scenarios.

Keywords: PACS; Grid computing; Fault tolerance; Disaster recovery

1. Introduction

Picture Archiving Communication Systems (PACS) has long been established as a major clinical image repository system. Although significant strides have been made to improve the operations of PACS, there still lacks a robust and cost-effective solution for maintaining continuous data backup and availability. During a downtime event, it is possible to lose access to pertinent clinical image data for a hospital PACS. Furthermore, if the event is severe enough, such as a disaster, PACS data can be lost. Because PACS is widely used in many hospitals, it has become a mission critical clinical image system that requires 24/7 clinical operations and access to PACS data [1]. Some common disaster scenarios include earthquake, fire, flood, sabotage, or any combination of these that results in the complete destruction of the archive storage and server. With the increasing presence of fully filmless hospitals it becomes more and more crucial to provide solutions to protect the PACS data. Some reasons for this include the fact that in

* Corresponding author. Tel.: +1 310 448 9120; fax: +1 310 448 94411.
E-mail address: brentliu@usc.edu (B.J. Liu).

0895-6111/$ - see front matter © 2005 Published by Elsevier Ltd.
doi:10.1016/j.compmedimag.2004.09.004

a filmless environment, only one data copy of the PACS exam may be available in PACS. In addition, future HIPAA (Health Insurance Portability and Accountability Act of 1996) requirements may necessitate a disaster recovery solution for PACS data. Other factors include the importance of historical PACS exams for accurate diagnosis of the current exam. Even though new PACS exams can be acquired during a disaster scenario, previous exams may not be accessible if the archive storage and server has been destroyed. The overall damages and costs associated with a destroyed PACS archive storage and server is comparable to losing the entire onsite film archive of the hospital department.

Component technologies of PACS have matured over the past several years, from image acquisition devices to display workstations, archive servers, networks, and, most significantly, the establishment of the DICOM standard and the Integrating the Healthcare Enterprise (IHE) initiative. One of the major components in PACS is the archive server and storage that serves as the clinical image data repository system. This component archives all clinical images for seven years, as mandated by the recently adopted HIPAA. Even though this component supports vast storage capacity, it is not enough to assure compliance that a single image within the archive cannot be lost under any circumstances. Fault-Tolerance, long a necessity in most other applications of mission critical computing, is now expected in medical applications. Furthermore, practical field experience demonstrates that PACS archive servers can and will experience downtime without warning. The following is a brief review of some of the current strategies and solutions to provide clinical image data recovery during a downtime event.

2. Available clinical image recovery solutions

Current image data recovery solutions vary in the approach towards creating redundant copies of PACS data. For some hospitals with PACS, there is no image recovery solution in which case the single copy of PACS data is vulnerable to loss. The following are other solutions [2] listed in order as well as some of the advantages and disadvantages:

1. The Second Backup Copy
2. The Third Backup Copy
3. The Continuous Available (CA) PACS Server
4. The ASP Backup Archive Server

2.1. The second backup copy of PACS data

This solution features a secondary copy of each new PACS exam acquired. This secondary copy can be stored:

1. Onsite directly in the PACS storage device.
2. Onsite in a fireproof safe or storage compartment.
3. Offsite in a storage vault.

The latter is a strategy adopted by most data centers because it is cost-effective. However, this solution does not provide a real-time backup as replacement hardware is still needed and data needs to be imported.

2.2. The third backup copy of PACS data

PACS archives with large data media storage capacities such as digital tape usually have large data media elements that can store PACS data. These tape media can hold up to a few week's worth of PACS data. In this case, a secondary copy may not be enough to protect against a downtime event because should a disaster occur right when the most recent tape is being filled to capacity and has not yet been sent offsite, the hospital could lose up to a few weeks' worth of image data. It is necessary to create a third copy to cover the turnover period between secondary copy tapes. As in the second copy solution, the third copy can be stored onsite or offsite in a data storage vault. Again, although this solution covers the up-to-date backup data, it still does not provide a real-time backup as replacement hardware is still needed and data needs to be imported.

2.3. Continuous availability (CA) PACS archive server

The previous solutions only protect the data stored on media but does not address the fact that should the archive server encounter a downtime event, clinical image data will be inaccessible. Recently, a relatively new concept of CA PACS archive server design has emerged [3,4]. This concept is designed to:

1. Minimize manual intervention.
2. Avoid the change of daily management and workflow operation routines.
3. Allow users to continue retrieving image data within seconds to minutes.
4. Be cost effective.

The CA PACS archive server features a triple modular redundant (TMR) system with a high level of integration and an elegantly simple voting mechanism to achieve the continuous availability for a PACS server and is implemented with external mass storage devices of RAID and DLT tape library. This design provides CA for the primary PACS server but it becomes costly as a second backup archive solution.

2.4. ASP model backup archive offsite

Another solution is to provide a short-term CA backup archive server using the Application Service Provider (ASP) model at an offsite location [5,6]. The primary hospital site and the backup archive site are connected via a broadband connection (e.g. T1, OC-3, or Internet 2). The ASP backup archive provides instantaneous, automatic backup of acquired PACS image data and instantaneous recovery of stored PACS image data all at a low operational cost because it utilizes the ASP business model. In addition, should the downtime event render the network communication inoperable, a portable solution is available using a Data Migrator. The Data Migrator can populate PACS exams that were stored on the backup archive server directly onto the clinical PACS within hours to allow the Radiologists to continue to read with previous PACS exams until new replacement hardware arrives and is installed or when the scheduled downtime event has been completed.

Although both of the above solutions exist, there are limitations. For example, the CA archive server has eluded widespread implementation because manufacturers have been confined by current Information Technology (IT) standards. The image recovery procedures for the ASP model backup archive are tedious and require manual intervention. In addition, the ASP backup server is one backup site serving potentially multiple hospital sites. It can quickly outgrow the cost-effectiveness and easily manageable support features as more hospital sites are incorporated. In order to have a design serving an enterprise level clinical data recovery, an innovative approach is needed. By utilizing the Data Grid technology, these inherent limitations of the current design would be alleviated.

3. Grid technology

Grid computing is the integrated use of geographically distributed computers, networks, and storage systems to create a virtual computing system for solving large-scale, data-intensive problems in science, engineering, and commerce [7–13]. Furthermore, a grid is a high-performance hardware and software infrastructure providing scalable, dependable, and secure access to any applications that utilize the grid. Unlike distributed computing and clustered computing, the individual resources in grid computing maintain administrative autonomy and are allowed system heterogeneity. This leads to better scalability and robustness. However, because of this, the resources within the grid must adhere to agreed-upon standards to remain open and scalable. A formal architecture, composed of five layers has been previously created to assure this standardization [9]:

1. *Fabric Layer*. This is the lowest layer and includes the physical devices or resources (e.g. computers, storage systems, networks, sensors and instruments).
2. *Connectivity Layer*. This layer above the fabric layer includes the communication and authentication protocols required for Grid network transactions (e.g. exchange of data between resources and verification of the identity of users and resources).
3. *Resource Layer*. This layer contains connectivity protocols to enable the secure initiation, resource monitoring, and control of resource-sharing operations.
4. *Collective Layer*. This layer above the Resource Layer contains protocols, services, and APIs (Application Programming Interface) to implement transactions among resources (e.g. resource discovery, and job scheduling).
5. *User Application layer*. This is the highest layer and calls on all other layers for resources.

Layers 2 and 3 are sometimes considered the core middleware while layer 4 is referred to as the user-level middleware. Grid computing is based on an open set of standards and protocols (e.g. the Open Grid Services Architecture (OGSA)) [11,12]. Fig. 1 shows these five layers and their relationships.

The grid computing technology provides the user with the following types of service [8,10]. All the following services adhere to the grid architecture and concepts presented in this paper:

(a) *Computational Services* support specific applications on distributed computational resources, such as supercomputers. A grid for this purpose is often called a Computational Grid.
(b) *Data Services* allow the sharing and management of distributed datasets. A grid for this purpose is often called a Data Grid. This is the major focus of this paper and henceforth, any reference to the Data Grid will be describing these types of services.
(c) *Application Services* allow access to remote software and digital libraries, and provide overall management of all applications running.
(d) *Knowledge Services* provide for the acquisition, retrieval, publication and overall management of digital knowledge tools.

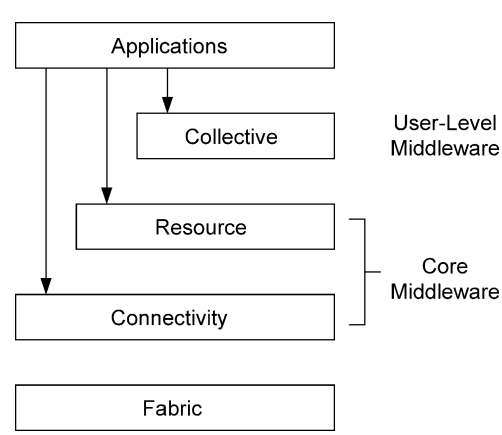


Fig. 1. The five layers of the grid technology.

Currently, there are several large-scale grid projects underway worldwide, for example, the Ninf from the Tokyo Institute of Technology; Globus from ANL (Argonne National Laboratory) and Information Science Institute (ISI), USC; Gridbus from University of Melbourne; European Datagrid, and many others [14]. However, there is only limited investigation of the impact of this emerging technology in biomedical imaging with an exception in a project called e-Diamond: a Grid-enabled federated database of annotated mammograms [15]. In this project, only preliminary thought was published mostly related to mammography as a field, which has not touched on the utilization of the power of grid computing. In this paper, the design and implementation of a Data Grid for clinical image recovery is presented. The Globus 3.0 toolkit co-developed by ANL and ISI, USC will be utilized as a guide and basis for implementing some of the software components within the Data Grid architecture [16].

4. Data grid development

4.1. Data grid and globus toolkit 3.0

The Data Grid concept presented needs both the Grid technology and PACS/DICOM technology. In particular, some PACS/DICOM resources are integrated into the five layers of the Grid architecture. For example, DICOM SCU (service class user), SCP (service class provider), and Query/Retrieve, are key components integrated within the five layer architecture. The Data Grid concept architecture with PACS/DICOM resources is shown in Fig. 2. It illustrates basic architecture already developed. The following describes some of the resources integrated within each layer.

1. *Fabric Layer*. This layer consists of a DICOM compliant fault-tolerant (FT) backup Archive Server and a PACS simulator at the research laboratory (IPI: Image Processing and Informatics); two SANs (Storage Area Network) at two PACS clinical sites; and a communications network system including LAN (local area network), Internet 2, broadband WAN (wide area network-T1), and wireless networks.

2 and 3. *Connectivity and Resource Layers (Core Middleware) using the Globus Toolkit 3.0*. This layer consists of a set of services in the Globus Toolkit 3.0 (GT3) created at the Argonne National Laboratory, and ISI (Information Science Institute), USC. GT3 is an open and free toolkit based on Open Grid Service Architecture (OGSA) mechanisms, which has the same five layer grid architecture. GT3 provides a set of services, such as Grid Security (GSI), remote job submission and control (GRAM), high-performance secure data transfer (GridFTP), Replica Location Service (RLS) and other core tools for building the Core middleware layer. GT3 and GT2 (a former version of Globus Toolkit 3.0) are common amongst a large user community in Grid research and applications.

4. *Collective Layer (User-Level Middleware)*. This layer consists of services that interact between the User Applications and the services in the Core Middleware, such as database service (to find the best available database in the Data Grid), information service (to monitor the current active services in the Data Grid), and data service (to find the physical address of the logical data) as well as other services. In this layer, GT3 has only the Information Service (Fig. 2, shaded), all others,

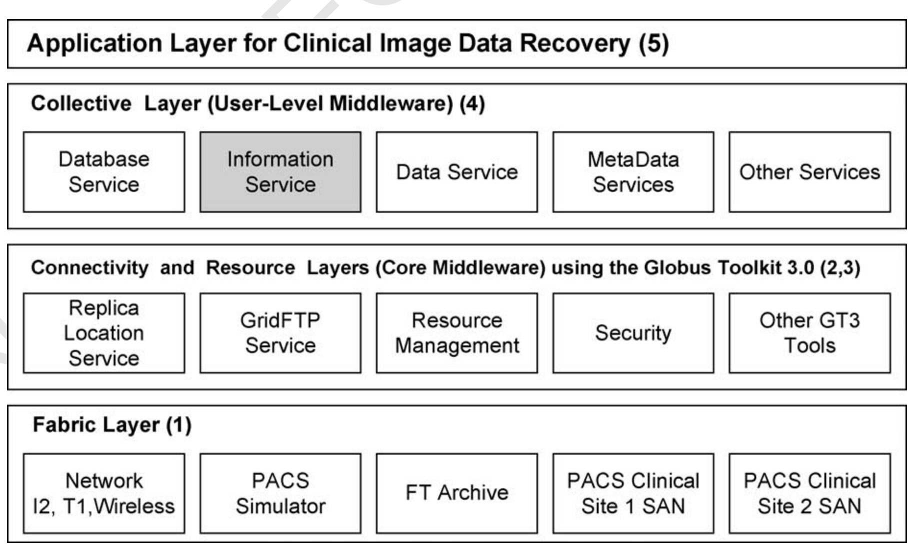


Fig. 2. Five-layer architecture and the contents of the Data Grid.

such as Database, Data, Metadata, and Other Services are currently not available. We have developed these software services in conjunction with our image data recovery application.

5. *Data Grid Application Layer*. This layer consists of several applications, such as the recovery of clinical PACS image data, and using a Personal Digital Assistant (PDA) to perform remote control workflow of clinical PACS data during the recovery.

4.2. A data grid design example and its major image data storage resources

The Data Grid Design example presented in this paper consists of three sites. However, the Data Grid design is not limited to three sites and can be expanded to multiple sites. The first site is the IPI Lab where the major resources are the PACS Simulator and the DICOM Fault-Tolerant backup Archive. Both components will be the resources in the Data Grid. The second and the third sites are the Saint John's Health Center, Santa Monica, CA (SJHC) and the Healthcare Consultation Center II (HCCII) at the University of Southern California, Health Sciences Campus. These clinical sites already have previous research collaborations with IPI Laboratory making them logical selections for the initial clinical evaluation. In addition, both sites have a clinical PACS with a SAN archive system. Fig. 3 describes the three-site configuration of this Data Grid design example. A partition of each SAN, which does not handle the clinical PACS image data, will be used as backup archive resources in the Data Grid (Fig. 3 'P2'). It is important to note that the SAN partitions belonging to each of the two sites are completely independent and the data stored in these partitions are orthogonal and separate from the clinical data partitions that are integrated with each of the respective clinical PACS. A clinical workstation outside

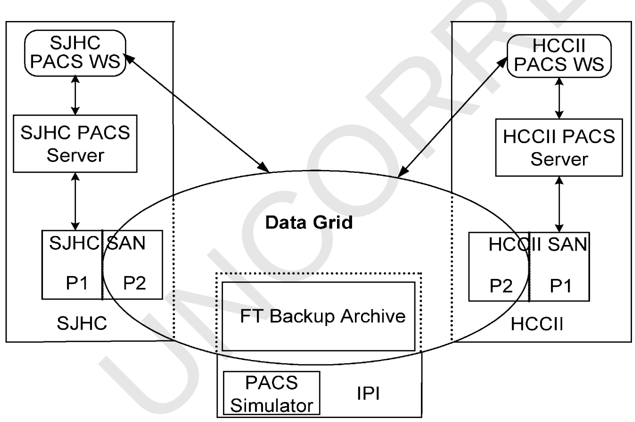


Fig. 3. Three-site configuration of the data grid: configuration of the data storage systems for three sites that comprises of image data storage resources the data grid. Workstations outside of the Data Grid can access the grid for particular services.

Table 1
Backup policy in the data grid

Site	Clinical image data	Backup copy 1	Backup copy 2
SJHC	SJHC SAN (P1)	HCCII SAN (P2)	IPI FT Backup Archive
HCCII	HCCII SAN (P1)	SJHC SAN (P2)	IPI FT Backup Archive
IPI Laboratory	PACS Simulator Archive	SJHC SAN (P2)	HCCII SAN (P2)

of the Data Grid will be able to access the grid for services using a DICOM resource.

4.3. The backup image data storage in the data grid example

Clinical images generated in the three sites, IPI, SJHC, and HCCII are backed up by the Data Grid using the following protocol. There are always two backup copies of the image data acquired from any site in the Data Grid. For example, image data acquired from SJHC PACS will have two backup copies in the Data Grid. One is stored in HCCII SAN (P2), and the second in the IPI FT backup archive. Similarly, image data acquired from HCCII PACS will be backed up in the SJHC SAN (P2) and IPI FT backup archive, and data from IPI PACS Simulator will be backed up in the SJHC SAN (P2) and HCCII SAN (P2). The IPI PACS Simulator has a connection to the clinical PACS of Childrens Hospital Los Angeles/USC for clinical image data to include a third clinical PACS site. Table 1 shows the backup policy in the Data Grid. A database in the Data Grid based on the PACS data model is used to track every patient data that comes in contact with the Data Grid.

Fig. 4 illustrates an example of the backup procedure from SJHC to the Data Grid and the steps are described below in more detail.

1. After an examination is completed at the SJHC CT scanner, the modality sends out two copies of the image

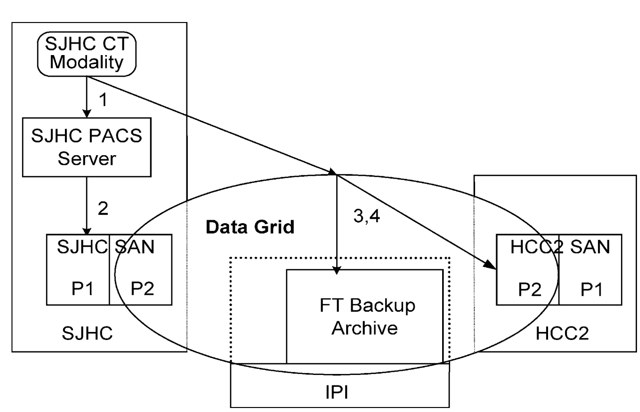


Fig. 4. The backup procedure of image data from SJHC CT Modality to two storage sites (IPI FT Backup Archive and HCCII SAN (P2)).

data, one to the SJHC clinical PACS Server, and the second to the Data Grid.

2. The SJHC PACS Server receives the first copy of the image data and stores it in the SJHC SAN (P1) as its own clinical archive.
3. The Data Grid activates a resource to receive the second copy of the image. The Data Grid monitors the status of the resource and can activate a second one if the first resource fails. After receiving the image data, the resource automatically sends two copies, one to the IPI FT Backup Archive and the second to the HCCII SAN (P2). Since the Data Grid distributes redundant copies to two different storage sites, a single-point-of failure is avoided. After the image data has been successfully archived in the two storage sites, the physical location is logged within the Data Grid and physical storage information of each copy of image data is kept.
4. The resource adds patient information of the image data (not the image data itself) to a DICOM Data Model database in the Data Grid which replicates this database and distributed to the FT backup Server at IPI, the SJHC SAN (P2), and the HCC SAN (P2). Three copies are used to avoid a single-point-of-failure of the database. We have developed the resources for this task.

4.4. An image data recovery scenario in the data grid example

Using the backup Image Data Storage and Image Recovery model in Fig. 3, we can simulate an image data recovery scenario. The scenario is that SJHC PACS archive goes down during a disaster and the physician from a SJHC workstation needs to query/retrieve the backup image data from the Data Grid immediately for review. The following is the design of a service-driven configuration and workflow for the image data recovery procedure.

Step 1: Find the Patient Information from the most optimal database copy in the Data Grid

1. A DICOM Q/R SCU (Service Class User) in a SJHC workstation sends a Query/Retrieve request to the Data Grid.
2. The Data Grid receives this request and finds from the copies of the database in the grid, which holds the desired patient record.
3. The Data Grid surveys the status (e.g., availability and latency) of all the database resources (copy of the database at HCCII SAN (P2), copy of the database at IPI, and copy of the database in SJHC SAN (P2) within the Fabric Layer) and determines the most optimal selected database location. The selected database gets the patient information and returns to the DICOM Q/R SCU in the Application Layer. This step is necessary because the user will have to select studies and images from the patient information directory, which is a standard DICOM Q/R operation from the PACS workstation.

Step 2: Retrieving Image Data from the Data Grid

4. After the user selects a set of studies from the patient directory, the DICOM Q/R SCU in the Application Layer requests the Data Grid to retrieve the patient image data from the most optimal image data database where this image data resides in terms of the best latency and efficiency.
5. Because there are two copies in the two storage sites (HCCII SAN (P2) and IPI FT Backup Archive) in the Data Grid, it determines which replica is the best choice to perform the retrieval based on network connection. It is noted that this optimal location may be different from the previous optimal location for patient information where the patient information is located because a very large amount of image data will be transferred in this latter operation. Network availability, file sizes and types will be major determined factors.
6. The Data Grid selects the best storage site (e.g., IPI Archive) and initiates a DICOM move client to request the DICOM server to transmit the file from the specified storage site to the SJHC workstation. Note that Step 2 is performed completely within the Data Grid and is seamless to the user.

Step 3: Other Data Grid Tasks

7. The Data Grid can also perform tasks to synchronize data should storage resources encounter downtime within the Data Grid.

All these services accessing the Fabric Layer will also need to be authenticated through a Security tool in the Connectivity and Resource layer (Core Middleware). There are always at least two copies of these services (including User-Level Middleware and Core Middleware) within the grid to avoid a single-point-of-failure. Image data Security methodology being developed will be used in this core middleware.

4.5. Development of applications and services: current status

Within the five-layer architecture of the Data Grid, GT3 will be used for Connectivity and Resource Layer services, and Information Services in the Collective Layer. We have completed or are planning to complete the following:

- Implement three copies of the database in the Fabric Layer to avoid the single-point-of-failure,
- Modify the IPI Archive, as well as the HCCII SAN and SJHC SAN partitions, and use some of these storage resource for the Data Grid,

- Support the connectivity of the Data Grid by connecting resources with the scalable networks,
- Implement the image data integrity algorithm in the Data Grid, and use the Grid computing resource to achieve acceptable image signature and envelope computation requirement,
- Develop the services in the Collective Layer, and modify the Applications, such as DICOM query/retrieve client/SCU, DICOM storage client, PACS Simulator, FT Backup Archive that are currently available in the IPI Lab, to support the Data Grid.

4.6. Implementation and evaluation

Phase One development will be within the laboratory research environment. This is currently being implemented. The laboratory testbed will be within the IPI Laboratory. In the laboratory environment, three simulated PACS will be implemented to evaluate the Data Grid. Resources including the PACS Simulator [17], the FT Backup Archive [5] will be used. The PACS Simulator developed at IPI has been previously used for testbed evaluations. It is a fully functioning PACS including PACS workstations, modality simulator and PACS archive server. The FT Backup Archive will simulate the SAN technology since it can be partitioned into multiple storage repositories. PACS clinical images with scrambled demographic patient data will be used to evaluate the Data Grid. To summarize, the three simulated PACS are:

1. System 1. Workstation 1 uses a PACS Simulator and its Archive Server.
2. System 2. Workstation 2 uses a PACS Simulator and the FT Backup Archive to simulate SAN technology.
3. System 3. Workstation 3 uses the PACS Simulator and the FT Backup Archive to simulate SAN technology.

The Data Grid in the laboratory evaluation is composed of three archive resources: the archive of the PACS Simulator, and two FT Backup Archive partitions. The failover scenario is to simulate a downtime to System 1 archive server and the workstation 1 will request a copy of the PACS data from the laboratory testbed Data Grid. Any system bottlenecks will be observed and collected.

Phase Two is the integration of one of the clinical sites in the design and evaluation of the Data Grid. The clinical site will replace System 1 in the laboratory testbed architecture. Again, the failover scenario will be to simulate a downtime to the clinical archive and request a copy of the PACS data from the Data Grid.

Finally, Phase Three is the integration of a second clinical site within the Data Grid to replace System 2 in the laboratory testbed environment to achieve the final configuration that was discussed and shown in Fig. 3.

5. Discussion

Utilizing the Data Grid for clinical image recovery yields the following benefits:

- Continuous Availability (CA) with 99.999% uptime.
- Automatic management, support, and recovery within the Data Grid.
- Automatic Image Recovery of PACS data that is seamless to the user.
- Applications/Users interface the Data Grid at a single entry point.
- Scalable and robust architecture utilizing open standards and heterogeneous hardware/software resources.
- DICOM compliant for clinical PACS image data.
- Cost-effective solution

For a true picture of the financial impact and savings utilizing the Data Grid for clinical data recovery, a general cost example is provided. This cost is based on current quotes from a reputable PACS vendor for utilizing SAN technology. Although it is obvious that costs may vary, this cost example provides a general view of potential savings that can be achieved. The general cost to replicate a second backup copy storage of PACS images for a community-sized hospital for one year is approximately $467,000. This cost includes the necessary server and archive hardware for 5.0–7.0 TB of data, which is the amount of PACS data generated in one year for a community-sized hospital. Utilizing the Data Grid for clinical data recovery, a federated facility of the Data Grid would only have to provide additional storage space costs to support the storage, which amounts to approximately $114,000 or 24% of the above-mentioned cost as quoted by the same PACS vendor assuming the primary PACS archive is the first price quoted above. As stated previously, in both cases, SAN technology is utilized for the cost comparisons. Therefore, a community-sized hospital could save approximately $353,000 or 76% to achieve the fully redundant data storage backup with the addition of Continuous Availability (CA) with 99.999% uptime, which a second backup copy storage may not even provide. These cost savings would only increase over time and as more and more facilities pool their resources into the Data Grid.

6. Summary

The Data Grid architecture concept presented in this paper is the result of an evolutionary process in system design. This concept has been developed based on an appreciation of the problem under consideration and an understanding of the potential power of grid computing technology. The overarching construct of a federation of PACS archives serving as cooperative backup archives for one another can be effectively realized utilizing grid technology. In this design, only a small fraction of the PACS data archive resource is needed from each federated member. Furthermore, the massive overhead burden in system design development and operation is mitigated by the public domain nature of most of the Data Grid technology, which inherently would be utilized by each of the resources that take part in the Data Grid. Essentially, any federated member can link to this Data Grid with minimal cost (as it serves as a CA backup archive for the other members) while receiving the tremendous benefit of CA image data recovery.

References

[1] Huang HK. PACS: basic principles and applications. Hoboken, NJ: Wiley; 1999.

[2] Liu BJ, Cao F, Zhou MZ, Mogel G, Documet L. Trends in PACS image storage and archive. Comput Med Imaging Graph 2003;27: 165–74.

[3] Huang HK, Cao F, Liu BJ, Zhang J, Zhou Z, Tsai A, Mogel G. Fault-tolerant PACS server. SPIE Med Imaging 2002;4685-44:316–25.

[4] Huang HK, Cao F, Liu BJ, Zhang J, Zhou Z, Tsai A, Mogel G. Fault-tolerant PACS server design. SPIE Med Imaging 2001;4323-14: 83–92.

[5] Liu BJ, Huang HK, Cao F, Documet L, Sarti DA. A fault-tolerant back-up archive using an ASP model for disaster recovery. SPIE Med Imaging 2002;4685-15:89–95.

[6] Liu BJ, Huang HK. ASP backup plan for PACS. Adv Imaging and Oncol Administrators 2003;13:55–61.

[7] What is grid computing, http://www-1.ibm.com/grid/about_grid/what_is.shtml.

[8] Grids and Grid technologies for wide-area distributed computing, Mark Baker, etc. SP&E; 2002.

[9] The grid: a new infrastructure for 21st century science. http://www.aip.org/pt/vol-55/iss-2/p42.html.

[10] Computational Grids. The grid: blueprint for a new computing infrastructure.: Morgan-Kaufmann; 1999 [Chapter 2].

[11] The physiology of the grid: an open grid services architecture for distributed systems integration. http://www.globus.org/research/papers.html#OGSA.

[12] The anatomy of the grid: enabling scalable virtual organizations. http://www.globus.org/research/OGSA.

[13] The physiology of the grid: an open grid services architecture for distributed systems integration. http://www.globus.org/research/papers.html.

[14] Bernman F, Fox G, Hey T. Grid computing. Hoboken NJ: Wiley; 2003.

[15] Brady M, Gavaghan D, Simpson A, Parada MM, Highnam R. eDiamond: a grid-enabled federated database of annotated mammograms. In: Bernman F et al, editor. Grid computing. Hoboken, NJ: Wiley; 2003. p. 923–43.

[16] Globus Toolkit 3 Core White Paper. http://www-unix.globus.org/toolkit/documentation.html.

[17] Zhou Z, Huang HK, Cao F, Liu BJ, Zhang J, Mogel G. Educational RIS/PACS simulator. SPIE Med Imaging 2003;5033-18.

Dr Brent Liu earned a PhD degree from the UCLA Biomedical Physics Graduate Program and performed research as a Post Doctorate. He is currently has a joint appointment with the Departments of Radiology, Keck School of Medicine, and Biomedical Engineering, Viterbi School of Engineering. He is also a senior research staff member of the Image Processing and Informatics Laboratory located at Marina del Rey. He has implemented fully filmless PACS in a clinical setting within the Imaging Department of both a high-profile community hospital (Saint John's Health Center, Santa Monica) and a high-profile academic hospital (UCLA) that has multiple campus sites and is currently advising multiple hospitals on their PACS process, including the USC Health Science Campus. His research areas of interest include Medical Imaging Informatics, Picture Archiving and Communication Systems (PACS) clinical uptime and usability, new PACS technology, Disaster Recovery for PACS, design and implementation of high-resolution image display workstations, next generation Internet and its clinical applications, and advances in the area of image processing and information management for healthcare including Security and HIPAA-compliance related issues.

Zheng Zhou, a PhD candidate of Biomedical Engineering Department at University of Southern California (USC), is a research assistant at Image Processing and Informatics (IPI) Laboratory, Radiology, USC now. He has 6 years experience in developments and researches on PACS and Medical Imaging Informatics. His current interest is on medical image security, grid computing for medical imaging, and medical imaging informatics.

DTD 5
ARTICLE IN PRESS

Computerized Medical Imaging and Graphics xx (2004) 1–13

Computerized Medical Imaging and Graphics

www.elsevier.com/locate/compmedimag

Informatics infrastructure of CAD system

Ewa Pietka[a,*], Arkadiusz Gertych[a], Krzysztof Witko[b]

[a]*Division of Biomedical Electronics, Institute of Electronics, Silesian University of Technology, ul. Akademicka 16, PL 44 100 Gliwice, Poland*
[b]*Central Teaching Hospital, Medical University of Silesia, ul. Medyków 14, PL 40 752 Katowice, Poland*

Received 30 March 2004; revised 14 September 2004; accepted 30 September 2004

Abstract

A computer aided diagnosis (CAD) system requires several components which influence its effectiveness. An image processing methodology is responsible for the analysis, database structure archives and distributes the patient demographics, clinical information, and image data. A graphical user interface is applied in order to enter the data and present it to the user. By designing dynamic Web pages a remote access to the entire is granted. The computer aided diagnosis system includes three layers, which might be installed on various platforms. Elements of the application software are designed independently. Integration of all components is another issue discussed in the presented paper. Implementation of a computer aided diagnosis system improves and accelerates the analysis by giving to the user objective measurement tools. It also standardizes the decision-making process and solves the problem of replicability. Finally, it permits a set of images and features to be collected and recognized as a medical standard and be applied in education and research.

Keywords: Computer aided diagnosis; Data archive and distribution; Remote data access; Bone age assessment

1. Introduction

In clinical application we deal with problems which have to be solved in a fast and objective way. However, human observation is influenced by internal (coming form the observer) as well as external (often independent of the observer) impacts. The objectivity of classification is restricted by the receptivity of human senses which are influenced by the experiences or level of training, psychological conditions (tiredness, hurry, etc.), as well as external conditions (lighting, destructive noise, etc.). A failure in perception questions the entire recognition process. The recognition process itself, influenced also by the above-mentioned conditions, may cause a slow down and/or lead to a false diagnosis. The demand of a computer added diagnostic system has become stronger and stronger.

A computer aided diagnosis system requires two components which influence its effectiveness. The first one is associated with an image processing methodology, the other deals with the data visualization or, in general, with a user-friendly graphical interface. An overall 'intelligent' workstation design becomes an important chain in the imaging informatics.

Image processing is a research area, handling the radiographs acquired by a certain radiographic procedure. It performs an image interpretation, which leads the extraction of a desired type of information. Due to the type of resulting information, image processing can be grouped into three classes:

- image manipulation which transforms the image (i.e. enhancement, suppression, matching, etc.)
- extraction of a certain anatomical region and/or features
- extraction of a certain type of information related to a medical abnormality (i.e. existence of nodules or extraction of features).

In the CAD these functions are performed automatically or they require a user interaction.

Image enhancement/suppression leads to increasing a contrast of one area with respect to another or the extraction of edges [1–3]. Another approach may separate two types of tissues (i.e. soft tissue or bony structures) using a spatially variant control function or a dual energy technique [4,5].

* Corresponding author. Tel.: +48 32 237 1040; fax: +48 32 215 21739.
E-mail address: pietka@csk.katowice.pl (E. Pietka).

0895-6111/$ - see front matter
doi:10.1016/j.compmedimag.2004.09.016

Suppression of noise artifacts and scattering effect [4,6] leads to the reduction of the unwanted noise and improves the image quality.

Extraction of a certain anatomical region is often problem-dependent and, due to some predefined configuration, can be effectively applied to one type of anatomical structures. A commonly used methodology is referred to as segmentation. Certain methods may be applied to the entire hand image or selected regions. Implemented to chest images, segmentation procedure may employ pattern recognition techniques or gray level histogram to characterize anatomical regions [7]. Other segmentation approaches find borders of one anatomical region, such as the heart borders, lungs, or rib edges. Segmentation methods developed for other anatomies include hand wrist images [8] or mammography [9]. A deep analysis of both phalangeal and carpal bone regions in the hand wrist image can be found in [10,11]. An analysis of phalangeal regions and epi-metaphyseal regions is presented in Refs. [12–14]. In Ref. [13] the author takes advantage of segmentation algorithm based on point distribution model method. These methods require user's interaction in order to indicate of bony structure to be segmented. Thodberg [15], implementing the technique of active appearance model in the segmentation of metacarpals, developed a computer aided diagnosis of osteoporosis and arthritis.

Segmentation is often followed by feature extraction stage. It may include measures reflecting the size and shape of certain anatomical structures, frequency-domain findings, statistical parameters, etc. Size measurements including heart size measure [16] in chest images, size of nodules [17], phalanges [18], carpal bones [19], size and shape of epiphyses [20], sharpness of edges after applying a wavelet transfer [21] or Fourier analysis in hand regions [22] may serve as examples.

These manually or automatically segmented images are often subjected to an analysis in which a search for certain abnormalities is performed. Many examples of investigated diseases may be given. Some of them are lung nodules [5,23], interstitial disease [24], pneumothorax [25], pulmonary opacities. Some methods have been developed for detection of masses in mammograms [26], densitometry, detection of osteoporosis [27], bone fusion. In the bone age assessment the final result is obtained by assigning the hand image into a certain stage of development. In Refs. [12,13, 28] region of interest is classified to one of eight group reflecting stage of skeletal maturity [22,29,30].

All efforts to develop a computer aided imaging diagnosis system will be useless unless an appropriate image display and graphical user interface can be implemented. A workstation design has become another important issue. Several workstation design approaches have been developed from a layer approach including application, logical, and physical layers, to radiologist-oriented workstation based on documented working interactions and task analysis to identify required procedures performed during the diagnosis. At all system design levels the functional requirements have to be specified for different clinical applications. Four typical areas of different needs are discussed most frequently:

- workstation designed for review used in intensive care unit (ICU) or by clinicians
- workstation designed for display and manipulation of gray scale multiple image studies such as CT (computed tomography) and MR (magnetic resonance)
- workstation designed for image display and manipulation with color enhancement (used in nuclear medicine and ultrasound)
- workstation designed for image display and manipulation of high resolution such as CR (computer radiography) and digitized films
- physician's desk top workstation for research, education, teaching, and review.

An additional factor which has to be considered while planning a workstation is its diagnostic implementation, related to the primary or secondary diagnosis. A primary diagnosis is performed in the radiological department as well as some clinician's offices. A secondary review station is typically used in preview rooms, operating rooms, and conference rooms. The basic distinction between primary and secondary viewing stations is related to the technical parameters of the monitor to be applied and the type of delivered images (i.e. full resolution or compressed). The primary diagnostic unit should feature higher image quality (spatial resolution, contrast, luminescence) and can be based on full resolution images only.

Many studies have performed a detailed comparison between soft copy display versus a hard copy for chest radiographs of adults or newborns, bone resorption, etc. The spatial resolution itself has been investigated by comparing soft copies of varying resolution (512, 1024, and 2048 B). A 2048-line monitor is pointed by all of these studies as the one comparable with a film. The CT, MR, and US could be viewed on a 1 KB monitor without a loss of initial resolution. However, the limitation of number of slices, displayed at a time may effect the diagnosis. This is of particular importance while comparing different studies. In clinical application four up to eight monitors have been implemented.

Functional requirements include basic software routines existing in most of commercially available as well as self-developed workstations. Image manipulation tools allow window/level, zoom/roam, flip and rotate, annotation, cine loop, viewing multiple studies, length and area measurements, magnifying glass, profile and histogram analysis, local density measurement. Standard image processing functions include edge enhancement, smoothing, 3D rendering, multiplanar reconstruction. Due to the multimedia information within a patient folder, a multimedia approach is taken for the design of a medical workstation.

The requirement of interface to other systems of hospital information is followed by audio, video, and graphics hardware for acquiring and processing of multimedia data. It includes the graphical workstation itself as well as a microphone, speaker, and video camera. The video signal offers the possibility of distributed cooperative work, remote consultation and education as well as dictating and retrieving reports.

Currently implemented CAD workstations do often not include components which might be helpful in clinical applications or teaching and training programs. Two of those missing components seem to be of big importance, at least in applications, which require a comparison with a set of reference cases. The database would then be required for storage and distribution of these cases. Remote access to the data as well as processing server would make the system opened to other health care centers. A structure of the CAD system is discussed in the current study.

2. Informatics infrastructure components—overall schema

Computer aided diagnosis (CAD), being basically a tool, which assists the radiologists in performing the image diagnosis, has recently switched to a more sophisticated application system. Image processing functions as well as pattern recognition procedures require often a set of data which needs to be stored in a database system as a set of standards or data repository. Interactions, performed by physicians during the diagnostic process, draw the attention towards a user-friendly graphical interface. Integration of image processing and manipulation tools, the database structure, and graphical user interface, forces a multilayer CAD system to be designed. In this study a three-layer system is presented. It contains (Fig. 1) a data layer, an application layer, and a presentation layer.

At the first layer a database management system (DBMS) is selected. It is responsible for the data archive and distribution. There are certain features that the database engine has to provide. They do not differ significantly from a general database system, yet some of them are of a particular importance.

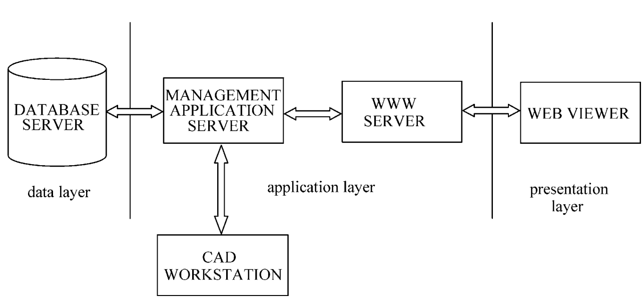


Fig. 1. Multilayer structure of CAD system.

Authorized access to the data, shared among various users and processes, has to be granted remotely. Security constrains define rules to be checked whenever access is attempted to sensitive data. Different constrains need to be established for each type of access (retrieve, insert, delete, etc.) to each piece of information in the database.

Inconsistency need to be avoided particularly when a standard set of data is built. All data updates have to be propagated. Integrity (i.e. ensuring that the data in the database is correct) should be maintained. Centralized control of the database can help in avoiding such problems by permitting the data administrator to define (and the DBMS to implement) integrity constraints.

Discussing the database structure one has to keep in mind, that a medical database may contain various types of information. Beside the numeric and alphanumeric data, coded in a string format, date and hour format, etc. patient record contains also images (compressed and uncompressed), regions of interest extracted from a diagnostic image, processed images, with extracted features, biomedical signals again with extracted features, etc.

The application layer contains three elements (Fig. 1): a management application server, a WWW server, and a CAD workstation. Tasks to be performed by the application layer include the database access (text and images), data presentation granted by the graphical user interface (GUI), and a remote access to the entire CAD system permitted by designing dynamic Web pages. The CAD workstation performs the image processing procedures. It usually contains two steps. The first one suppresses the background noise and removes undesired structures or enhances the regions processed or viewed at the following stage. Then, selected features are extracted and subjected to a recognition procedure. It yields the pattern assessment to a certain class. The results being stored in the database, transferred to the management application server and via the WWW server are accessible through the Web viewer.

Presentation layer—by means of a Web viewer, permits an easy, remote access to the system. At a user site, it serves as an interface to the entire system. The informatics infrastructure is discussed by using an example of the computer assisted bone age assessment. Three levels of the system are shown. The data is organized into a relational database. The software application permits an access to the data and its adjustment. The image processing system results in extracted features as a basis for the assessment of skeletal maturity.

The organization of the paper follows the implementation order of phases of the entire project. Usually the image analysis software (CAD) is being developed first. Then, the database appears to be necessary. After the design is finished, a graphical user interface is applied in order to enter the data and present it to the users.

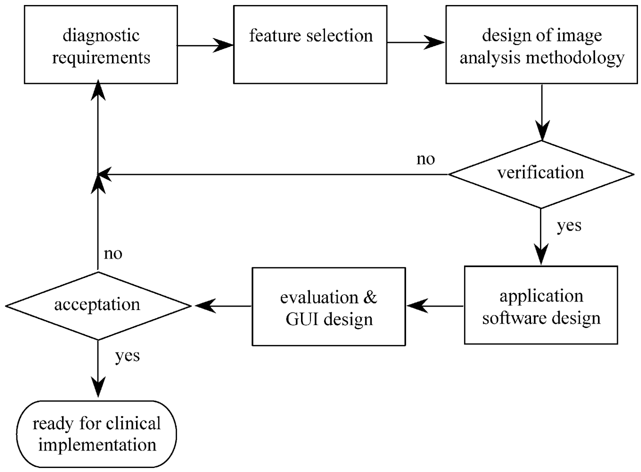


Fig. 2. Phases of the CAD workstation design.

3. CAD workstation

A design of a CAD workstation (Fig. 2) starts with an overall analysis of the diagnostic process. Then, features, important in the diagnostic process, are selected. Next, the image processing methods are verified considering the design possibility and performance efficiency of algorithms able to extract these features. Since this stage is a compromise between the diagnostic requirements and the possibility of their fulfillment, this loop may be circled several times. After a processing schema has been accepted, the methodology is designed and evaluated based, firts, on testing cases and then, clinically. Meanwhile, a graphical user interface is designed. As an example of the CAD workstation structure, a computer assisted bone age assessment is presented.

The bone age assessment analysis is a procedure frequently performed in pediatric radiology. In the medical diagnosis a left hand wrist radiograph is compared with atlas patterns. The comparison is based on the entire hand image [31] or selected regions of interest [29]. The choice of a corresponding pattern assigns the bone age. Due to the radiological analysis of the hand radiogram, areas of biggest importance are located at the joints. Epiphyses, increasing in size and changing its shape, reflect the stage of development in children before puberty (Fig. 3(a)–(c)). Later, the stage of fusion between the epiphyses and metaphyses (Fig. 3(d) and (e)) delivers diagnostic information.

According to Kirks [32] statement: *the more distal the better*, the distal and middle joints are to be analyzed. Other approaches consider also carpal bones, yet, their analysis requires robust and time consuming procedures without offering any persistent diagnostic information at the later stage of development [19]. These diagnostic requirements have pointed the epiphyseal areas as regions of interest to be processed for the extraction of findings describing the shape and size of epiphyses and stage of their fusion with the metaphyses.

The image analysis starts usually with its standardization that includes the background suppression, removal of markers, labels, and artifacts. The latest is related to an inaccurate focus of X-ray beam above the cassette. The noise suppression methods applied to the entire image should permit the body area to be separated from the background. At these stage only general methods, neither very time consuming nor requiring heavy computation power, may be applied. Then, the body area (in our application the hand region) has been separated from the background [33]. Shrinking the analyzed area to the hand location, the epiphyseal regions of interest (EROI) of the II–IV phalanges are located. The flow chart of the EROI processing is shown in Fig. 4. The implementation of filters is optional, yet a study has shown [34] that this step shortens the segmentation time by 50%.

Extracted regions of interest include two types of anatomical structure. One refers to soft tissue, the other one to bones. Before separating both anatomies by means of segmentation procedures, the region has to be subjected to a more robust than at the preprocessing stage, noise suppression procedure. Since, the exposure parameters as well as the acquisition track are unknown and changeable from one image to another, it has been assumed that each region is a summation of uncorrupted image and the additive mixture of impulse and zero mean Gaussian noise.

Due to the lack of any information about image acquisition system, the image noise distribution is assessed by using sample image regions surrounding the hand area (Fig. 5). For 100 image samples a Kolmogorov α-test with the confidence level of 0.5 has been performed. Due to the assumption that radiographs are distorted by Gaussian noise with unknown variance, there is no reason to reject a null hypothesis about normal noise distribution. The presence of impulse noise components is occasionally caused by scratches and non-uniformly spread photographic emulsion.

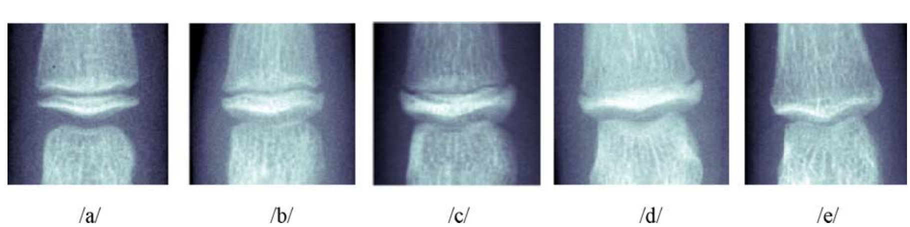


Fig. 3. Stages of development /a/, /b/, /c/ early stage, /d/, /e/ later stage.

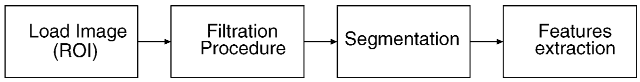


Fig. 4. The flowchart of the EROI processing.

The influence of this type of noise can be reduced by a median filter [35] that replaces each pixel by the median of its neighborhoods. All eight neighborhoods may be chosen, yet applying a star shaped area requires fewer calculations. The additive noise is suppressed by an adaptive Lee filter [36]. Parameters of the Lee's filter such as local gray level mean and local variance are estimated from the input image and delimited by a filter with a 3×3 mask. The Lee's filtration procedure requires also the variance of noise to be known. If it is not given, the algorithm uses an estimated variance. In this approach a noise estimator [37] has been implemented. The local image variance in a 3×3 window shifted over the filtered image is calculated. The noise variance is then proportional to the mean value of image gray levels filtered with a second order derivative operator [37,38].

The role of a median filter is to suppress the impulse noise ('salt and pepper'), however, Lee's filter smoothes the non-uniformity within the soft tissue and bony structure. The filtration procedures reduce both impulse and additive Gaussian noise.

Image segmentation is a difficult and very important problem in image processing. Its accuracy influences the overall image analysis and pattern recognition. The method, applied to certain applications, dependents on the type of images. Statistical approaches based on the minimum mean-squared error criterion, maximum likelihood, and maximum a posteriori estimation is used in the current study. Gibbs random field techniques have been applied to estimate the maximum a posteriori probability [39].

The segmentation separates the bony structure and cartilage from soft tissue. The procedure consists of two steps: (1) a preliminary clustering step applying the c-means algorithm and (2) segmentation with Gibbs random fields and estimation of the intensity function [40,41]. This procedure [42] yields a binary image in which the bony structures appear as white objects, whereas other areas (background and soft tissue) become black. Segmentation procedure requires the noise variance to be known. An inaccurately set value may cause a segmentation failure. The segmentation method has been adjusted to the hand structure by estimating the adaptive noise variance [37,38], and by a hierarchical multiresolution implementation (HMR) of a segmentation algorithm (Fig. 6) [40]. The HMR solution takes advantage of the pyramid construction of images at different resolutions by filtration and image decimation by a factor of two. At each level of the pyramid, the algorithm uses the previously segmented image generated at the lower level and enlarged by a factor of two. The noise variance at the lower resolution is calculated as half of the noise variance value found at the higher resolution. In this study a three-level-pyramid is performed on a filtered image (Fig. 6). A cascade of previously applied median and Lee filters preserves the edges, removes impulse artifacts and smoothes the tissue structure. In comparison to its non-hierarchical version, such approach significantly reduces the execution time [41]. As a result a binary image is obtained. The separation of epiphyses and metaphyses is followed by extraction of features which reflect the size and shape of epiphyses [42]. The analysis yields good results in children before puberty. When the epiphyseal fusion starts, the separation of both structures is impossible. At this stage a wavelet transform is implemented. The horizontal detailed component of the second level of decomposition yields a set of features reflecting the degree of fusion [42].

The final stage in the computer aided diagnosis is to assign the image to a certain class or to a few classes with a degree of membership if a fuzzy logic is applied. Two methods have been implemented in the computer aided bone age assessment. The first one corresponds to the atlas matching approach [34]. Features, extracted from the diagnosed image are compared with corresponding findings form the reference set. The comparison is based on various metrics. Three of them (i.e. Euclidean (ρ_1), Minkowski (ρ_2), and city block (ρ_3) normalized distances metric have been tested [43]. The recognition process consists of two steps: first a raw assessment indicates three closest age groups. Then, four closest matches are searched within selected age groups. The results are shown in Section 6.

The second classifier is based on the Mamdani fuzzy reasoning system [44]. The knowledge is represented as

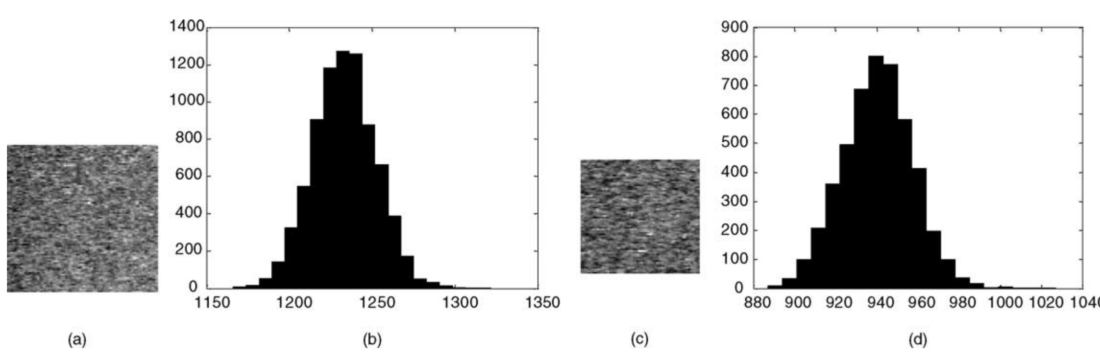


Fig. 5. Background noise distribution in samples of hand radiographs: (a), (c) background areas and their histograms (b), (d).

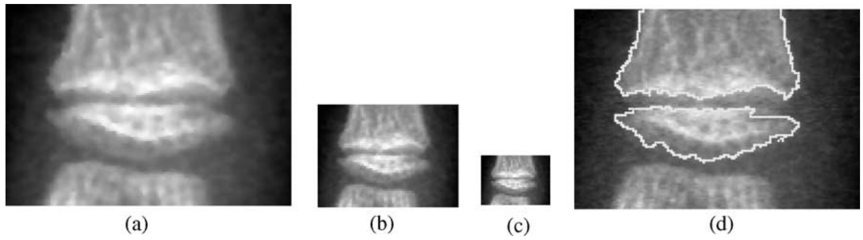


Fig. 6. The pyramid of filtered EMROI: (a) filtered input image (highest resolution), (b) middle resolution, (c) lowest resolution, (d) segmented image with extracted contour.

a set of *if-then* rules linking the features with the skeletal age. A connection between input variables and implications is interpreted by a minimum operator whereas, the aggregation of rules applies the maximum operator. A quantization of the input and output space is based on age groups just for the sake of simple natural linguistic interpretation. Since the discrimination power of a feature may differ in various age groups, the domains of input variables are divided into a fuzzy set describing feature values corresponding to one or several consecutive age groups. Fuzzy sets of the output variable (refer to as the bone age) represent classes corresponding to the years of age. More details can be found in [42,45].

A clinical implementation of a CAD workstation requires also some functions which prevent failure of the image analysis due to unpredictable events. First, the image standardization has to be verified. Each image-processing algorithm is able to recognize certain image content. In clinical practice the standardization of the image content causes many problems. A CR image may be rotated at any angle or flipped. Artifacts, markers or transparent borders caused by blocking of the collimator that touch or overlap the anatomical structure may lead to a failure of the image analysis. Non-standardized images are rejected. The second group of errors detected results from the image processing procedures. The error detectors verify the segmentation result and the range of extracted features.

4. Database structure

In the computer assisted diagnosis of medical images the data structure is heterogeneous. It contains alphanumerical structures as well as imaging structures. The alphanumeric data usually describes patient demographics as well as clinical information required in a particular diagnostic field served by the CAD. The data is transferred from other information systems or extracted from the image data at the phase of image processing. The image data contains diagnostic images written in DICOM (Digital Imaging and Communication in Medicine) or compressed (for example JPEG) images. In some cases a storage and distribution of selected regions of interest might be required.

A database design contains two steps. The database design starts with a logical phase and is followed by a physical phase. Although decisions made at the physical level may have impact back on the logical level, yet in this study the logical structure has been made without paying any attention to the physical level. After the logical design is completed, further normalization is performed to build the physical structure.

While starting the database design we are primarily concerned with what the data *is* rather than how it will be *used* [46]. This means, that at the design time the data is application-independent, i.e. the data is hardware-independent, operating-system-independent, DBMS-independent, user-independent, language-independent, etc. This means also, that no compromises for performance reasons will be made.

The first step at this level requires a definition of the entire set of data that is to be stored in the database. Then, a functional dependency is being built. Since in the bone age assessment a comparison between the left hand radiogram and atlas patterns is performed, the database contains a set of cases considered as a medical standard. The data can be used during the manual image analysis performed by the radiologist, training, and consultation. It should be ready to serve as a reference data set at any time and any place it is required by any user (physician or student). The database will also be searched for cases by a computer system, which performs a pattern recognition procedure. The search may be based on an atlas-like analysis in which features or image pattern are compared. Another method of comparison is based on a pattern classifier available after a training session is completed. In this case the data is entered at the training level. In both modes an adjustment of the data is required and performed by a data administrator.

Due to the medical requirements, two types of cases are being considered for storage in the database. One is the standard set used as a reference in medical diagnosis. The other group includes clinical cases for research and education purpose.

The data is organized into groups related to the information requited and collected during the clinical examination. The first group contains demographic data, i.e. birth date, the chronological age at the time the exam has been performed, gender, and race. For reference cases no patient ID is stored. Yet, if data is considered in the clinical

race	gender	age group	chronological age	birth date	height	weight	trunk	Tanner	diagnostic image	compressed image	ROI1	SoD1	FV[11]
												SoD2	FV[11]
											ROI2	SoD1	FV[11]
												SoD2	FV[11]
											ROI3	SoD1	FV[11]
												SoD2	FV[11]
											ROI4	SoD1	FV[11]
												SoD2	FV[11]
											ROI5	SoD1	FV[11]
												SoD2	FV[11]
											ROI6	SoD1	FV[11]
												SoD2	FV[11]

Fig. 7. Data structure. ROI (Region of Interest), SoD (Stage of Development), FV[11] (Feature Vector).

(diagnostic) procedures or in the electronic patient record, patient ID should be added to the database.

The second group includes clinical data: height, weight, trunk, Tanner coefficient. The third group is related to the data required by the image processing system: age group, diagnostic and compressed images, identifier of the region of interest, and features extracted from that region of interest.

The following step in the logical design is to set the functional dependency. It is the many-to-one relationship from one set of attributes to another. The data structure is shown in Fig. 7. The set of data for a particular case includes the demographics, diagnostic image and compressed image. Then, a several number of regions of interest are assigned to one image. Next, two stages of development may be considered. Each of them requires a set of features f1 to f11 extracted from a region of interest [42].

After the information to be stored in the database has been defined, the data is subjected to three normalization procedures. Each decomposition step is reversible. This ensures a lossless final decomposition.

The data has been subjected to a normalization procedure. First three normal forms are considered. In the first normal form [46], in every legal value of the relation variable, every tuple contains exactly one value for each attribute. The second normal form (2NF) turns the structure into a configuration in which every non-key attribute is irreducibly dependent on the primary key. In the third normal form (3NF) every non-key attribute is non-transitively dependent on the primary key.

A function dependency has been generated. Three main branches have been obtained. First, a case record variable is projected onto the patient's data. Then, a join between image and the case record, compressed image, and a region of interest is created. Finally, the region of interest is linked with the features vector.

5. Application layer

Application layer supports four areas which permit a CAD to be implemented in a clinical practice. First, the image processing software is responsible for accurate analysis of the diagnosed image. Then, extracted features and selected regions of interest as well as the image itself or its compressed version is archived in a database and accessed by authorized users in a secure mode. Next, the communication between the image processing procedures and database server need to be managed. Finally, a user-friendly interface is required to support the interaction and display of images and the processing results. If an open architecture is to be ensured, DICOM compliance interface is necessary to retrieve images form a PACS (Picture Archiving and Communication System) server.

Two main components are implemented to support these functions: the image analysis server and the management application server (Fig. 8). In order to preserve an open architecture, the image processing application server has been separated from the management server. It permits various image analysis systems to be incorporated in the application layer, yet requires a standardization of the communication process. Since the CAD workstation has already been presented (Section 3), three other components are to be discussed. First, the data management and communication with the database server satisfies the storage requirement. Then, the communication between the management application server and the CAD workstation solves

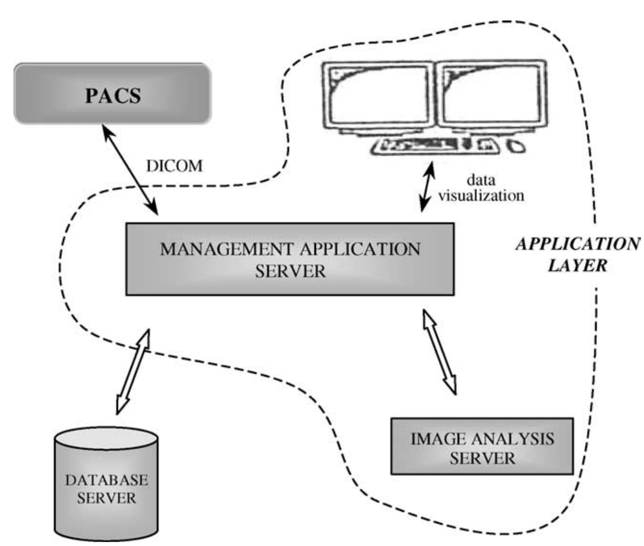


Fig. 8. Application layer structure and functions.

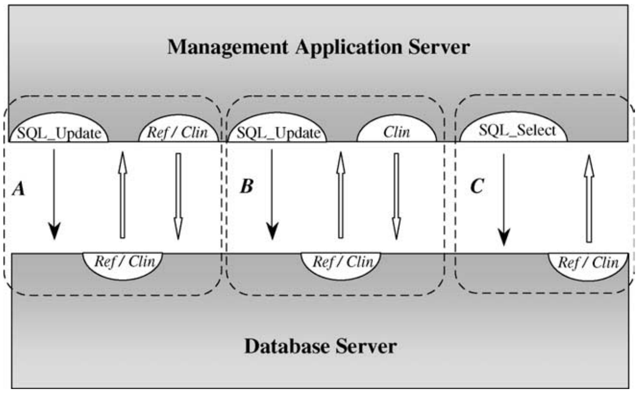


Fig. 9. Communication between Management Application Server and Database Server for three groups of users (see also Table 1). Ref and Clin stand for reference and clinical cases, respectively. **A**. Data administrator, **B**. Radiologist, **C**. Student.

the data transfer problem. Finally, the graphical user interface gives a user friendly access to the entire system.

The data management server authorizes each access to the database. As a standard the SQL (structured query language) is used. Three types of the data access mode are required. Two of them support the update and require the *SQL update* query (Fig. 9). They permit new data to be added into the database. The third access mode permits a read only mode and grants to the user the *SQL select* query only.

As it has already be mentioned, two types of cases are stored in the database. One is the standard set used as a reference in medical diagnosis. The other group includes clinical cases for research and education purpose. The significance of both types of data differs. When a clinical case is added to the database, no changes are made in the overall performance of the recognition system. It is not the case, when a reference set of data is adjusted. A case, added to the standard set, influences the classification procedure. A wrong update may lead to a loss of accuracy in the medical diagnosis. Thus, the users have to be organized into two groups with two various access permissions. One group, usually radiologists who perform the image analysis (group B in Fig. 9), can insert clinical cases, pathological or with no pathology, yet, cannot appoint them as reference cases. The second group of user (data administrators) has the authority to assign a case as a standard and update the reference set of data. They also share the responsibility for the accuracy of the result of the image classification analysis (group A in Fig. 9).

The read only mode is given to students and remotely logged users (group C in Fig. 9) who may only review the data. The range of features describing a particular class is also available. The data access modes are summarized in Table 1.

At the current stage of development the communication between the management application server and the CAD workstation is based of the push mode (Fig. 10).

Table 1
Access modes of various groups of users

User group	Description
A. Data administrator	Read/write mode to reference and clinical data set
B. Radiologist	Read/write mode to clinical data set only (no insert of reference image)
C. Student	Read mode only

A diagnostic image is transferred to the CAD workstation and its ID is written in a *first-in-first-out* queue. After being processed, the compressed image, located regions of interest, and extracted features are communicated back and stored in a buffer. In order to notify a user that new results have been delivered, a worklist is updated (Fig. 11).

The main goal of a graphical user interface (GUI) module is to guide a user through the diagnostic process in a quick and friendly manner and deliver necessary date at the time it is required. Several functions are performed (Fig. 12).

Images are to be transferred from a PACS server, an acquisition modality (computer radiography, digital radiography, computer tomography, ultrasound, etc.), or from a scanner. In order to communicate with PACS components, a DICOM compliant module is installed. It supports the send/receive as well as the query/retrieve DICOM functions [47].

The images are then transferred to the CAD workstation. An operation send/receive is used. After being processed, all extracted components are communicated to the management application server. The data record is stored in a buffer. Adjustment of the worklist notifies the user that new results are available for verification.

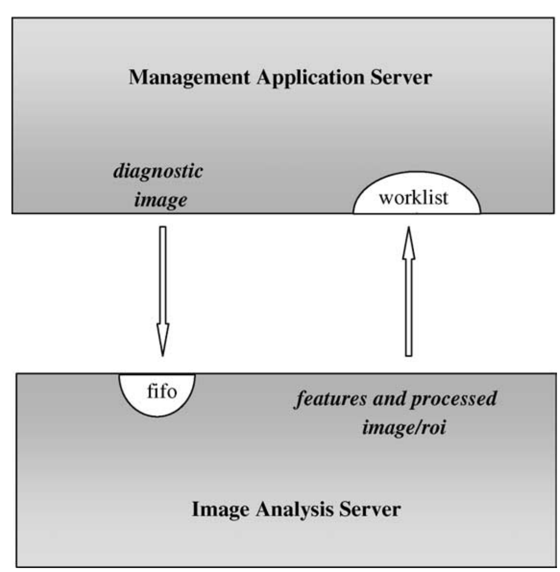


Fig. 10. Communication between Management Application Server and CAD workstation.

Fig. 11. Worklist at the Management Application Server.

A query/retrieve function permits the data to be displayed at the user site in various formats. The entire reference data set can be scrolled (Fig. 13) or a certain class of images may be selected. Images are shown in a compressed JPEG format. A display of the entire set of images belonging to one class opens up the possibility of performing an inter- and intra-class comparison study. If diagnostic quality is desired, a click on a certain image results in a DICOM image display. Due to the resolution, only one image can be viewed at a time.

After the image analysis is completed and the results are transferred to the management application server, the image data together with the selected regions of interest, extracted features, and patient data are displayed (Fig. 14). For comparison, four closest matches from the reference data set are shown. Each of them can be displayed together with the extracted features. This presentation of data assists the radiologist in performing a bone age assessment. The result, given by a computer analysis, has to be approved and may be corrected or abandoned and performed manually by the radiologist.

Once the results are being accepted, the data is stored in the database. The marker is set according to the type the case. If the case is assumed to be normal it becomes a candidate for a reference set of data. Yet, only a data administrator has got the permission to update the reference data set.

The management application server may also be applied just for reviewing a certain group of cases (including the reference set). All users, who have the access permission, may use this option. Since it does not permit the database update, a limited set of data could also be delivered in the Internet.

Another useful path is to adjust the standard set of data according the former analysis. If a data record has become a candidate for the reference set adjustment, after a double check, its marker may be changed (by a data administrator only).

6. Results

The evaluation of the image analysis algorithm is performed at two stages. At the first one the accuracy of measures and extracted findings are tested. If various

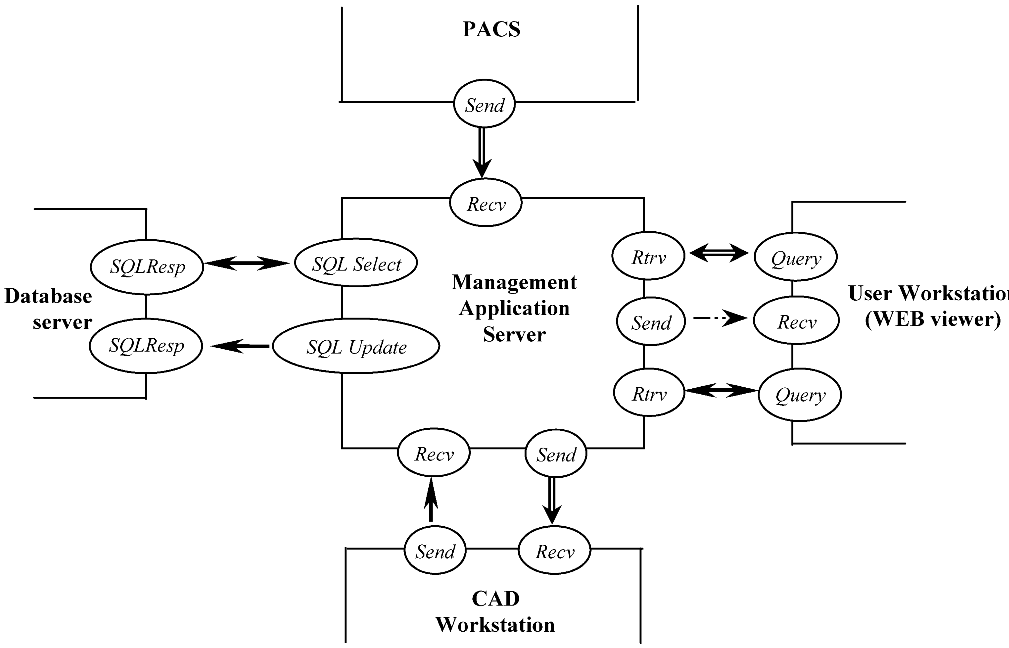


Fig. 12. Functions performed by Management Application Server.

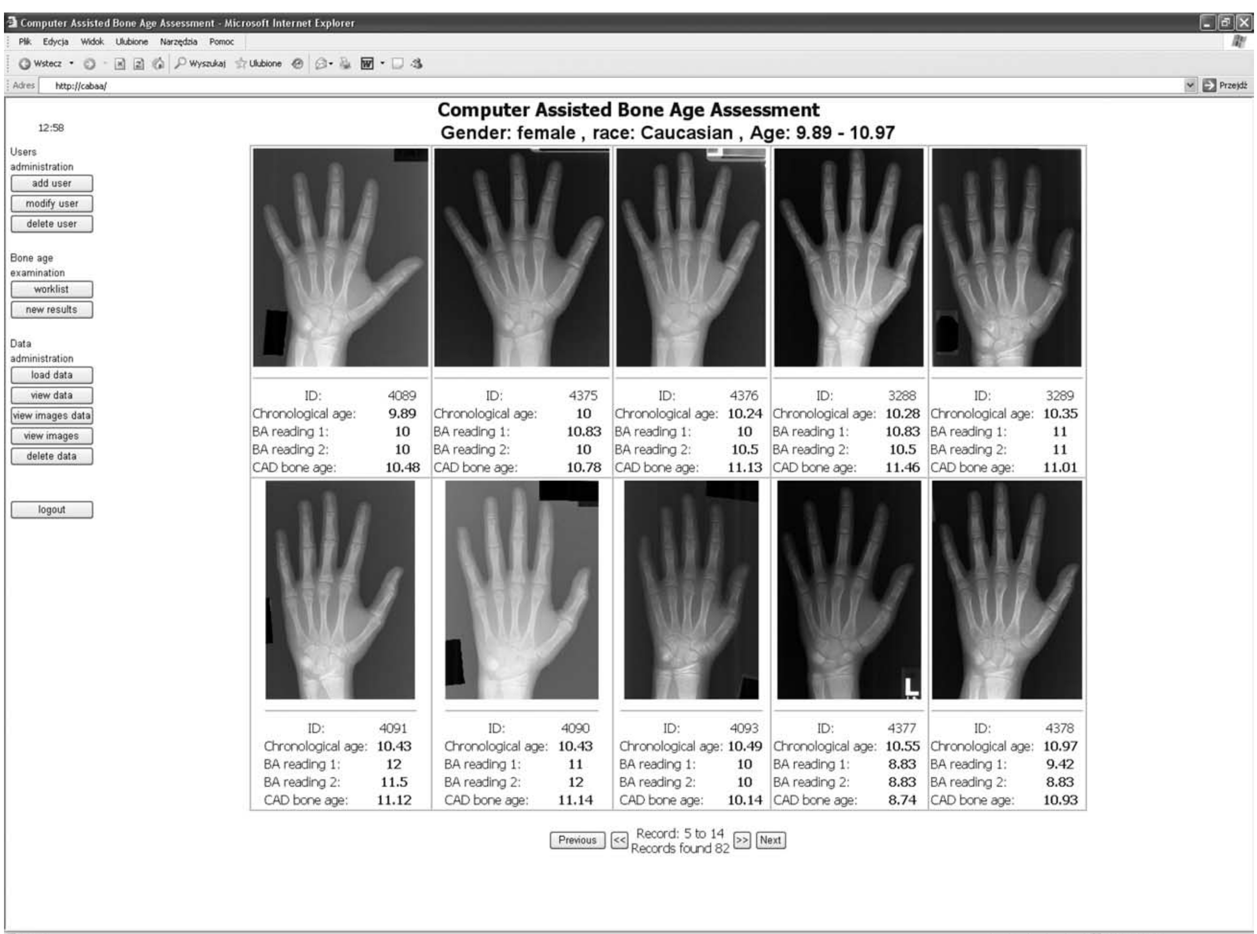


Fig. 13. Reference data set display.

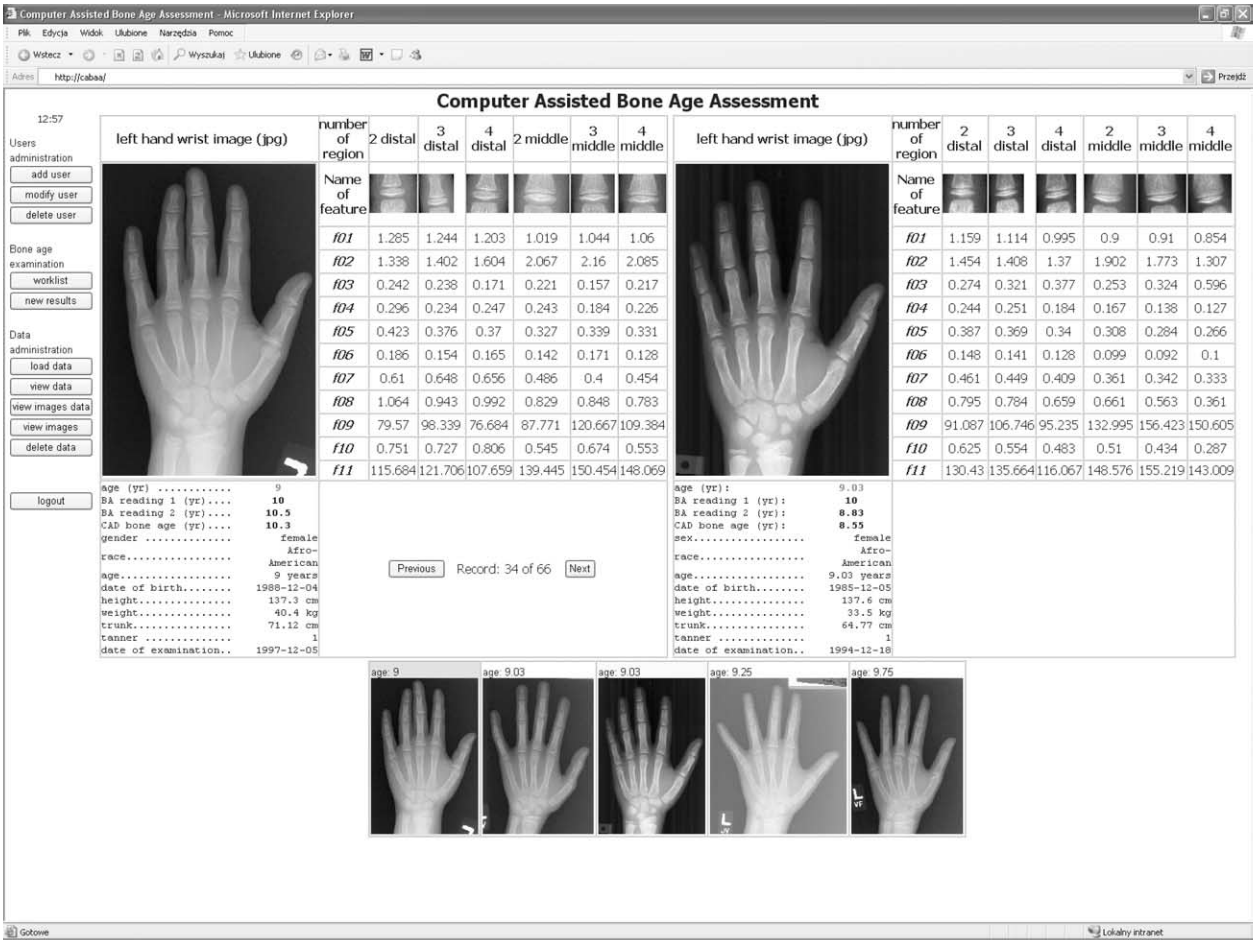


Fig. 14. Image analysis results displayed together with four best matches (see JPEG file).

methods are compared, the features extracted by means of these methods might also be compared with each other.

In the bone age assessment each region of interest analysis results in a feature vector of eleven elements. Comparison of two segmentation approaches applied to the ephiphyseal region of interest (EROI) is presented in Table 2. The evaluation involves 1851 automatically extracted regions of interest. As the most accurate method the adaptive segmentation with hierarchical multiresolution approach has been found. Better results have been found in all tested parameters. Implementation of the multiresolution pyramid reduces also the segmentation time that influences the measurements performed on overall regions of interest.

The second stage evaluates the classification result. For this purpose, the image database has been randomly divided into two parts. One group of images has served as a testing set, the other as a reference set. Since the images originate from a normal population, the bone age corresponds to the chronological age. Results of classification accuracy versus used metrics are presented in Table 3. A lower than 1.5 year difference between the assessed bone age and the chronological age for all metrics (i.e. Euclidean (ρ_1), Minkowski (ρ_2), and city block (ρ_3) normalized distances metric has been noticed in 70% cases, whereas, in less than 4.5% the difference has exceeded 3 years. This relatively high

Table 2
Results of features extraction accuracy and relative processing time

Segmentation algorithm	% of regions with correctly calculated features	% of regions with no features	% of regions with incorrectly calculated features	Relative calculation time
Adaptive	86.5	3.5	10	1.66
Adaptive with hierarchical multiresolution (HMR)	91.1	1.5	7.4	1

Table 3
Classifications results

Bone age (BA) distance	BA≤0.5 (%)	0.5<BA ≤1 (%)	1<BA ≤1.5 (%)	1.5<BA ≤2 (%)	2.5<BA ≤3 (%)	BA>3 (%)
ρ_1	28.9	22.6	17.3	12.8	13.9	4.5
ρ_2	25.7	23.8	18.9	12.4	15.1	4.1
ρ_3	27.1	22.6	17	15.1	13.5	4.5

variance reflects also the distribution of developmental stage in normal population.

The implementation of computers in the medical diagnosis improves the analysis in three different areas. Firstly, it increases the objectivity by using quantitative features instead of a visual interpolation. These features are extracted automatically rather than manually in order to accelerate the analysis and make the measurements independent of the external conditions (light), measurement inexactness, or errors (calibration errors, etc.). Secondly, a transparent algorithm is defined to standardize the decision-making process in the feature analysis. This again increases the objectivity and solves the problem of replicability. Finally, it permits a set of images and features describing them to be collected and recognized as a medical standard.

7. Summary

An overview of the informatics infrastructure of a computer-added diagnosis in presented. It permits an integration of image processing and manipulation tools, the database structure, and graphical user interface, forces a multilayer CAD system to be designed. In this study a three-layer system is presented. At the first layer a database management system (DBMS) is selected. It is responsible for the data archive and distribution. Then, the application layer contains three elements (Fig. 1): a management application server, a WWW server, and a CAD workstation. Tasks to be performed by the application layer include the database access (text and images), data presentation granted by the graphical user interface (GUI), and a remote access to the entire CAD system permitted by designing dynamic Web pages. Finally, the presentation layer—by means of a Web viewer, permits an easy, remote access to the system. At a user site, it serves as an interface to the entire system. The informatics infrastructure is discussed by using an example of the computer assisted bone age assessment.

References

[1] Canny J. A computational approach to edge detection. IEEE Trans Pattern Anal March Intell 1986;PAMI-8:679–714.
[2] Petron M, Kittler J. Optimal edge detectors for ramp edges. IEEE Trans Pattern Anal March Intell 1991;PAMI-13:483–91.
[3] Sanada S, Doi K, MacMahon H. Image feature analysis and computer-aided diagnosis in digital radiography. Automated delineation of posterior ribs in chest images. Med Phys 1991;18:964–71.
[4] Wagner RF, Macovski A, Nishimura DG:. Effect of scatter in dual-energy imaging: an alternative analysis. IEEE Trans Med Imaging 1989;8:236–44.
[5] Kelcz F, Zink FE, Peppler WW, Kruger DG, Ergun DL, Mistretta CA. Conventional chest radiography vs. dual-energy computed radiography in the detection and characterization of pulmonary nodules. AJR 1994;162:271–8.
[6] Barnes GT, Wu X, Sanders PC. Scanning slit chest radiography: a practical and efficient scatter control design. Radiology 1994;190: 525–8.
[7] Sherrier RH, Johnson GA. Regionally adaptive histogram equalization of the chest. IEEE Trans Med Imaging 1987;MI-6:1–7.
[8] Michael DJ, Nelson AC. HANDX: a model-based system for automatic segmentation of bones from digital radiographs. IEEE Trans Med Imaging 1989;8:64–9.
[9] Lucier BJ, Kallergi M, Qian W, et al. Wavelet compression and segmentation of digital mammograms. J Digital Imaging 1994;7: 27–38.
[10] Efford ND. Knowledge-based segmentation and feature analysis of hand-wrist radiographs. School Computer Studies Research Report Series. Report no. 94.31; 1994.
[11] Snel JG, Venema HW, Grimbergen CA. Detection of the carpal bone contours from 3-D MR images of the hand wrist using a planar radial scale-space snake. IEEE Trans Med Imaging 1998;17(6):1063–72.
[12] Wastl S, Dickhaus H. Computerized classification of maturity stages of hand bones of children and juveniles. Proc IEEE EMBS, 1996. 4.5.3: Image Pattern Analysis. p. 1155–6.
[13] Al-Taani AL. Classification of hand bones for bone age assessment. Proc IEEE EMBS, vol. 2; 1996. p. 1088–91.
[14] Piętka E, McNitt-Gray MF, Huang HK. Computer-assisted phalangeal analysis in skeletal age assessment. IEEE Trans Med Imaging 1991; 10:616–20.
[15] Thodberg HH. Hands-on experience with active appearance models, Proc SPIE, 2002.
[16] Nakamori N, Doi K, Sabeti V, MacMahon H. Image feature analysis and computer-aided diagnosis in digital radiography. Automated analysis of size and lung in chest images. Med Phys 1990;17:342–50.
[17] Glatt LA, Longbotham HG, Arnow T, et al. Application of weighted-majority minimum-range filter in the detection and sizing of tumors in mammograms. Proc SPIE 1992;1652:477–88.
[18] Pietka E, McNitt-Gray MF, Huang HK. Computer-assisted phalangeal analysis in skeletal age assessment. IEEE Trans Med Imaging 1991; 10:616–20.
[19] Pietka E, Kaabi L, Kuo ML, Huang HK. Feature extraction in carpal-bone analysis. IEEE Trans Med Imaging 1993;12:44–9.
[20] Gertych A. Epiphyseal diameters in computer aided bone age assessment. International workshop on control and information technology, Ostrava, IWCIT'99; 1999. p. 178–84.
[21] Pośpiech-Kurkowska S, Gertych A, PiętkaE, Cao F, Huang HK. Wavelet decomposition based features in description of epiphyseal fusion, analysis of biomedical signals and images. Proceedings of 15th Euroconference Biosignal 2000, Brno; 2000. p. 246–8.
[22] Tanner JM, Gibbons R. A computerized image analysis system for estimating tanner-whitehouse 2 bone age. Horm Res 1994;42:282–7.
[23] Giger ML, Ahn N, Doi K, MacMahon H. Computerized detection of pulmonary nodules in digital chest images: use of morphological filters in reducing false-positive detection. Med Phys 1990;17:861–5.
[24] Asada N, Doi K, MacMahon H, et al. Potential usefulness of an artificial neural network for differential diagnosis of interstitial lung diseases: pilot study. Radiology 1990;177:857–60.
[25] Sanada S, Doi K, MacMahon H. Image feature analysis and computer-aided diagnosis in digital radiography. Automated detection of pneumothorax in chest images. Med Phys 1992;19:1153–60.
[26] Wu Y, Giger ML, Goi K, et al. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. Radiology 1993;187:81–7.
[27] Roeske JC, Giger MG, Dixon LB, Doi K. Computerized analysis of osteoporosis on bone radiographs. Radiology 1990;177(P):317.
[28] Dickhaus H, Habich R, Wastl S, Maier C, Gilli G, Schonberg D. A PC-based system for bone age assessment. Proceeding of the EMBEC'99; 1999. p. 1008–9.
[29] Tanner JM, Whitehouse RH. Assessment of skeletal maturity and prediction of adult height (TW2 method). London: Academic Press; 1975.

[30] Tanner JM, Gibbons RD. Automatic bone age measurement using computerized image analysis. J Pediatr Endocrinol 1994;7:141–5.
[31] Greulich WW, Pyle SI. Radiographic atlas of skeletal development of hand wrist. 2nd ed. Stanford, CA: Standford University Press; 1971.
[32] Kirks DR. Practical pediatric imaging. Diagnostic radiology of infants and children. 1st ed. Boston/Toronto: Little, Brown & Company; 1984 [chapter 6; p. 198–201].
[33] Pietka E, Gertych A, Pospiech S, Huang HK, Cao F. Computer assisted bone age assessment: image preprocessing and ROI extraction. IEEE Trans Med Imaging 2001;20:715–29.
[34] Gertych A. Computerized bone age assessment based on clinical hand radiographs. PhD Thesis. Gliwice, Poland: Silesian University of Technology; 2002.
[35] Lim JS. Two-dimensional signal and image processing. Englewood Cliffs, NJ: Prentice-Hall; 1990.
[36] Lee JS. Refined filtering of image noise using local statistics. Comp Vis Graph Image Process 1981;15:380–9.
[37] Immerkaer I. Fast noise variance estimation. Comput Vis Image Understand 1996;64(2):300–2.
[38] Gertych A, Piętka E, Cao F, Huang HK. Computer assisted bone age assessment: region of interest segmentation. Proceeding SYMBIOSIS 2001, Szczyrk; 2001. p. 67–71.
[39] Li SZ. Markov random field modelling in computer vision. Tokyo: Springer; 1995.
[40] Pappas TN. An adaptive clustering algorithm for image segmentation. IEEE Trans Signal Process 1992;40(4):901–14.
[41] Gertych A, Piętka E. An automated segmentation and features extraction from hand radiographs. Arch Theor Appl Inform 2003; 15(3):315–26. ISSN 0867-2121.
[42] Piętka E, Pospiech-Kurkowska S, Gertych A, Cao: F. Integration of computer assisted bone age assessment with clinical PACS. Comp Med Imaging Graph 2003;27(2/3):217–28.
[43] Duda RO, Hart PO, Stork DG. Pattern classification. 2nd ed. New York: Wiley–Interscience; 2001.
[44] Mamdani EH. Application of fuzzy algorithms for control of simple dynamic plant. Proc IEEE 1974;121:1585–8.
[45] Pośpiech-Kurkowska S, Piętka E, Cao F, Huang HK. Fuzzy system for the estimation of bone age from wavelet features, Pro Biosignal: 2002. p. 441–3.
[46] Date CJ. An introduction to database systems. Reading, MA: Addison-Wesley; 2000.
[47] Huang HK. PACS—basic principles and applications. Wiley–Liss; 1999.

Ewa Pietka received the master degree in 1979, the PhD in 1987 from the Silesian University of Technology in Poland, and the DSc in 1997 form the Polish Academy of Sciences. In 1979 she started working for the Silesian University of Technology. In 1989 Dr Pietka joined the staff of the Medical Imaging Departure at the University of California, Los Angeles. In 1993 she received an appointment at the University of Geneva, Switzerland. In 1996 she joined the Silesian University, School of Medicine and in 1999—the Silesian University of Technology. Currently she is a Professor at the Silesian University of Technology and at the University of Southern California, Los Angeles. Her research activities center around image analysis, pattern recognition, picture archiving and communication system, and medical information systems.

Arkadiusz Gertych received master degree in 1995 and the PhD degree in electrical engineering in 2003 from the Silesian University of Technology, Gliwice, Poland. In 1996 he started working for Silesian University of Technology in Department of Biomedical Electronics. He is currently working at University of Southern California, Los Angeles. His current research activities include computerized medical diagnosis support, data analysis, signal and image processing.

Krzysztof Witko received the master degree in 2002 from the Silesian University of Technology in Poland. He received an appointment and is currently working for the Teaching Hospital, Medical University of Silesia. He is also a PhD student at the Silesian University of Technology. His research activities center around picture archiving and communication system, and medical information systems.

ELSEVIER

Computerized Medical Imaging and Graphics xx (xxxx) 1–10

Computerized Medical Imaging and Graphics

www.elsevier.com/locate/compmedimag

Web-based electronic patient records for collaborative medical applications

Jianguo Zhang[a,*], Jianyong Sun[a,1], Yuanyuan Yang[a,1], Xiaomeng Chen[a,1], Lili Meng[b,2], Ping Lian[c,3]

[a]*Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China*
[b]*Xinhua Hospital Affiliated to Shanghai Second Medical University, Shanghai, China*
[c]*PLA 85 Hospital, Shanghai, China*

Received 28 April 2004; revised 30 August 2004; accepted 3 September 2004

Abstract

We developed a web-based system to interactively display electronic patient records (EPR), such as DICOM images, graphics, and structure reports and therapy records, for intranet and internet collaborative medical applications. This system has three major components, a C/S (client/server) architecture for EPR data acquisition and authoring, and a Web B/S architecture for data delivering. The Web viewer of this system integrates multi-media display modules and remote control module together to provide interactive EPR display and manipulation functions for collaborative applications. We have successfully used this system two times to provide teleconsultation for severe acute respiratory syndrome (SARS) patients in Shanghai Infection Hospital and Xinhua Hospital. During the consultation, both the physicians in infection control area and the experts outside the control area could use this system interactively to manipulate and navigate the EPR objects of the SARS patients to facilitate a more precise diagnosis. This paper gives a new approach to create and manage image-based EPR from actual patient records, and also presents a novel method to use Web technology and DICOM standard to build an open architecture for collaborative medical applications. The system can be used for both intranet and internet medical applications such as tele-diagnosis, teleconsultation, and distant learning.

Keywords: DICOM-based images; Electronic patient record; Web technology; Collaborative applications

1. Introduction

During the past years, using PACS (picture archiving and communication system) diagnostic workstations and Web technologies as a means to access digital image data have been implemented with different architectures [1,2]. Also, more and more multi-media medical documents or records are used in hospitals and medical community, and these documents usually are stored in different formats or systems. Most medical applications need to access these medical documents through intranet or internet. Also, there are a lot of collaborative medical applications needing to share or interactively exchange medical information through the networks [3,4].

Collaborative medical applications are often happened in some medical practices, procedures and education encounters. For example, local and remote doctors collaboratively discuss cases in telemedicine procedures [5,6]. Medical students or residents study or learn cases from medical experts through network [7]. For infection diseases, physicians and experts in different departments or areas of an infection control hospital can study patient records collaboratively through networks without the concern of being infected. They can concentrate in case discussion with patient medical records combining with image manipulation, and show results to remote participants through the network.

* Corresponding author. Tel: +86 21 5551 0087.
E-mail addresses: jzhang@mail.sitp.ac.cn (J. Zhang), nobm@sina.com (J. Sun), whyyy78@163.com (Y. Yang), among@online.sh.cn (X. Chen), mlili@citiz.net (L. Meng), telmed@icareyou.com (P. Lian).
[1] Tel.: +86 21 65420850x22218.
[2] Tel.: +86 21 65796770.
[3] Tel.: +86 21 62138277.

0895-6111/$ - see front matter © 2004 Published by Elsevier Ltd.
doi:10.1016/j.compmedimag.2004.09.005

For more difficult or complicated cases, the medical record usually contains medical images with other related records, which come from hospital information system (HIS) and other clinical information systems (CIS). For these reasons, the medical imaging informatics research and applications trends are to develop and build electronic patient record (EPR) with images for medical applications [8], and this effort will benefit the hospitals and medical institutions on their collaborative healthcare activities.

In this paper, we first review our previously developed interactive teleradiology system architecture, and discussed its advantages and disadvantages in collaborative applications. Second, we present a new design of web-based interactive system architecture and its major components, which support EPR display and manipulation and operate in a central mode for collaborative applications. The new system can be integrated with PACS and other hospital clinical information systems, and provides a web interface to allow access to the EPR either through intranet or Internet for different applications. Third, we discuss the work and data flows of the web-based EPR system in collaborative applications. We also present the advantages of using the central operation mode for the collaboration comparing to the point-to-point mode. Finally, we give an example of the how the new web-based EPR system was used in collaborative application for SARS (Severe acute respiratory syndrome) diagnosis.

2. Interactive teleradiology

Several years ago, we developed a cost-effective real-time teleconsultation system in a clinical DICOM PACS environment for collaborative health care application [5]. This system consisted of two sites: a referring physician site and an expert center, for point-to-point teleradiology service mode [5]. Both sites could display all the DICOM images and could remotely synchronize image manipulation during consultation using remote dual cursors. This system allowed real-time collaborative consultation of serious or difficult cases with high resolution and large volume medical images in a limited bandwidth network environment. The system relied both on standardized technologies, such as DICOM and TCP/IP network protocols, and on innovative techniques such the advanced message routing concept for remote control functionality. Also the dual cursors of the system could synchronize their operation and image processing results to make both participants feel no distance barrier and no misunderstanding between them during image study. Fig. 1 shows the basic architecture of this real-time interactive teleconsultation system.

However, there are two drawbacks in this system. First, it only supported DICOM image consultation, and could not receive other medical records. Secondly, the consultation operation mode was point-to-point, it would make collaborative application more complicated if third party was interested in or invited into the collaborative applications. This was because the system had to re-send images to a third party. In order to overcome these drawbacks, we design and develop a new web-based system to provide electronic patient records (EPR) collaboration with several new functionalities including a central operation mode for intranet and internet healthcare providers.

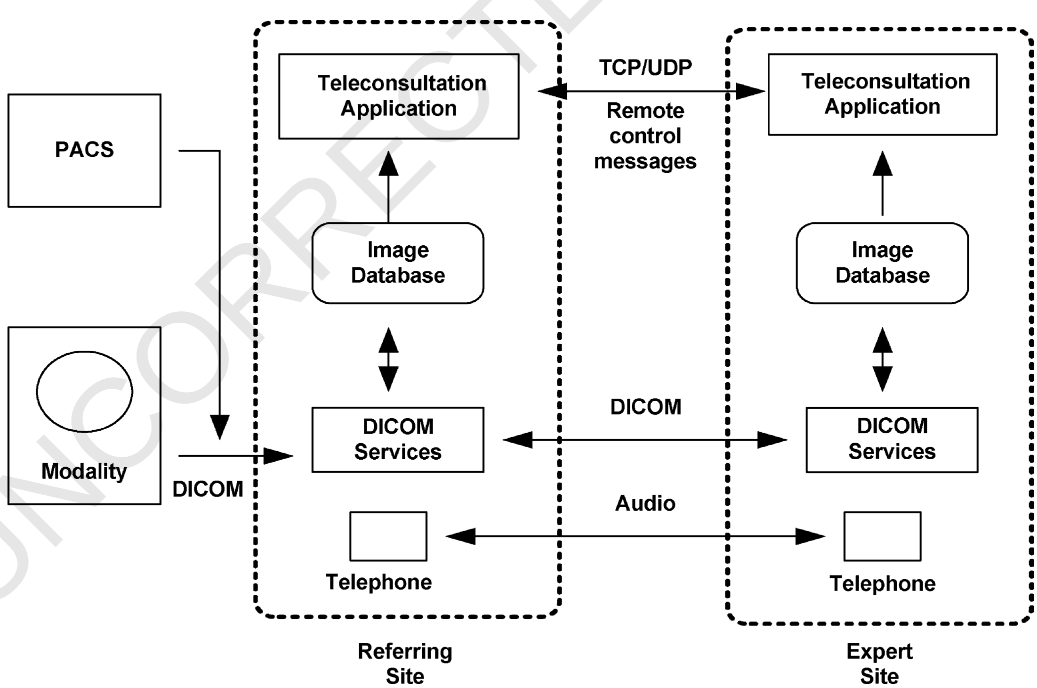


Fig. 1. The software architecture of the real-time teleconsultation system with point-to-point collaboration operation mode.

3. System architecture and major components

3.1. Software components and architecture

There are three major components in the designed EPR collaborative system: EPR gateway (EPR-GW), EPR server (EPR-Server) including the EPR repository server and web server, and EPR Web viewer (EPR-Viewer). Currently, there are two typical application architectures used in networked information systems: two tiers client/server (C/S) architecture and three tiers browser/servers (B/S) architecture. In order to avoid the drawback of the point-to-point, we designed the EPR collaborative system by using the combination of the C/S and B/S architectures, as shown in Fig. 2. The EPR data acquisition from the HIS/PACS/CIS to EPR gateway and the data transmission from the EPR gateway to EPR server adopt C/S architecture. The EPR data delivery from the EPR server to EPR Viewers uses the B/S architecture. The EPR data acquired from the HIS, PACS or CIS first have to pass to the EPR gateway to check their patient and study headers information to make sure the medical records of same patient linked together to provide patient-oriented applications. Then, the EPR data are sent to EPR server. Most medical records are in DICOM formats such as images, graphics, curves and RT records, but the text-based records are in XML (The Extensible Markup Language) format. The collaboration operation on the EPR data between any two web users is follows the central mode which means that all the consulted medical records would be sent to, managed and retrieved from the central EPR server. In Section 4, we will present the work and data flows of this web-based collaborative consultation system to explain how it works.

3.2. Information management model of EPR server

There are two parts in the EPR server: EPR repository server and web server. The EPR repository server uses Oracle object relational database (Enterprise Edition 9i) storing patient electronic medical records. It uses the general DICOM information object model as EPR information model to manage the medical records in the database. In order to manage other complex medical objects such as used in oncology treatment procedures, the information model also has the extension for radiation therapy (RT) objects defined by DICOM Standard. Fig. 3 shows the diagram of the data models used in our EPR database. The up-left part of the diagram relates to the general medical record management, e.g. DICOM images, reports, and other medical data, the down-right part covers the RT object management. We choose Microsoft Internet Information Server (IIS) as the web server, and the Internet Explore (5.0 or higher) as the default browser to be supported by our web-based EPR system. The web server provides *http* communication protocol interface to let users to access EPR by using Web browsers.

3.3. Component-based EPR web viewer

We use component software technology to develop a web-based image processing and display component to visualize and manipulate various DICOM images in a Web

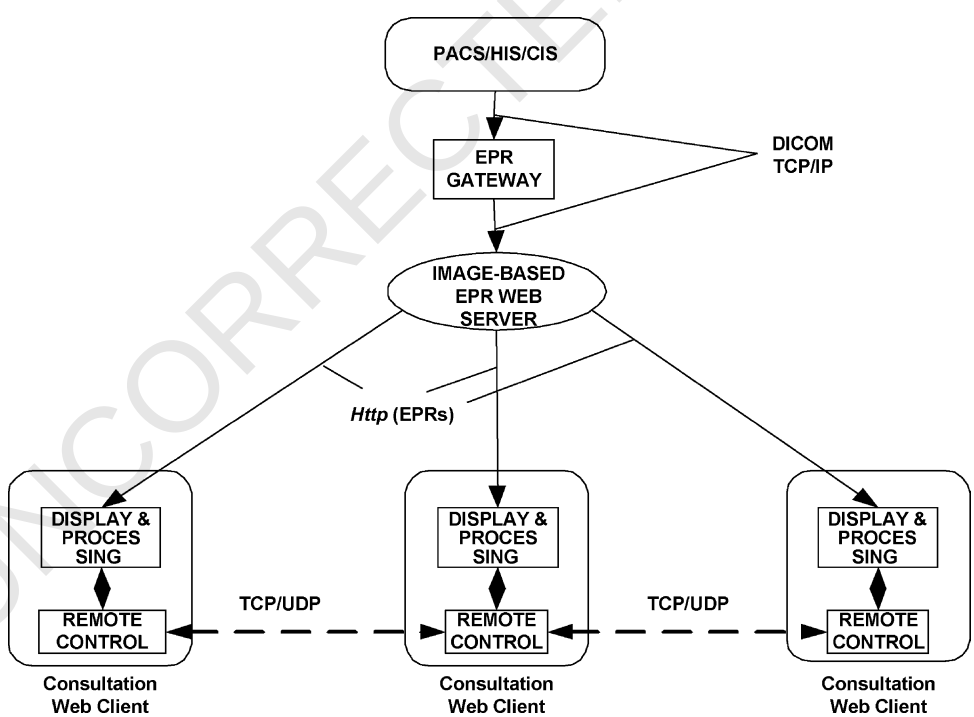


Fig. 2. The software architecture of the web-based EPR system with central collaborative operation mode.

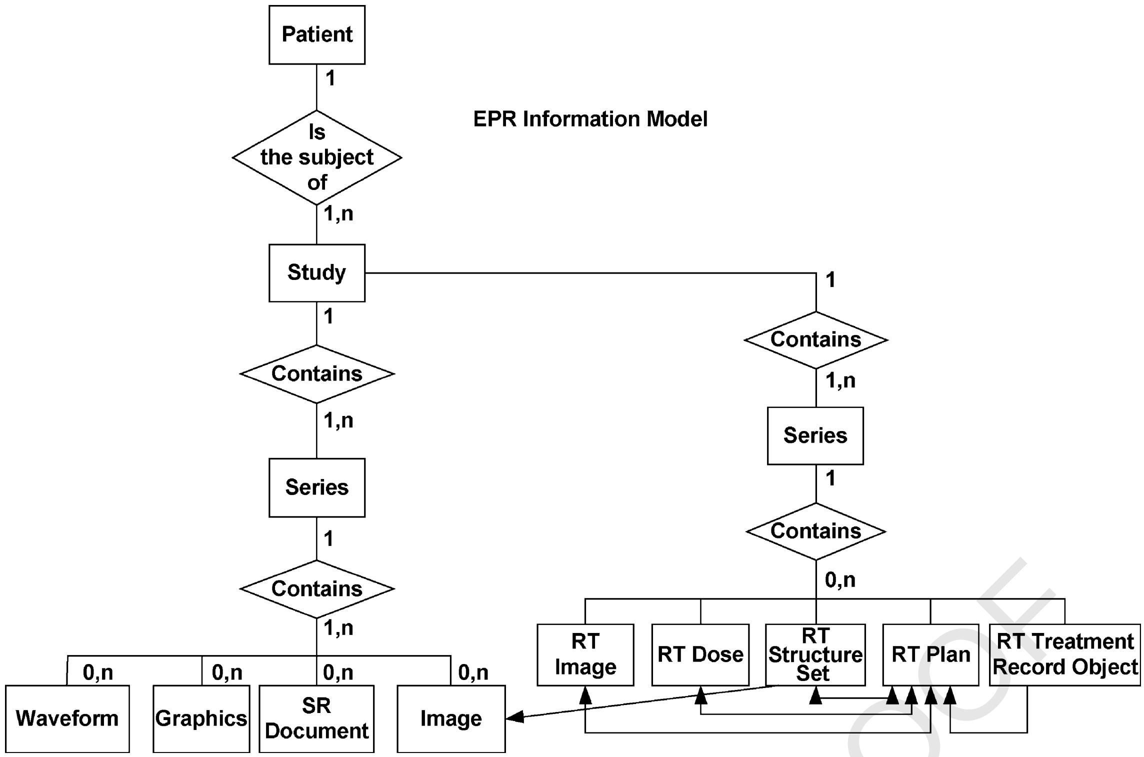


Fig. 3. The DICOM-based information model used for EPR management in the EPR repository server.

browser [9]. Since this component has an open and scalable architecture with multi-thread processing capability, we integrated multi-media display and manipulation modules and functions into this component to display the electronic patient records, as indicated in the Fig. 4. We also introduced the remote control module, developed in our previous interactive teleradiology system [5], into the display and processing component to enable this new component have collaborative operation functions with the EPR objects, such as using dual cursors to perform image manipulation and processing functions of window/level, zoom, overlay, orientation, and measurements. The remote control module sends and receives operation messages generated from local and remote sites, and dispatches the messages to multi-media display modules to synchronize their operations on the EPR objects. We integrate this new component into a Web browser to create web-based EPR Viewer, as we did it in our web-based PACS image distribution [9]. This allows users to use dual cursors, provided by the EPR viewer, to synchronize their operations

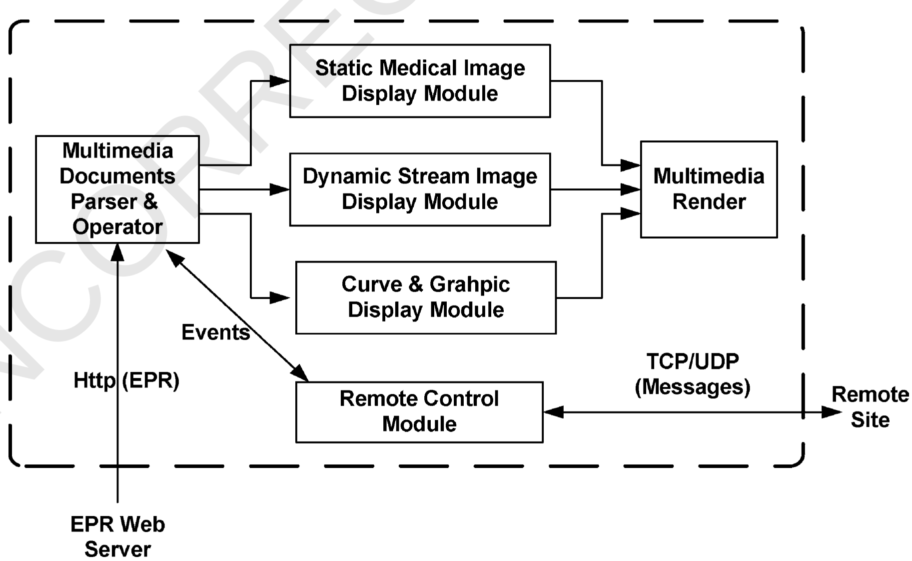


Fig. 4. The component architecture of the multi-media display modules integrated with the remote control module in the Web viewer.

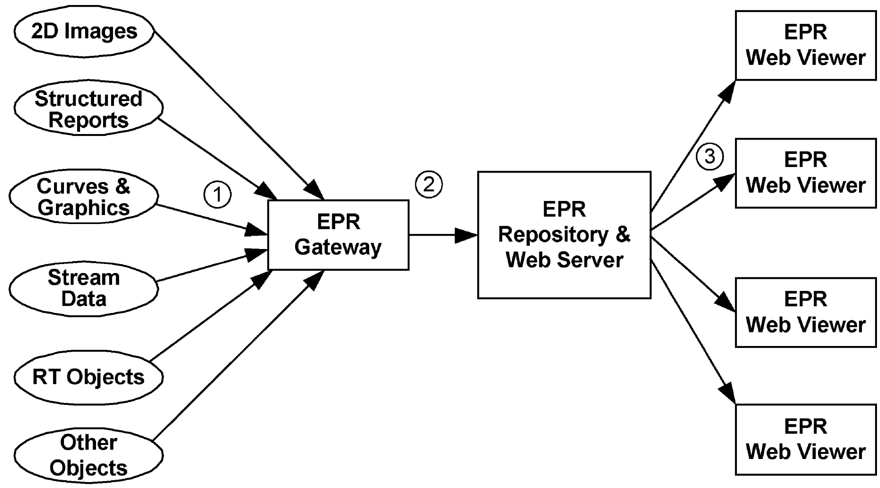


Fig. 5. The operation procedures from EPR data acquisition to final display.

to understand each other on studied cases. The communication relation between the web viewer and the web server is same with our web PACS [9].

4. Data flow and central collaborative operation mode

The data and work flows of the web-based EPR collaborative system presented in this paper are shown in Fig. 5. There are three steps to view and play EPR objects for the collaborative applications. (1) The EPR data, retrieved from hospital PACS, HIS, and other clinical information systems, are sent to the EPR gateway for data formatting if they are not already in DICOM format or XML. (2) The image or DICOM based medical data are sent to the EPR Repository Server for archiving from EPR gateway. (3). Any authorized Web user can access the EPR web server, by using the EPR Web viewer, to get the EPR data and to display them. Fig. 6 shows a GUI (graphic user interface) of the EPR Web viewer, which displays the ECG curve, DSA dynamic image, CT images, RT structure set of same patient.

Comparing to the point-to-point operation mode mentioned in Section 2 of this paper, the collaborative operation

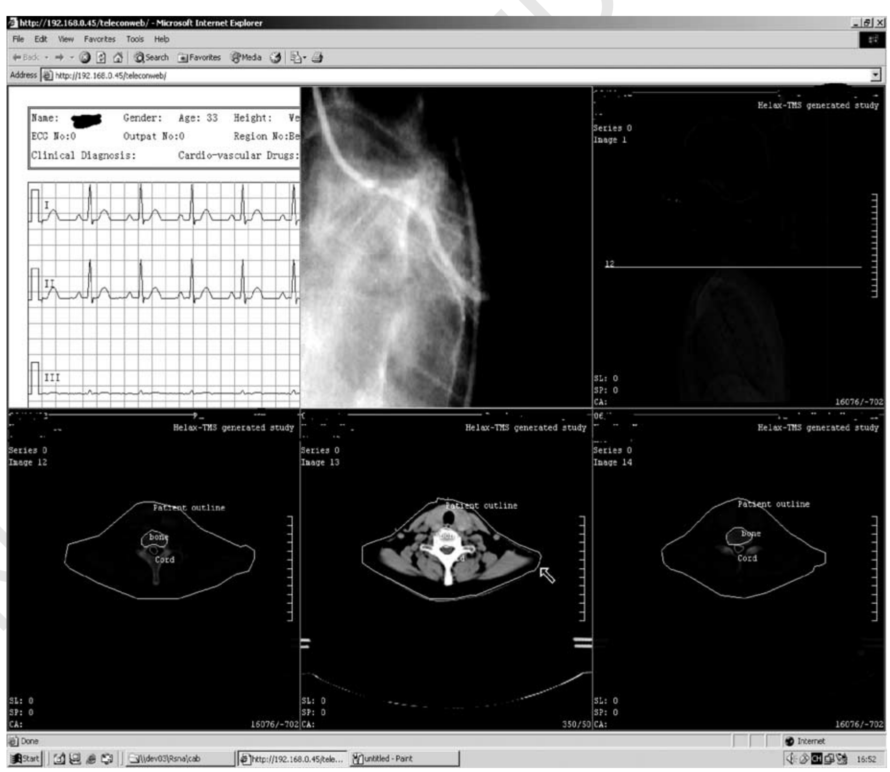


Fig. 6. The EPR display GUI of the Web viewer with ECG, dynamic angiographic images, RT images and objects.

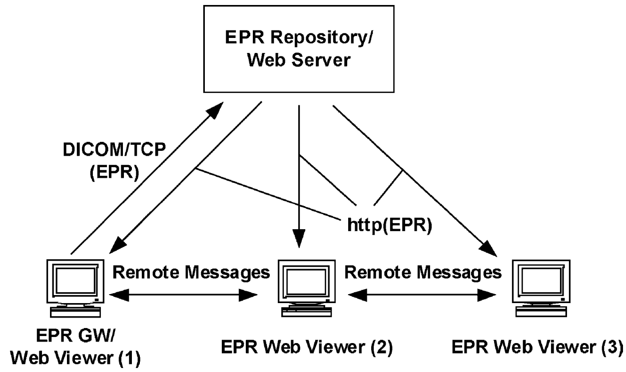


Fig. 7. The work and data flows of the web-based EPR system for collaborative applications.

mode in this web-based EPR system can be considered as a central mode, since any user or a client gets the collaborative EPR data from the central EPR Server. Fig. 7 gives an example of how the system operates with three collaborative participants. First, the medical image, records or objects are sent to the EPR Repository Server from the client 1 through DICOM or TCP/IP protocol by using EPR gateway. Later on, client 2 or 3 can get the EPR data through *http* protocol from the web server and manipulate the data interactively with client 1, or between them by using the message routing and controlling function through the remote control module integrated in multi-media display and processing component. So, the collaborative operation procedures of the central mode are much simpler and easier comparing to the point-to-point mode. There is only one data authoring and sending operation before the data are ready to all potential participants.

5. Preliminary application of the web-based EPR system for SARS consultation

5.1. Application background and network connection

Severe acute respiratory syndrome (SARS) is a respiratory illness that has been reported in Asia, North America, and Europe. During the winter of 2002 through the spring of 2003, World Health Organization received reports of > 8000 SARS cases and nearly 800 deaths. The main way that SARS seemed to spread is by close person-to-person contact. The virus that caused SARS was thought to be transmitted most readily by respiratory droplets (droplet spread) produced when an infected person coughed or sneezed. Most of the SARS cases in China occurred by infection in hospitals or among travelers. To protect the physicians, experts and nurses from the SARS during the diagnosis and treatment procedures, the infection control mechanisms were built in SARS hospitals in China, e.g. medical workers in SARS control area could not contact outside people and even medical records could not bring out the control area. Usually, there were three areas in the SARS hospital: the infection area, in which SRAS patients were stayed and taken cared by clinical SARS physicians; non-infection area, in which the radiologists and other experts of the hospital worked to support the SARS diagnosis and treatment happened in the infection area; and the data center where the HIS, PACS, CIS located. Also, there was an off-site SARS expert center out side the SARS hospital, where the other hospital experts can provide teleconsultation services to help the SARS hospital on SARS diagnosis and treatments. In order to make radiologists, physicians and experts both inside and out side control area collaborate efficiently on SARS image diagnosis without concerning the infection, we installed the web-based EPR system in a SARS hospital and an off-site expert center in Shanghai to provide tele-consultation for SARS diagnosis.

The web-based EPR collaborative system was implemented in Shanghai Infection Hospital and Shanghai Xinhua Hospital affiliated to the Second Shanghai Medical University in May 2003. The infection area and non-infection area were in Shanghai Infection Hospital, and the off-site expert center was in Xinhua Hospital, the distance of the two hospitals was about 10 km. There were three Web EPR viewers or clients installed in these two hospitals: two were located in the infection area and non-infection area of the Infection Hospital, and the third one was located in Xinhua Hospital, as shown in Fig. 8. With the Web architecture and the intranet and Internet connections of the system to different clients located in side and out side hospitals, the real-time teleconsultation with interactive EPR manipulation could take place between any two sites of the infection area, non-infection areas, and off-site expert center.

5.2. Data acquisition and transmission

Since any medical record of SARS patients, such as paper record and films, was not allowed to bring out of infection area according to the infection control mechanism, the digital image acquisition and transmission were absolutely demanded in the diagnosis and consultation procedures. Although most images of SARS patients could be obtained directly from CT and CR modalities in situ, there were still a large quantity of film images, which were transferred from other hospitals accompanying with patients' arrival. We used a laser digitizer (Array, 2905, Japan) to digitize these film images and sent the digitized film images together with direct digital images to the web-based PACS for archiving. Fig. 9 shows the digitized film images of a SARS patient in the infection area of the Shanghai Infection Hospital. Most film sizes were 14×17 in., and the image formats of digitized films were DICOM about $3560\times 4320\times 2$ b (the grayscale was 12 bits). The size of one digitized film images was usually 29.3 Mb.

Images transmitted inside SARS hospital were trivial with fast LAN (local area network) network connection for

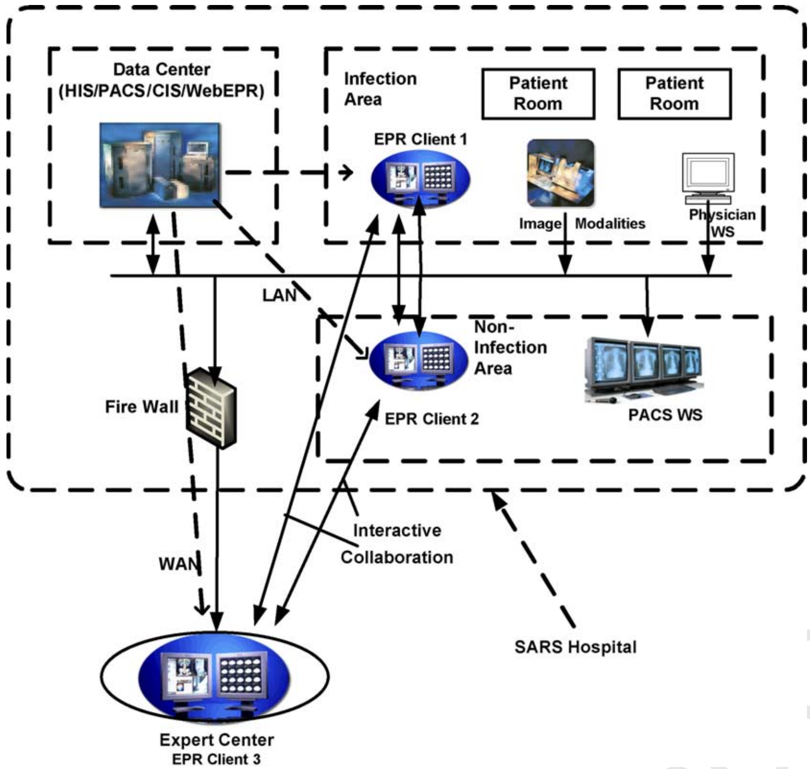


Fig. 8. The network connections of the web-based EPR system used in SARS consultation in infection control mechanism.

daily diagnosis and consultation between infection area physicians and non-infection area doctors. For the off-site consultation, all patient images had to be transmitted to the expert center through WAN (wide are network) before the consultation started. We used two WANs for image transmission, one was ATM network built by Shanghai local healthcare organization, and the other was ADSL owned by Shanghai Telecommunication Corp. The bandwidth of the ATM site to site was about 10 Mb/s. For the ADSL, the downloading bandwidth was 2 Mb, and uploading was 1 Mb/s. There were two consultation meetings held for SARS diagnosis between the Shanghai Infection Hospital and Xinhua Hospital during May 2003. The first consultation meeting was held using ADSL WAN, the image transmission time for one case took about one hour and more. The second consultation meeting was held using ATM WAN. The time required to transfer one case was less than 15 min. Tables 1 and 2 give the transmission results of the digitized film images from Shanghai Infection Hospital to Shanghai Xinhua Hospital by using ADSL and ATM, respectively. The other medical records of SARS patients acquired from the HIS were sent to the EPR repository before the image transmission. There were also ICU monitoring (intensive care unit) data used in the consultation, which were real-time transmitted from the infection area to the non-infection area and the off-site expert center through the LAN and WAN.

Fig. 9. The image acquisition used for SARS consultation with the web-based EPR system.

5.3. Interactive teleconsultation for SARS diagnosis

Since the diagnosis and treatment procedures for SARS were complicated, the physicians, radiologists and experts usually had to review all the medical records of a patient to make final decisions. The web-based EPR system not only provided EPR navigation tool to the users, but also enable them interactively to study the medical records. For example, both local and remote users can interactively manipulate the images and medical records with dual cursors to understand each other on the cases. Fig. 10 shows

Table 1
Transmission results of the digitized film images from Shanghai Infection Hospital to Shanghai Xinhua Hospital by using ADSL network connection

Case	Image type	Number of images	Total size (Mb)	Transmission time (mm:ss)	Used bandwidth (kbps)	Usage of band-width for 1 Mbps(%)
1	Second capture (SC)	11	322.3	85:21	515.58	50.35
2	SC	8	234.4	62:32	511.78	49.98
3	SC	10	293.5	78:04	513.31	50.13

Table 2
Transmission results of the digitized film images from Shanghai Infection Hospital to Shanghai Xinhua Hospital by using ATM network connection

Case	Image type	Number of images	Total size (Mb)	Transmission time (mm:ss)	Used bandwidth (kbps)	Usage of band-width for 10 Mbps (%)
4	SC	20	561	14:48	5080	50.8
5	SC	20	558	14: 35	5100	51.0
6	SC	15	437	11:23	5120	51.2
7	SC	11	320	15:03	2830	28.3

the dual cursors pointing to a SARS chest image, which were controlled by local and remote users.

There were 8 SARS patients consulted from the remote expert center, and total 2.73 Gb image data were transmitted from the Infection Hospital to Xinhua Hospital in the consultation activities. During the consultation meeting, the web-based EPR system provided bi-directional remote control functionality to both sides on the EPR data processing and manipulation, and synchronized their operations on the patient medical records so that it is easy for them to do the case analysis and management. Fig. 11 shows an expert in the remote expert center, which was located in Xinhua Hospital, talking to the SARS physician about SARS patient images with the system.

6. Conclusions

Collaborative medical applications need to share or interactively to exchange medical information through the networks and most medical documents or records stored in different formats or systems. This paper presented a novel approach to design and develop a web-based electronic patient record system for collaborative medical applications. The electronic patient records supported by this

Fig. 10. The GUI of the Web viewer with dual cursors on the chest images.

Fig. 11. An expert in the off-site expert center interactively talking to the clinical SARS physician located in the infection area of the SARS hospital about a SARS patient images by using the web-based EPR system.

system cover most DICOM image-based objects and medical records. This system has three major components, use a C/S architecture for EPR data acquisition and authoring, and a Web B/S architecture for data delivery. The web viewer of this system integrated multi-media display modules and remote control module together to provide interactive EPR display and manipulation functions for collaborative applications. The web-based architecture of the system makes the collaboration operation procedures much simpler and easier to the users.

We have installed this system in Shanghai Infection Hospital and Shanghai Xinhua Hospital to provide teleconsultation services for SARS diagnosis between the SARS physicians and experts in May 2003. There were two consultation meetings held for SARS diagnosis between the Shanghai Infection Hospital and Xinhua Hospital during May 2003. The first consultation meeting was held using ADSL WAN, and the second consultation meeting was held using ATM WAN. There were 8 SARS patients consulted from the remote expert center, and total 2.73 Gb image data were transmitted from the Infection Hospital to Xinhua Hospital in the consultation activities. During the consultation meeting, the web-based EPR system provided bi-directional remote control functionality to both sides on the EPR data processing and manipulation, and synchronized their operations on the patient medical records so that it is easy for them to do the case analysis and management. The consultation results showed that this web-based EPR system could provide multi-sites interactive consultation services between SARS clinical physicians, so that radiologists and off-site experts did not have to be concerned with SARS infection. Also, the central collaborative operation mode simplified the consultation procedures in image and medical record delivery and display comparing to the point-to-point mode.

This paper also gives a new approach to create and manage image-based EPR from actual patient records, and also presents a novel method to use web technology and DICOM standard to build an open architecture for collaborative medical applications. The New web-based EPR system provides new method to enhance the value of interactive and collaborative studying of EPR for difficult cases, especially for SARS diagnosis. The on-time interactive communication features of the system can manipulate EPR objects which, in turn, can improve the efficiency and the quality of collaborative healthcare. The system can be used for both intranet and Internet medical applications such as tele-diagnosis, teleconsultation, and distant learning.

Acknowledgements

This research was supported in part by the National Nature Science Foundation of China (Grant No. 39925009), China '863' High Tech. Plan (Contract No. 2003AA0010), and Shanghai Sci. and Tech. Plan (Contract No. 03DZ19709), and Chinese Academy of Sciences. The authors would like to thank the colleagues of Shanghai Infection Hospital and Xinhua Hospital for data acquisition, transmission, and system testing.

References

[1] Kim Y, Horii SC. Hand Book of Medical Imaging, vol. 3. Bellingham, Washington: SPIE Press; 2000.

[2] Huang HK. PACS and imaging informatics. Display workstation. New York: Wiley-Liss Press; 2004 [Chapter 11].

[3] Sakusabe T, Kimura M, Onogi Y. On-demand server-side image processing for web-based DICOM image display. SPIE Proc 2000; 3976:359–67.

[4] Cao X, Hoo KS, Zhang H, et al. Web-based multimedia information retrieval for clinical application research. SPIE Proc 2001;4323:350–8.

[5] Zhang J, Stahl JN, Huang HK, et al. Real-time teleconsultation with high resolution and large volume medical images for collaborative health care. IEEE Trans Inf Technol Biomed 2000;4:178–86.

[6] Stahl JN, Zhang J, Huang HK, et al. Teleconferencing with dynamic images. IEEE Trans Inf Technol Biomed 2000;4:88–95.

[7] Huang C. Changing learning with new interactive and media-rich instruction environments: virtual labs case study report. Comput Med Imaging Graphics 2003;27:157–64.

[8] Ratib O, Swiernik M, McCoy JM. From PACS to integrated EMR. Comput Med Imaging Graphics 2003;27:207–16.

[9] Zhang J, Sun J, Stahl JN. PACS and web-based image distribution and display. Comput Med Imaging Graphics 2003;27:197–206.

Jianguo Zhang received his BS and MS degrees in 1984 and 1987 from Optical department, Shangdong University, and PhD in 1991 from Chinese Academy of Science. From 1994 to 1998, He was a research fellow in the Laboratory for Radiological Informatics, Department of Radiology, University of California at San Francisco. Currently, He is Director and Professor of Laboratory for Medical Imaging Informatics in Shanghai Institute of Technical Physics, Chinese Academy of Science. His major interests include PACS technologies, medical image processing and visualization, imaging informatics, and medical information system integration and applications.

Jianyong Sun received his BS from Electronic Engineering department of Shandong University, and MS and PhD degrees in 2001 and 2004 from Shanghai Institute of Technical Physics, Chinese Academy of Sciences. Currently, he is assistant professor in the Laboratory for Medical Imaging Informatics of Shanghai Institute of Technical Physics. His interests include image display and visualization, image processing and teleradiology.

Yuanyuan Yong received her BS and MS from Computer Science department and Biomedical Engineering department in 1999 and 2002 from South Central University for Nationalities. She joined the Laboratory for Medical Imaging Informatics of Shanghai Institute of Technical Physics in 2002. Currently, she is studying as PhD student in the graduate school of Chinese Academy of Science. Her interests include telemedicine and multimedia database.

Lili Meng BS, the chief of Information Technology Department of Xinhua hospital and Children Healthcare Center in Shanghai. Her interests include hospital information management and telemedicine for infection disease.

Ping Lian MD, co-director of the 85th hospital of People's Liberation Army (PLA) in Shanghai and the director of the Development Center for Telemedicine Technologies of PLA. His interests include clinical applications of telemedicine technologies, tele-education for medicine and hospital information management.

ELSEVIER

Computerized Medical Imaging and Graphics xx (xxxx) 1–7

Computerized Medical Imaging and Graphics

www.elsevier.com/locate/compmedimag

HIPAA compliant auditing system for medical images

Zheng Zhou*, Brent J. Liu

Image Processing and Informatics (IPI) Laboratory, Radiology Department, University of Southern California, Marina del Rey, CA 90292, USA

Received 16 March 2004; revised 17 August 2004; accepted 30 September 2004

Abstract

As an official regulation for healthcare privacy and security, Health Insurance Portability and Accountability Act (HIPAA) mandates health institutions to protect health information against unauthorized use or disclosure. One such method proposed by HIPAA Security Standards is audit trail, which records and examines health information access activities. HIPAA mandates healthcare providers to have the ability to generate audit trails on data access activities for any specific patient. Although current medical imaging systems generate activity logs, there is a lack of formal methodology to interpret these large volumes of log data and generate HIPAA compliant auditing trails.

This paper outlines the design of a HIPAA compliant auditing system (HCAS) for medical images in imaging systems such as PACS and discusses the development of a security monitoring (SM) toolkit based on some of the partial components in HCAS.

Keywords: HIPAA; Security; HIPAA compliant auditing system; Auditing; monitoring

1. Introduction

Health Insurance Portability and Accountability Act (HIPAA) [1,2] of 1996, Public Law 104–191, was officially instituted on April 14, 2003. HIPAA mandates healthcare providers to be HIPAA compliant by April, 2005. The goal of HIPAA is to set and enforce the standards to protect the privacy and security of health data. Currently, there are four types of standards in HIPAA: Transaction and Code Set Standards, Identifier Standards, Privacy Standards and Security Standards. Among them, HIPAA Security Standards [3] are to protect the confidentiality, integrity and availability of electronic health information, and to protect the information against unauthorized use or disclosure by utilizing administrative, physical and technical safeguards. The technical safeguard defines the technical methods to assure the security of the health data. One such method proposed by HIPAA is audit trail, which can record and examine information system activities. HIPAA requires health institutions to have the ability to generate audit trails on data access activities for a specific patient on demand. Specifically, HIPAA compliant audit trails require following information for the health data access [4]:

- Identification of the person who access the data
- Identification of the data that is accessed
- Location of where the data is accessed
- Date and time when the data is accessed
- Types of access (create, read, write, modify, delete)
- Status of access (success or failure)

Most current health information systems have no such ability to generate HIPAA compliant audit trails, even though they generate activity logs. Pertinent auditing information should be extracted from these logs to create audit trails. However, there is a lack of a formal methodology to interpret the potential large volumes of these log data to generate HIPAA compliant audit trails.

Some efforts have been achieved by developing HIPAA compliant auditing tools for general health information systems [5–7]. These auditing tools generate audit trails by recording the health data transactions or changes in logs and extracting the pertinent auditing information from these logs on demand. This method is applicable for health information systems that have all the data transactions or data flow controlled by a centralized server, such as Radiology

* Corresponding author. Tel.: +1 310 448 9437; fax: +1 310 448 9441.
E-mail addresses: zhengzho@usc.edu (Z. Zhou), brentliu@usc.edu (B.J. Liu).

0895-6111/$ - see front matter © 2004 Published by Elsevier Ltd.
doi:10.1016/j.compmedimag.2004.09.009

information system (RIS) [8]. However, the data flow is much different in integrated medical imaging systems, such as Picture Archiving and Communication System (PACS). There is no single component that can control and record the data flow of all components in PACS. For example, PACS archive server, as the central server of PACS, has no control of the workflow of the CT modality, and vice versa. This makes it very difficult for these auditing tools to record all the data transactions and changes in PACS. Therefore, a HIPAA compliant auditing tool for integrated medical imaging systems needs to be tailored to the complex workflow.

In this paper, we focus on designing a HIPAA compliant auditing system (HCAS) for medical images in PACS. HCAS can generate HIPAA compliant audit trails of image data access for a specific patient on demand. It can also automatically monitor the data flow of PACS facilitating the detection of unauthorized image access and other abnormal activities. In addition, this paper will discuss the development of a security monitoring (SM) toolkit based on partial components in HCAS.

2. Deign of HIPAA compliant auditing system (HCAS)

Our goal is to design a HIPAA complaint auditing system (HCAS) for medical images in PACS with the log data from PACS components as input sources.

2.1. Design criteria

In order to apply the HCAS in various PAC systems generating different format log files, the HCAS must be independent from any individual PACS. For this reason, we defined the design criteria as follows:

a. HIPAA compliant. The HCAS should always be able to generate the HIPAA compliant auditing trail report in terms of who access it, when, where, what are accessed, access status, and access types.
b. Open and extensible. The HCAS should provide interfaces for integration of new auditing or monitoring techniques. The HCAS should be able to support current HIPAA auditing requirements and accommodate new HIPAA additions in the future without affecting the already existed components.
c. Portable. This HCAS always assumes there is existing logs generated by PACS and makes use of these logs. The HCAS itself will not develop any logs for PACS. Hence, this makes the HCAS not tie to any individual PACS.
d. No interruption on the workflow of PACS. Any interruption on the workflow of PACS should be avoided when designing the HCAS.

2.2. Design development of HCAS

According to the aforementioned design criteria, HCAS was designed as a four-layer system shown in Fig. 1. The first layer (the lowest layer) is the Record layer, consisting of all the logs in PACS components. By logically separating PACS logs from other components in HCAS, independence from PACS can be achieved. The second layer is the Audit layer, which includes a centralized auditing database and other audit data analysis and interpretation tools. HIPAA compliant audit trails can be generated based on the auditing database. This layer also enables us to automatically monitor the data flow of PACS, which greatly assists PACS management. The third layer is Notification layer, which has a Notification component sending warning or alert messages of abnormal events to end users, such as PACS administrators. Finally, in the fourth layer, end users can decide to take certain actions against these abnormal events.

2.2.1. Record layer

This first layer is the data resource layer, including various types of log data shown in Table 1. PACS application logs are the event logs generated by individual PACS application. For example, an image query/retrieve event in PACS archive server may include such information as time, local host name, DICOM Application Entity Title (AET), patient information and query/retrieve status. PACS user login logs record user login events in each individual PACS component. Other computer system logs generated in PACS components, such as application access logs, can also provide supplement information.

New logs can also be added to this layer. For example, an image integrity log can be added to record image data integrity verification events. Data integrity, as one requirement of HIPAA Security Standards, refers to protecting image data from being altered or destroyed by unauthorized users. A digital signature embedding (DSE) method has been developed to ensure the data integrity of medical images at IPI laboratory [9,10]. By recording signature verification time, local machine, and signature verification status in the integrity log, DSE method can provide logs for the HCAS to generate HIPAA compliant audit trails on the data integrity of image.

These logs provide the pertinent information needed to generate HIPAA compliant audit trails. However, to extract and interpret the pertinent information from thousands of log events requires proper methodology, which will be addressed in the second layer, the Audit layer.

2.2.2. Audit layer

As shown in Fig. 1, the Audit Layer is the heart of the HCAS. It collects the audit data from distributed PACS components and stores the data in a centralized auditing database. The database is then used for audit analysis and automatic monitoring. Currently, there are

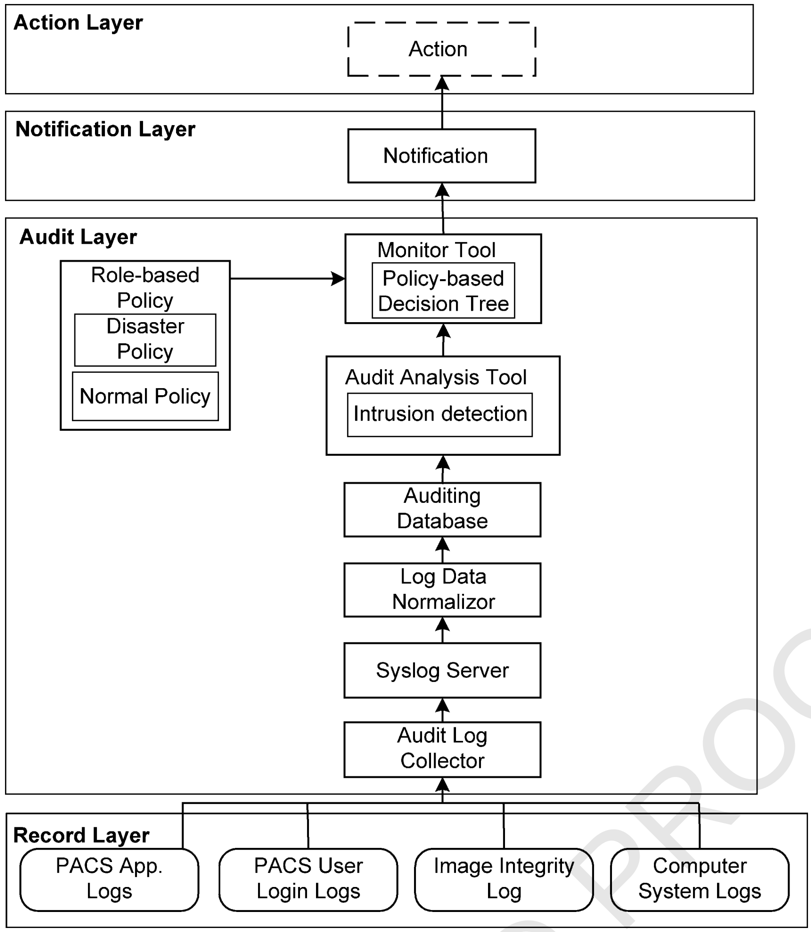


Fig. 1. HIPAA compliant auditing system (HCAS) for auditing of medical images in PACS.

seven components in this layer: Audit Log Collector, Syslog Server, Log Data Normalizer, Auditing Database, Audit Analysis Tool, Role-based Policy and Monitor Tool.

Audit Log Collector. Because the audit data is scattered within large volume of logs, a collector was designed to extract the pertinent data from logs and send the data to the centralized Auditing Database. PACS logs may be stored with different formats, such as database tables or textual files. The collector must support all these types of logs.

Syslog (System log) server. The pertinent data extracted from PACS logs are distributed in different PACS components connected by digital networks. In order to store them in the centralized Auditing database, a transmission mechanism is needed.

Currently, syslog [11] is a de-facto standard for transport and storage of event notification messages in UNIX systems, network devices and network applications. Syslog is a client-server mechanism. The clients can be configured to locally store event messages or directly send event messages to the server without local storage. Syslog uses User Datagram Protocol (UDP) to transfer event messages. This feature can be utilized to reduce the overhead added to the image transmission in PACS caused by event message communication, since PACS uses DICOM (Digital Imaging

Table 1
Types of logs existing in the HCAS record layer

Type of logs	Time	Location	User name	Status	DICOM AET[a]
PACS application logs	Operation time	Local operation machine	N/A	Status of this operation	DICOM AET of this application
PACS user login logs	User login time	Local machine user login	Login user name	Status of login	N/A
Image integrity logs	Integrity check time	Local machine does check	N/A	Status of integrity check	N/A

[a] AET, application entity title.

and Communications in Medicine) Protocol and Transmission Control Protocol (TCP).

For this advantage, syslog technology was adopted in the HCAS to transfer pertinent log data. The data is converted to syslog format by the syslog client in each PACS component. The client then sends the data to Syslog server, which will forward the data to the Log Data Normalizer.

Log Data Normalizer. The pertinent data extracted from PACS components might have different terminologies for the same object. For example, the name ‘film clerk’ in the CT modality might be named as ‘clerk’ in an MR modality. For this reason, a Log Data Normalizer was designed to normalize the data into common terms and then add them to the Auditing Database.

Auditing Database. In order to generate HIPAA compliant audit trails in a short time, a centralized database was designed to preserve all the obtained auditing data. The structure of database was designed based on the requirement of HIPAA compliant audit trails, including: who, when, where, what, how, and status. Patient information, such as name and id, and other relevant information are also included in the database. The advantages to use database technology to preserve the log data are:

a. *No loss of historical logs*: since all the logs generated in PACS components are obtained and stored in the database everyday, there is no loss of log data when these logs are overwritten and updated by PACS components.
b. *Centralized management of data access information for every patient*: the image data access events for an individual patient usually happen in multiple PACS components. For example, an event that a CT image is generated in a CT modality and another event that the same CT image is retrieved to viewing workstation for clinical review are related to the same patient. But these two events were recorded in two different logs at two separate PACS components. Without centralized database, the pertinent information needs to be extracted from these two components every time HIPAA compliant audit trails of image access for this patient is desired. Therefore, a centralized database enables us to quickly generate audit trails.

Audit Analysis Tool. Most current PACS lack a mechanism to dynamically monitor the data flow, which results in PACS management mostly relying on the experience of PACS administrators. A monitoring tool that can automatically analyze the data to find abnormal patterns and make decisions on the patterns would make PACS management much easier. To develop such a tool, the information of data flow of PACS needs to be collected and analyzed in real time.

With audit data collected in the auditing database, the HCAS can provide this ability using some data analysis techniques, such as Intrusion detection technology [12]. Audit Analysis Tool is the component in the HCAS to perform such data analysis functions.

Monitor Tool. After the Audit Analysis Tool finds abnormal patterns in the data flow of PACS, a Monitor Tool was designed to monitor the pattern, and make decisions whether it is an unauthorized data access for the abnormal pattern based on the Role-based Policy. Any pattern that violates the Policy would automatically cause a warning or alert result. For example, Audit analysis tool discovers an abnormal pattern of image query/retrieve by a PACS user ‘A‘, belonging to the role of ‘Film clerk‘, which was defined to have no image query/retrieve right in the Policy. Monitor tool automatically makes a decision that this is an unauthorized image query/retrieve and gives a warning message.

Role-based Policy. The Role-based Policy defines the roles for PACS users based on the roles they performed in the clinical environment, such as film clerk, PACS manager and Radiologists, and the image access rights for each role. Two types of policies, Normal policy and Disaster policy, are defined for two different conditions. Normal policy is for daily operation, whereas Disaster policy is defined for the emergency situations, such as earthquake, when normal policy can be bypassed.

2.2.3. Notification layer

Notification layer consists of a notification component, which receives the warning or alert messages from the Audit layer and notifies PACS end users of the unauthorized image data access and other abnormal activities.

2.2.4. Action layer

Action Layer is designed for PACS end users to take actions, such as access control, against the unauthorized image access and other abnormal activities.

This four-layer HCAS design enables PACS to generate HIPAA compliant audit trails of image data access for a specific patient on demand. Meanwhile, it can automatically monitor the data flow of PACS facilitating PACS management. With an open and extensible design, the HCAS can also easily incorporate new data analysis and monitoring techniques, and be extended to support future HIPAA requirements.

3. Development of a security monitoring (SM) toolkit

A security monitoring (SM) toolkit has been developed for automatic monitoring the data flow of PACS based on partial components of the Audit Layer in HCAS. As shown in Fig. 2, the SM toolkit currently includes such components as Audit Log Collector, Syslog Server, Auditing Database, Monitor Tool and Role-based Policy (Normal Policy).

A simple role-based policy was designed according to the roles of healthcare providers. The policy consists of three tables: role table, resource table and policy table.

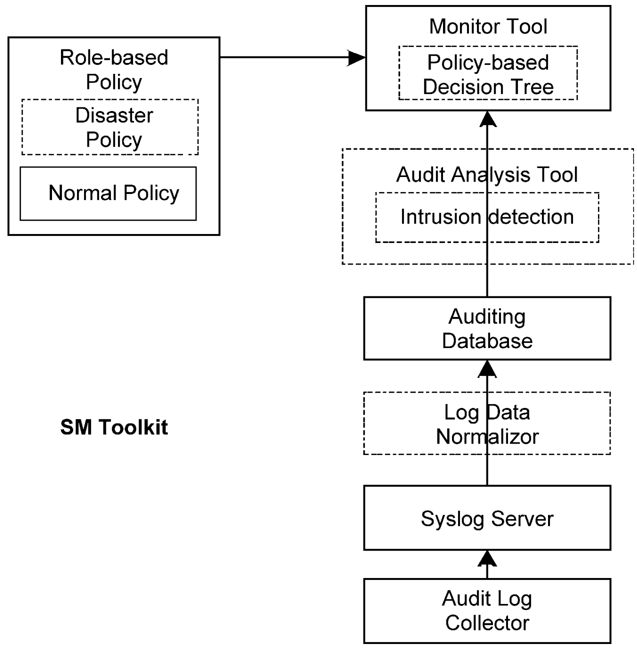


Fig. 2. Components and data flow of the security monitoring (SM) toolkit based on partial components of the Audit Layer in HCAS. The components with dotted line have not been implemented in the SM toolkit.

The role table defines different roles, such as radiologists. The resource table records PACS applications, such as the viewing software. The policy table, shown in Table 2, assigns each user the role and the resources this user can access based on the right of his/her role. For example, one entry in the policy table is user ‘Michael’, ‘System Administrator’, and ‘All’. ‘All’ means all the resources can be accessed by this user.

The toolkit can monitor the dynamic data flow of PACS. First, it collects pertinent auditing data from PACS application logs, PACS user login logs and other computer system logs. It then stores the log data in the Auditing Database. Next, the toolkit compares the user name in every record in the Auditing Database and the user name in the policy table. If match occurs, the toolkit further compares the application name in the database record and the application name in the policy table. If any comparison failed, the SM toolkit gives out a warning message of unauthorized image data access in its graphic user interface (GUI). Otherwise, a normal message is given out.

Table 2
Example of a role-based policy table

User name	Role name	Resources
Michael	System administrator	All
John	Clerk	Viewing software
Joe	Radiologist	All
Jessica	Technician	CT scanner

Currently, the toolkit lacks the ability to generate HIPAA compliant audit trails of image data access for a specific patient. This function is currently in development.

4. Evaluation of SM toolkit

In order to evaluate the impact of this toolkit in PACS, a laboratory-based PACS Simulator has been developed to simulate the data flow of clinical PACS.

4.1. Educational PACS simulator

An educational PACS (Picture Archiving and Communication System) Simulator [13,14] has been developed for the purposes of PACS trainings and application tests at IPI Laboratory. The Simulator can simulate the complete data workflow of clinical PACS from patient registration to exam ordering, and to image generation, image archive and display.

As shown in Fig. 3, the Simulator consists of six key components: RIS simulator, Acquisition Modality Simulator (AMS), DICOM gateway, PACS Controller, Viewing workstations, and a PACS monitoring system which will be discussed in more detail in the following paragraph. The PACS monitoring system is separate from the SM toolkit development presented in this paper. The AMS is connected to a clinical PACS and contains thousands of CT, MR, US, CR, and digitized film examinations in its local hard disk. The images are replenished continuously through the clinical PACS connection. With the patient information in the DICOM header of the image removed, these images can be used for simulation.

The PACS monitoring system is a software package developed to monitor all the image data flow going through the components of the PACS Simulator in real-time. The PACS monitoring system consist of clients and a centralized monitor server. The clients are installed in every component except the RIS Simulator to receive event message of each image data access activity generated by these components. The clients send the log data to the monitor server, which can display log data in a user interface for PACS users to trace image data workflow at each component in real-time. For example, the AMS simulated the CT chest image generation for patient ‘John Doe’. Meanwhile, the AMS also sent a log message of this generation activity to the local monitor client, which immediately forwards the message to the monitor server. The message includes such information as time and location of this activity, and patient name. The server then displayed the log message in its user interface (UI). When other Simulator components performed image data access activities, such as receiving the images, they also generated log messages and sent the messages to the monitor server. With all these log messages, the PACS monitoring system has collected the necessary log data for the SM toolkit, such as what, when and where

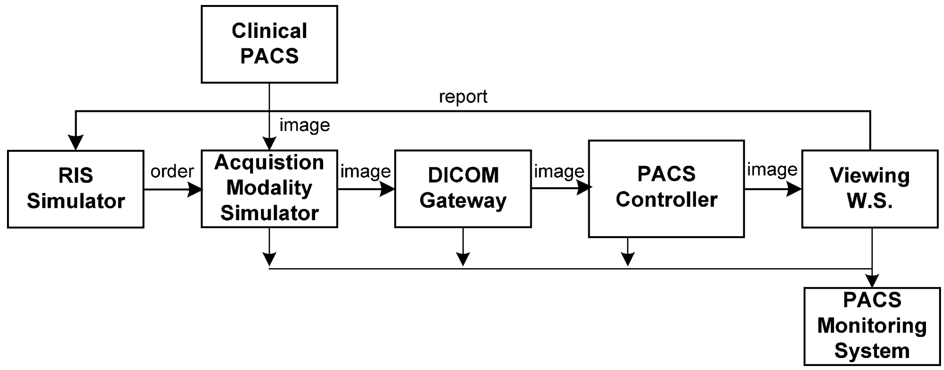


Fig. 3. Data workflow of the PACS simulator and PACS monitoring system.

images are accessed. However, PACS monitoring system has two limits:

1. No preservation of historical log data. Except a log database in the PACS controller, there is no preservation of any other log data in the monitoring system. The log database contains the information of every image data coming in and out the PACS controller.
2. No PACS user information. Additional resources need to be used to collect the information of users who performed the image access activity.

4.2. Evaluation set up

Because of the aforementioned limit of the PACS monitoring system, the SM toolkit can only monitor the data flow in the PACS controller utilizing the PACS monitoring system. However, it is crucial to monitor other components within the PACS Simulator, such as the DICOM gateway and viewing workstation. These two components are directly connected to the PACS controller. Because of this, their user login logs and computer system logs need to be collected by the SM toolkit as well. These data, along with log data already collected by the PACS monitoring client in PACS controller are used as input data for the evaluation of the SM toolkit. Fig. 4 shows the laboratory setup for the evaluation of the SM toolkit. The toolkit was connected to the PACS controller, DICOM gateway, and the viewing workstation. The pertinent information in the log database of PACS controller was collected and stored in the Auditing Database.

The user login information and application access information from the computer system logs of DICOM gateway and viewing workstations was also collected. The information was utilized to address the second limit mentioned above as to who performed the image data access activity. For example, user 'Michael' logged on a viewing workstation and made an image query/retrieve with the viewing software. A viewing software access event with the user name was recorded in computer system logs in the workstation. Meanwhile, PACS controller also recorded this query/retrieve activity in its log database. Comparing the time of the software access event and the time of the image query/retrieve and assuming that only a single user was using the viewing software at that time, these two events can be related together and the name of the user who did this query/retrieve can be obtained.

A graphical user interface (GUI) was connected to the SM toolkit for displaying the monitor information.

4.3. Test scenarios design

According to the perspective of clinical end-users, two test scenarios were designed for the evaluation.

- Test Scenario 1: Background scenarios
- Test Scenario 2: On-demand scenarios

Background scenarios are basically automatic storage functions, such as DICOM gateway sending images to PACS archive server, whereas on-demand scenarios are requests issued by end users, such as image query/retrieve at viewing workstations.

In order to simulate clinical 24/7 image automatic storage, a loop process was designed in the AMS to repeatedly send various types of modality images, such as CT, CR, MR and Ultrasound images, to DICOM gateway, which automatically forwards the images to PACS controller. The test scenario 2 was performed by on-demand query/retrieve from the viewing workstations. PACS

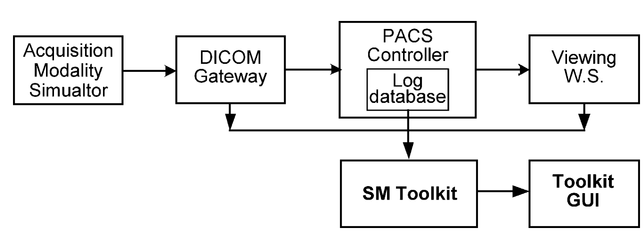


Fig. 4. Laboratory setup for evaluation of the SM toolkit.

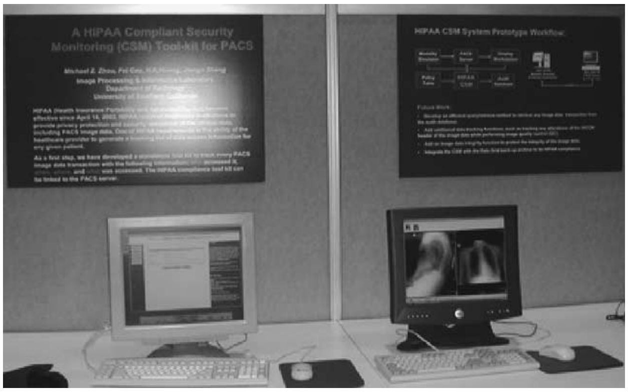


Fig. 5. Demonstration of the SM toolkit at InfoRAD exhibition of 2003 RSNA.

controller generates logs of image data access activities for both scenarios.

The laboratory evaluation is currently ongoing.

5. Results

We have successfully given a demonstration of the SM toolkit at the InfoRAD exhibition of 2003 RSNA (Radiology Society of North America) annual meeting. Fig. 5 shows the demonstration set up. The SM toolkit and it GUI were installed in a UNIX machine. The viewing software was installed in a PC to perform image query/retrieve, while the modality simulator software was installed in the same PC to perform automatic image storage.

6. Summary

The advent of HIPAA greatly impacts medical imaging systems, such as PACS, and even the entire health information systems. To be HIPAA compliant, every medical imaging system must satisfy the HIPAA requirement of audit trails.

In this paper, we proposed a HIPAA compliant auditing system (HCAS) for medical images in PACS. HCAS enables PACS to generate HIPAA compliant audit trails of image data access for a specific patient on demand. It also enables PACS to automatically monitor the image flow in the system, including detection of unauthorized usages of image data and other abnormal activities.

A security monitoring (SM) toolkit has been developed based on partial components of the Audit layer in HCAS. The toolkit can automatically monitor the image flow in PACS. Currently, the toolkit lacks the ability to generate HIPAA compliant audit trails, which is in developing now. A PACS Simulator has been developed for laboratory evaluation of the toolkit. The evaluation is currently ongoing.

Acknowledgements

This research is partially supported by a NIH Grant No. R01-LM06270.

References

[1] HIPAA, http://www.cms.hhs.gov/hipaa/hipaa2/general/background/pl104191.asp
[2] HIPAA, http://www.rx2000.org/KnowledgeCenter/hipaa/hipfaq.htm
[3] HIPAA Security Standard, 2003, http://www.cms.hhs.gov/hipaa/hipaa2/regulations/security/03-3877.pdf
[4] Cao F, Huang HK, Zhou XQ. Medical image security in a HIPAA mandated PACS environment. Comput Med Imaging Graph 2003; 27(2–3):185–96.
[5] HIPAA audit, http://www.datamirror.com/products/liveaudit/
[6] HIPAA audit tool, http://www.axolotl.com/press/20021113/
[7] HIPAA audit tool, http://www.peacefulpackers.com/it_solutions/hs2.htm
[8] Huang HK. PACS: basic principles and applications. Wiley–Liss; P. 291.
[9] Zhou X, Huang HK, Lou SL. Authenticity and integrity of digital mammography image. IEEE Trans Med Imaging 2001;20(8):784–91.
[10] Huang HK, Cao F, Zhou MZ, Mogel GT. Digital image envelope: method and evaluation. SPIE Med Imaging 2003.
[11] Syslog, http://www.loriotpro.com/Products/SyslogCollector/Syslog-DataSheet_ENv3.php
[12] White GB, Fisch EA, Pooch UW. Computer system and network security. Boca Raton: CRC Press; 1996.
[13] Law MYY, Zhou Z. New direction in PACS education and training. Comput Med Imaging Graphics 2003;27:147–56.
[14] Zhou Z, Huang HK, Cao F, Liu BJ, Zhang J, Mogel GT. RIS/PACS educational simulator. SPIE Med Imaging 2003;4:139–47.

Zheng Zhou, a PhD candidate of Biomedical Engineering Department at University of Southern California (USC), is a research assistant at Image Processing and Informatics (IPI) Laboratory, Radiology, USC now. He has 6 years experience in developments and researches on PACS and Medical Imaging Informatics. His current interest is on medical image security, grid computing for medical imaging, medical imaging informatics and PACS related applications.

Dr Brent Liu earned a PhD degree from the UCLA Biomedical Physics Graduate Program and performed research as a Post Doctorate. He is currently has a joint appointment with the Departments of Radiology, Keck School of Medicine, and Biomedical Engineering, Viterbi School of Engineering. He is also a senior research staff member of the Image Processing and Informatics Laboratory located at Marina del Rey. He has implemented fully filmless PACS in a clinical setting within the Imaging Department of both a high-profile community hospital (Saint John's Health Center, Santa Monica) and a high-profile academic hospital (UCLA) that has multiple campus sites and is currently advising multiple hospitals on their PACS process, including the USC Health Science Campus. His research areas of interest include Medical Imaging Informatics, Picture Archiving and Communication Systems (PACS) clinical uptime and usability, new PACS technology, Disaster Recovery for PACS, design and implementation of high-resolution image display workstations, next generation Internet and its clinical applications, and advances in the area of image processing and information management for healthcare including Security and HIPAA-compliance related issues.

RSNA 2004 PAMPHLET

Image Processing & Informatics (IPI) Laboratory
Department of Radiology

Staff & Collaborators

2004 RSNA

Presentations & Scientific Exhibits

November 28 – December 3, 2004
McCormick Place, Chicago

4676 Admiralty Way, Suite 601
Marina Del Rey, California 90292
310 448-9440 Office
310 448-9441 Fax

TABLE OF CONTENTS

QUICK REFERENCE FOR ORAL PRESENTATIONS

Date	Time	Location	Title	Author
11-29/04 Monday	**3:30PM to 3:50PM**	Mobile Computing Pavilion	**MCP16** ● One-Year Clinical Experience at Saint John's Health Center Utilizing the PDA as a Wireless Remote Control for Clinical image Distribution	**J. Documet, BS.**
11-29-04 Monday	**3:30PM**	InfoRAD Theater/Electronic Medical Record	**9305 EMR-i** ● Distribution of DICOM - RT ePR-based Radiotherapy Information	**M. Law, Ph.D.**
12-1-04 Wednesday	**11:40AM**	Rm N229 - Pediatric Muscular Imaging	**SSK16-08** ● How Accurate is the Greulich and Pyle Atlas for Bone Age Assessment of Children Today?	**H.K.Huang, D.Sc.**
12-1-04 Wednesday	**3:20PM**	Rm S404-CD Radiology Informatics	**SSM21- 03** ● A Consistent and Repeatable Clinical Protocol for High Performance International Internet-2 Connection in Remote Medical Image Transmission	**L. W. Chan, Ph.D.**
12-1-04 Wednesday	**12:25PM**	Hall D1, Lakeside Center	**092ORI-p** ● The Development and Implementation of a Prototype DICOM-based Radiotherapy Information System	**M. Law, Ph.D.**
12-2-04 Thursday	**10:00AM**	InfoRAD Theater PACS Work-station Classroom	**9604 PACS-i** ● Web-based Collaborative Medical Applications with Image-Based Composite Objects (IBCO)	**J. Zhang, Ph.D.**
12-2-04 Thursday	**11:30AM**	InfoRAD Theater Web Classroom	**9710 NT-i** ● The Data Storage Grid: The Next Generation of Fault-Tolerant Storage for Backup and Disaster Recovery of Clinical Images	**B.J. Liu, Ph.D.**
12-2-04 Thursday	**12:35PM**	Hall D1, Lakeside Center	**0927RI-p** ● Lossless Digital Signature Embedding (LDSE) for Image Integrity in a Fault-tolerant Clinical Off-site Backup Archive	**Z. Zhou, M.S.**
12-2-2004 Thursday	**12:55PM**	Hall D1, Lakeside Center	**0929RI-p** ● A HIPAA Compliant Architecture for Securing Clinical Images	**B.J. Liu, Ph.D.**
On-Going INFO RAD Exhibits: 8:00AM thru 5:00pm Sun, Mon, Tue, Wed, Thu 8:00am thru 12:45pm Fri				
11-28-04 to 12-3-04	**On-going**	InfoRAD Lakeside Center	**9113 DS-i** ● An Intelligent Detection of Idiopathic Scoliosis In Chest Computed Radiography Under Picture Archiving and Communication System (PACS) Environment	**F.H. Tang, Ph.D.**
11-28-04 to 12-3-04	**On-going**	InfoRAD Lakeside Center	**9114DS-i** ● Remotely Accessible Computer Assisted Skeletal Maturity Assessment	**E. Pietka, Ph.D.**
11-28-04 to 12-3-04	**On-going**	InfoRAD Lakeside Center	**9221 ED-i** ● Collaborative Augmented Reality Environment for Liver Visualization	**C. Choy, Ph.D.**
11-28-04 to 12-3-04	**On-going**	InfoRAD Lakeside Center	**9613 PACS-i** ● Multi-dimensional Intelligence Integrated Server for Distributed Picture Archiving and Communication Systems	**F. H. Tang, Ph.D.**
11-28-04 to 12-3-04	**On-going**	InfoRAD Lakeside Center	**9618 PACS-i** ● Educational RIS/PACS Simulator Integrated with HIPAA Compliant Architecture (HCA) for Auditing	**Z. Zhou, M.S.**
11-28-04 to 12-3-04	**On-going**	InfoRAD Lakeside Center	**9710 NT-i** ● The Data Storage Grid: The Next Generation of Fault-Tolerant Storage for Backup and Disaster Recovery of Clinical Images	**B. J. Liu, Ph.D.**
11-28-04 to 12-3-04	**On-going**	InfoRAD Lakeside Center	**9305 EMR-i** ● Distribution of DICOM - RT ePR-based Radiotherapy Information	**M. Law, Ph.D.**
11-28-04 to 12-3-04	**On-going**	InfoRAD Lakeside Center	**9604 PACS-i** ● Web-based Collaborative Medical Applications with Image-Based Composite Objects (IBCO)	**J. Zhang, Ph.D.**
11-28-04 to 12-3-04	**On-going**	InfoRAD Lakeside Center	**9902 PDA-i** ● One-Year Clinical Experience at Saint John's Health Center Utilizing the PDA as a Wireless Remote Control for Clinical image Distribution	**J. Documet, BS,**

MOBILE COMPUTING PAVILION PRESENTATION

Date: Monday, November 29, 2004
Time: 3:30 PM - 3:50 PM
Location: Mobile Computing Pavilion

MCP16
One-Year Clinical Experience at Saint John's Health Center Utilizing the PDA as a Wireless Remote Control for Clinical image Distribution
J. Documet, BS, Marina Del Rey, CA ● B.J. Liu, PhD ● L. Documet, PhD (documet@usc.edu)

Last year an application to perform wireless remote control of PACS image distribution utilizing a Personal Digital Assistant (PDA) was presented. This paper provides the clinical experiences that have been encountered after a year of operation at Saint Johns Health Center. This application is used to distribute PACS exam data to diagnostic/review workstations, a PACS web server, a teleradiology system and a CD burning device. In addition, a preview of the examination to be routed is displayed as a thumbnail to the user. This new feature provides a way for the users to verify that the chosen exam is the correct one before distribution. Clinical reviews form radiologists and other clinical staff was utilized to develop a user-friendly application. The easiness to operate the application simplifies PACS exams distribution.
LEARNING OBJECTIVES: 1. Demonstrate the capability to manage and distribute PACS image data from a PDA with a wireless network connection. 2. Provide a thumbnail capability of the examinations in the PDA as a preview before distribution. 3. Provide clinical review from clinical users.

EDUCATION EXHIBITS

Oral Presentation

Date: Wednesday ● December 1, 2004
Time 3:20 PM
Location: Room: S404-CD
SESSION: Radiology Informatics (Internet 2, Image Compression)

SSM21- 03
A Consistent and Repeatable Clinical Protocol for High Performance International Internet-2 Connection in Remote Medical Image Transmission
L. W. Chan, PhD, Hong Kong, China ● H. K. Huang, DSC ● F. Cao, PhD ● Z. Zhou, MS ● M. Law ● F.H. Tang, PhD, I (orlchan@polyu.edu.hk)

Internet-2 is high performance advanced network initiated by the University Corporation for Advanced Internet Development (UCAID) with the effort of US research universities. The high bandwidth, low latency, quality of service (QoS), multicast function and security assurance of Internet-2 facilitate remote applications such as rapid distribution of large-volume medical images and interactive teleconsultation. In North America, over 200 research universities and laboratories are connected by Abilene backbone that currently has 10 Gigabit-per-second capacity and about 50 ms round-trip-time across the US continent. However, global medical image and knowledge exchange requires a tremendously long and expensive connection from international partners outside US to a Chicago-based optical infrastructure for international Internet-2 access, called StarLight. With limited local support, the bandwidth of the existing Connection from the Hong Kong Academic and Research. Network (HARNET) to StarLight is 45 Megabit-per-second (Mbps) only. This connection has been used for daily medical image exchange between the Hong Kong Polytechnic University (PolyU) and the University of Southern California since October 2002. In October 2003, a. multiple end-point video-teleconference connecting PolyU, University of Hawaii and University of Pittsburgh was successfully held. Such bandwidth is sufficient for video streaming. When it comes to remote synchronized image manipulation, the high latency of the international connection becomes a large barrier to synchronize data and instructions between remote sites. Our Hong Kong Internet-2 community and Web100 project have been designated to derive tuning protocols to tackle this problem by optimizing the TCP/IP parameters. The preliminary nary result shows that the network throughput with tuning is 14 Mbps, while it was only 1.5 Mbps before tuning. Since the tuning protocol should be consistent and repeatable, in the clinical environment, the protocol is required to be self-contained in the kernel of operating system. The methodology, clinical implementation, performance and advantage of the proposed tuning protocol will be presented and discussed.

SCIENTIFIC EXHIBITS

Oral Presentation:

Date: Wednesday ● December 1, 2004
Time: 11:40 AM
SESSION: Pediatric (Musculoskeletal Imaging)
Location: Room N229

SSK16-08
How Accurate is the Greulich and Pyle Atlas for Bone Age Assessment of Children Today?
H. K. Huang, DSC, Marina Del Rey, CA ● J. W. Sayre, PhD ● V. Gilsanz, MD, PhD, (hkhuang@aol.com)

PURPOSE: Bone age assessment (BAA) is a frequent procedure performed on pediatric patients to evaluate growth and is key for growth disease diagnosis/management. Most common method for BAA is matching by a left wrist radiograph against a small reference set atlas of normal standards developed by Greulich & Pyle (G&P) in the 1950s from data collected by the Brush Foundation Study. Due to increasing racial diversity and changing nutrition/behavior, the G&P method may not accurately reflect skeletal development for today's children. This study compares the difference in BAA using G&P method with a large set of racially diverse normal hand images of present-day children.
METHOD AND MATERIALS: We have developed a digital hand atlas with 1,080 images collected from boys and girls of European, African, Hispanic and Asian descent at the Childrens Hospital of Los Angeles; 5 images per age group of pre-pubertal children and 10 images during puberty of uniform distribution. All relevant patient information have also been collected and stored along with the digital images. The digital atlas contains features of 3 fingers including bone growth regions and phalange geometric parameters.
RESULTS: Two pediatric Radiologists performed BAA. There are two significant results: 1) The collected normal data was validated by three existing standards. 2) A difference of approximately 10 months lower in bone age was noted between both the digital atlas mean and the G&P mean over the 5-14 year old children as compared to the children of the Brush Foundation Studies. Plots and tables of various comparisons among ethnic origin and gender will be exhibited.
CONCLUSIONS: This discovery of the differing bone age assessments is puzzling since better nutrition and living environments today should produce the opposite effect. Possible explanations for these findings are related to at least one of four factors: true population change and ethnicity; differences in patient recruitment methods; quality of image presentation; and quality/training of the readers. This study will be further continued by enlarging the sample sizes and increasing statistical power and considering regional patient distribution effect, which may have skewed analysis.

Poster Sessions:

Date: Wednesday ● December 1, 2004
Time: 12:25 PM
Location: Hall D1, Lakeside Center

0920RI-p
The Development and Implementation of a Prototype DICOM-based Radiotherapy Information System
M. Law, PhD, Hong Kong, China ● H.K. Huang, DSC ● L.WC. Chang, PhD ● X. Zhang, PhD ● J. Sun, PhD, (ormaria@polyu.edu.hk)

PURPOSE: DICOM-RT objects were ratified by 1999 but were neither totally nor widely used because of its complexity and its tediousness in implementation. The objective of this research is to develop a prototype DICOM-based Radiotherapy Information System for the radiotherapy community to appreciate the advantages of a common standard for radiotherapy information and to persuade for its implementation.

METHOD AND MATERIALS: The workflow of a generic radiotherapy department and user requirements were reviewed, from which the data flow was designed. A prototype of the system consisting of a Radiotherapy Modality Simulator, a RT Gateway, a RT Archive Server, a RT Web Application Server and the Web client was designed. DICOM-RT objects and non-DICOM files were collected from various RT-related imaging and radiotherapy systems. All non-DICOM files were converted into the DICOM format. The data files were then organized into DICOM-based virtual patients to test out the data flow through the system. Usability testing, including connectivity with Treatment Planning System and analysis by typical end-users was conducted for evaluation of the system. The evaluation results were used for iterative modifications to the system.
RESULTS: 1. A prototype of the system was developed and implemented 2. Eleven windows with graphic user interface for the web client were designed for use by the end-users who are radiation oncologists and radiation therapists. 3. Data of ten virtual patients were input to the RT archive server, and reviewed at the web client workstation. The data flow was verified at the RT Archive Server and the RT Web Application Server. The data integrity was reviewed and verified at the web client. 4. Transmission of DICOM-RT objects from a treatment planning system using DICOM export to the RT archive server was successful. 5. End users were invited to perform the evaluation and their feedback was adopted to modify the system.
CONCLUSIONS: The development of the DICOM-based RT Information System will facilitate the communication of RT information between equipment from different vendors, and improve the productivity of the Oncology Department and the healthcare efficiency.

Date: Thursday ● December 2, 2004
Time: 12:35 PM
Location: Hall D1, Lakeside Center

0927RI-p
Lossless Digital Signature Embedding (LDSE) for Image Integrity in a Fault-tolerant Clinical Off-site Backup Archive
Z. Zhou, MS, Marina Del Rey, CA ● H.K. Huang, DSC ● B.J. Liu, PhD, (zhengzho@usc.edu)

PURPOSE: Image integrity is a vital issue in clinical off-site backup archive, where the image data are not under the security protection of local PACS anymore. Our previous work on image integrity uses a digital envelope method, which would permanently change the least significant bit (LSB) of some image pixels. This method is adequate for most current medical imagings since the LSB contains only noise. However, due to the higher density resolution of newer imaging modalities in the horizon, the LSB may also contain real data. We have developed a lossless digital signature embedding (LDSE) method, which can completely recover the original image pixels, and assure image integrity.
METHOD AND MATERIALS: The data to be embedded in this method is the digital signature (DS) generated from the original image. The DS, which assures the image integrity, is embedded in the LSB of the image pixels at the acquisition gateway. The image integrity can be verified when the embedded image is received by the off-site backup archive, or when it is retrieved by workstations. In order to completely recover the LSB of the original image changed by the embedding, a lossless data embedding method is developed. In particular, the LSB of the image pixels to be embedded is first determined and extracted before the actual embedding; the extracted bits are then compressed with a lossless compression method; the compressed LSBs and DS are finally concatenated together and embedded into these image pixels. The compressed LSBs can be extracted and de-compressed, thus completely restoring the image pixels.
RESULTS: A PACS simulator and a fault-tolerant clinical off-site backup archive were used to evaluate the performance of LDSE method. The embedded DS is extracted to verify the data integrity of various types of clinical imagings including CT CR, MRI, and US. The verification demonstrates that image pixels can be completely restored with LDSE method. Performance of the LDSE was acquired and tabulated.
CONCLUSIONS: A LDSE method has been developed to assure the image integrity in clinical off-site backup archive without inducing any changes to the original image pixels. The performance of LDSE was evaluated at laboratory and clinics.

Date: Thursday ● December 2, 2004
Time: 12:55 PM
Location: Hall D1, Lakeside Center

0929RI-p
A HIPAA Compliant Architecture for Securing Clinical Images
B.J. Liu, PhD, Marina Del Rey, CA ● Z. Zhou, MS ● H.K. Huang, DSC, (brentliu@usc.edu)

PURPOSE: HIPAA Security Standards mandate health institutions to protect health information against unauthorized use/disclosure. This can be achieved by utilizing access control and generating audit trails. Although most current clinical image systems (eg, PACS) have components that generate log files, there is a lack of methodology to obtain and synthesize pertinent data from the large volumes. We have designed and developed a HIPAA Compliant Architecture specifically for tracking and auditing the image workflow of clinical imaging systems such as PACS.
METHOD AND MATERIALS: The HIPAA Compliant Architecture developed comprises of Four Layers: Record Layer, Audit Layer, Notification Layer, and Action Layer. The Record Layer consists of various log data (eg, PACS and RIS applications logs) as input sources. The Audit layer includes a centralized auditing database as well as audit analysis and interpretation tools. Information is collected from the Record Layer and normalization is performed prior to storing in the database. Analysis is performed on the database by Intrusion detection to discover abnormal access patterns. Role-based policy is utilized to aid in the decision of whether the abnormal activity is unauthorized or necessary under special circumstances. The Notification Layer receives the warning/alert messages from the Audit Layer and notifies PACS end users. The Action layer is designed for the PACS coordinator or manager to take actions such as limiting access control and is not included in the development of this architecture.
RESULTS: As an initial step, a software toolkit was implemented based on the above HIPAA Compliant architecture. it was implemented within a PACS Simulator located in the IPI Lab at USC. Scenarios were developed where different user types performed legal/illegal access of PACS data in the system. Results were evaluated based on whether scenarios were correctly identified and documented. Integration and implementation pitfalls were also noted and included.
CONCLUSIONS: A HIPAA Compliant Architecture was developed and designed to provide clinical imaging systems with the ability to assist in the protection and enforcement of unauthorized access to clinical image data.

INFORAD EXHIBITS

9113 DS-i
An Intelligent Detection of Idiopathic Scoliosis In Chest Computed Radiography Under Picture Archiving and Communication System (PACS) Environment
F.H. Tang, PhD, Kowloon, China ● L.W.C. Chan, PhD ● P. Wong, BS, (orfhtang@polyu.edu.hk)

Collection of digital image database in Picture Archiving and Communication System (PACS) has provided a useful means for improved image content accessibility. Idiopathic scoliosis is sometimes found in chest radiography of university students. In order to conduct retrospective studies of such disease, extraction of image features based on image content becomes crucial. To identify images with scoliosis, chest images are retrieved from the PACS image server and processed by an application server for content analysis. Automatic detection of image content can be enhanced by the fuzzy logics for the segmentation of vertebrae on the digital images retrieved from the PACS system. The intelligent classifier then identifies cases of scoliosis automatically. In addition to human interpretation, our method provides a more sensitive and efficient tool for detection of the abnormal feature in chest images.
LEARNING OBJECTIVES: 1.By studying this exhibit, the learner will understand how fuzzy logic can help in detection of spine. 2.The learner will be aware of how computer detection method can be applied to Chest computed radiography.

9114 DS-i

Remotely Accessible Computer Assisted Skeletal Maturity Assessment

E. Pietka, PhD ● A. Gertych, PhD, Gliwice, Poland ● K. Witko, MS, (pietka@ps.edu.pl)

Computer aided skeletal maturity assessment, being basically a tool, which assists the radiologists in performing the image diagnosis, has recently switched to a more sophisticated application system. Since image processing functions as well as pattern recognition procedures require a set of features, automatically extracted from hand radiographs, to be stored in a database system as a set of standards. A comparison with corresponding findings extracted from radiographs of normally developed subjects permits the bone age to be assessed. Interactions, performed by physicians during the diagnostic process, draw the attention towards a user-friendly graphical user interface (GUI). The WEB technology implemented in designing the GUI permits the system to be accessed remotely. Via the GUI data can be accessed in read only or read/write mode. Depending on the user permission access, the SQL select or SQL update query can be granted. LEARNING OBJECTIVES: (1) become familiar with the implementation of WEB technology implemented in designing the graphical user interface, (2) study the CAD design for the computer aided skeletal maturity assessment, (3) remote access to the standard data collected from a normal population, (4) review the image analysis methodology.

9221 ED-i

Collaborative Augmented Reality Environment for Liver Visualization

C.S. Choy, PhD, Hong Kong, China ● F.H. Tang, PhD, (mccliff@polyu.edu.hk)

Visualizing of the complex structure of liver is important in training for surgeons. While 3D physical model is good for discussions and sharing of knowledge, the model can be highly complicated and expensive to make, especially if it has to reveal soft tissues, bones and vessels and their connections, and different cross-sectional views. Also, physical model requires physical presence of participants, which makes remote collaboration difficult. In this paper, we will demonstrate a system for collaborative augmented reality, in which participants with stereoscopic video-see-through head-mounted displays can see a shared virtual space with 3D virtual liver. Through a tangible device, participants can dissect, manipulate and add annotations to the virtual liver to communicate more effectively with others. In case when participants are separated physically, they can also supplement with audio communications.
LEARNING OBJECTIVES: (1) To allow participants wearing head mounted displays (HMD) to visualize a shared virtual 3D liver with details including soft tissues, bones and blood vessels (2) To allow participants with suitable tangible devices to dissect, to manipulate parts of the virtual liver, to add virtual pointers, and to add annotations for collaborative discussion.

9613 PACS-i

Multi-dimensional Intelligence Integrated Server for Distributed Picture Archiving and Communication Systems

F.H. Tang, PhD, Kowloon, China ● A. Lee, PhD ● L.W.C. Chan, PhD ● M. Law, PhD, (orfhtang@polyu.edu.hk)

A coordination server (CS) is developed to integrate PACS servers distributed at different locations using web-based technology to facilitate data inter-change while the autonomy of individual PACS servers is preserved. In addition to ad-hoc information query and retrieval from the CS servers both at local and remote sites, an intelligence-enhanced program for automatic detection can be pushed from the CS to the PACS servers distributed at different locations to detect incidence of the disease at their location. When a case is detected in a distributed server, the result is sent back to the CS where a directory is created to keep track of patient records. Then the incidence of the disease in that location can be known. When further analysis is required, the case can be retrieved from the distributed PACS. Thus the intelligence-enhanced PACS would help to detect and monitor the spread of a particular disease at different locations.
LEARNING OBJECTIVES: 1.By studying this exhibit, the learner will understand the working principle of coordination server for PACS. 2. The learner will be aware of the feasibility of using distributed computing method for detection of particular image pattern.

9618 PACS-i
Educational RIS/PACS Simulator Integrated with HIPAA Compliant Architecture (HCA) for Auditing
Z. Zhou, MS, Marina Del Rey, CA ● B.J. Liu, PhD ● HK. Huang, DSC ● J. Zhang, PhD, (zhengzho@usc.edu)

Many educational courses have been designed for training radiologists and allied healthcare providers to operate PACS workstations. However, there are yet tools available for training of PACS concept and workflow analysis. In 2002 RSNA we demonstrated an educational RIS/PACS Simulator for such purpose. In 2003 we added the Web distributions. The Simulator has seven components simulating a typical clinical RIS/PACS: RIS simulator, modality Simulator, DICOM gateway, PACS server, viewing workstations, monitoring system, and the web server. Due to the HIPAA mandates of healthcare providers to have the ability to generate audit trails on the data access for any specific patient, we add in this year a new feature in the Simulator with HIPAA compliant Architecture (HCA) for auditing the image flow of PACS. The added HCA, which extracts pertinent information from various PACS log data, stores it in a centralized audit database, and performs HIPAA audit analysis based on the audit database.
LEARNING OBJECTIVES: 1. Trace image data access for a specific patient through HCA in the forms: who, when, where, what, and how, 2. Monitor the user actions on the RIS/PACS data from HCA, 3. Observe clinical RIS/PACS operation, component by component, with the HIPAA trails, 4. Induce failure in a component to observe its impact on the entire RIS/PACS operation.

9902 PDA-i
One-Year Clinical Experience at Saint John's Health Center Utilizing the PDA as a Wireless Remote Control for Clinical image Distribution
J. Documet, BS, Marina Del Rey, CA ● B.J. Liu, PhD ● L. Documet,BS (documet@usc.edu)

Last year an application to perform wireless remote control of PACS image distribution utilizing a Personal Digital Assistant (PDA) was presented. This paper provides the clinical experiences that have been encountered after a year of operation at Saint Johns Health Center. This application is used to distribute PACS exam data to diagnostic/review workstations, a PACS web server, a teleradiology system and a CD burning device. In addition, a preview of the examination to be routed is displayed as a thumbnail to the user. This new feature provides a way for the users to verify that the chosen exam is the correct one before distribution. Clinical reviews form radiologists and other clinical staff was utilized to develop a user-friendly application. The easiness to operate the application simplifies PACS exams distribution.
LEARNING OBJECTIVES: 1. Demonstrate the capability to manage and distribute PACS image data from a PDA with a wireless network connection. 2. Provide a thumbnail capability of the examinations in the PDA as a preview before distribution. 3. Provide clinical review from clinical users.

Date: Monday, November 29, 2004
Time: 3:30 PM
Session: Electronic Medical Record
Location: InfoRAD Theater

9305 EMR-i
Distribution of DICOM - RT ePR-based Radiotherapy Information
M. Law, Hong Kong, China ● L.WC. Chan ● H.K. Huang, DSC ● X. Zhang ● J. Sun (ormaria@polyu.edu.hk)

This exhibit demonstrates a prototype Web-based DICOM-RT ePR radiotherapy information system implemented in our Laboratory. The prototype consists of a DICOM-RT Archive Server used for storage of DICOM- based radiation therapy information, a web client/server for information distribution. Graphic user interface for operation in radiotherapy was designed for the client windows. Patient information including treatment plans, images and therapy records in DICOM format (or translated to DICOM format if non-DICOM) is stored in the RT Archive Server and routed to the web server. The patient image and treatment information can be queried and retrieved at the client workstation. This exhibit demonstrates the retrieval of virtual radiation therapy patient treatment information including images from the web server to the web client.
LEARNING OBJECTIVES: 1.The use of DICOM standard for integration of radiotherapy patient image and treatment information. 2. The distribution of radiotherapy information using Web technology. 3. A model radiotherapy electronic patient record

Date: Thursday, December 2, 2004
Time: 10:00 AM
Location: InfoRAD Theater – PACS Workstation Classroom

9604 PACS-i
Web-based Collaborative Medical Applications with Image-Based Composite Objects (IBCO)
J. Zhang, PhD Shanghai, China ● J. Sun, PhD ● Y. Tan, MS ● Y. Yang, MS ● B. Gao, BS ● H.K. Huang, DSC, (jzhang@mail.sitp.ac.cn)

We have developed a Web-based system to interactively display and manipulate image-based composite objects (IBCO) including 2D, 3D and 4D images, graphics, and therapy records for intranet and Internet collaborative applications. The system consists of IBCO gateways, a Web-based IBCO Repository Server, and IBCO Viewers. The gateway is used to import IBCO, and export them to the IBCO Repository Server. The Server stores the IBCO, and has a Web-based interface to communicate with IBCO Viewers. The Viewer, plugged in a Web browser, can display and manipulate IBCO. During networked collaborative applications, both local and remote users can use Viewers to retrieve IBCO objects from the Server, and to display and manipulate these objects interactively. The Viewer has remote control mechanism on both local and remote objects to synchronize operative functions. We have successfully used this system in teleradiology, distant learning, and teleconferencing on cardiology for three months.
LEARNING OBJECTIVES: 1.How to create and manage IBCO from actual patient records; 2.Use Web technology and architecture for collaborative medical applications; 3.The value of using interactive methods with IBCO for difficult cases; 4.The advantages of on-time interactive communication in improving the efficiency and quality of collaborative healthcare.

Date: Thursday, December 2 , 2004
Time: 11:30 AM
Location: InfoRAD Theater – Web Classroom

9710 NT-i
The Data Storage Grid: The Next Generation of Fault-Tolerant Storage for Backup and Disaster Recovery of Clinical Images
B.J. Liu, PhD, Marina Del Rey, CA ● Z. Zhou, MS ● J. Documet, BS ● N. King, PhD ● H.K. Huang, DSC, (brentliu@usc.edu)

Grid Computing represents the latest most exciting technology to evolve from the familiar realm of parallel, peer-to-peer and client-server models. We have researched and developed a novel Data Grid testbed involving several federated PACS based on grid architecture. By integrating a grid computing architecture to the DICOM environment, a failed PACS archive can recover its image data from others in the federation in a timely and seamless fashion. The design reflects the five-layer architecture of grid computing: Fabric, Resource, Connectivity, Collective, and Application Layers. This exhibit will display the Data Grid architecture with three simulated federated PACS with an archive failure event and its timely recovery with 99.999% uptime. The successful demonstration of the Data Grid in the testbed will provide an understanding of the Data Grid concept in clinical image data backup as well as establishment of benchmarks for performance from future grid technology improvements.
LEARNING OBJECTIVES: 1.Introduction to Grid Technology and Five- Layer Architecture. 2.Understand the importance of clinical image recovery and Continuous Availability (CA 99.999%). 3. See the application of the Data Grid for clinical image recovery. 4. See the Data Grid application recover clinical image data in a simulated downtime event.