ELSEVIER

Computerized Medical Imaging and Graphics 29 (2005) 367–383

Computerized Medical Imaging and Graphics

www.elsevier.com/locate/compmedimag

Protecting patient privacy against unauthorized release of medical images in a group communication environment

Mingyan Li[a], Radha Poovendran[a,*], Sreeram Narayanan[b,1]

[a]*Department of Electrical Engineering, University of Washington, Seattle, WA 98195, USA*
[b]*Department of Radiation Oncology, University of Washington, Seattle, WA 98195, USA*

Received 9 November 2004; accepted 8 February 2005

Abstract

In this paper, we identify and study an important patient privacy protection problem related to medical images. Following Health Insurance Portability and Accountability Act (HIPAA) mandate on privacy protection of patients' medical records, efforts have been devoted to guaranteeing the confidentiality of data and medical images during storage and transmission via an untrustworthy channel. However, to our knowledge, there has not been any effort towards protecting against unauthorized release of images by an authorized recipient. In this paper, we study the problem of tracing illegally distributed medical images in a group communication environment and identify a set of design requirements that must be met. We propose a fingerprint model suitable for many-to-many multicast, that is computationally efficient and scalable in user storage and key update communication. Simulation results also show that our scheme is highly robust to typical medical image processing and collusion attacks, while yielding high quality watermarked images.

Keywords: Medical image security; Watermarking; Multicast; Fingerprinting; Privacy protection; DICOM; HIPAA

1. Introduction

Privacy protection of medical images has always been an important issue in the management of patients' medical records. As part of Health Insurance Portability and Accountability Act (HIPAA), a set of standards for privacy protection of health data issued by the federal government took effect on April 14, 2003 [1]. The HIPAA mandates hospitals, medical professionals, and other health providers to ensure 'confidentiality and integrity of individually identifiable health information, past, present or future'.

As digital technology pervades our society, a vast number of medical images now exist in electronic format for easy storage, maintenance, and retrieval. Ubiquitous wired and wireless networks make it possible to access and share data among medical personnel, to promote high quality care for patients. However, the convenience of data access and distribution poses a great threat on privacy of patients' information. Constant efforts are being made to provide security solutions [2–5] to ensure (i) medical image transmission cannot be accessed by unauthorized parties (confidentiality), (ii) images are not modified during transmission (integrity), and (iii) images have originated from the correct sources to the claimed receivers (authentication). Continuously updated Digital Imaging and Communication in Medicine (DICOM) standards provide guidelines to ensure authentication, integrity and confidentiality of medical images [2].

Security measures in DICOM [2] and the research on medical image security [3–5] focus on secure storage and secure transmission, before reception. However, after reception it is possible for a recipient to distribute a patient's data to unauthorized parties, hence violating the patient's privacy. To the best of our knowledge, currently no one has addressed the problem of guaranteeing patient's privacy after the data is accessed by an authorized recipient. Hence the HIPAA standards have not been fully addressed. We intend to fill the gap between the HIPAA mandate

* Corresponding author. Address: Department of Electrical Engineering, Paul Allen Center, AE 100R, Campus Box 352500, University of Washington, Seattle, WA 98195, USA. Tel.: +1 206 221 6512; fax: +1 206 543 3842.

E-mail address: radha@ee.washington.edu (R. Poovendran).

[1] The project was completed when he was a PhD student in the Department of Electrical Engineering at University of Washington.

0895-6111/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compmedimag.2005.02.003

and current privacy mechanisms by studying the problem of tracing unauthorized disclosures of medical images.

In order to understand the complexity of the problem, we first need to consider the commercial model involved. To provide high quality medical care, a team of collaborative physicians is often formed for a patient's case. These physicians exchange patients' data such as images, and diagnostic reports via public networks. The medical personnel can be geographically dispersed, and hence naturally form a group communication network. Since members of the group should be able to exchange opinions, one-to-many (a single sender multiple recipients), or many-to-many (multiple senders and multiple recipients) multicast communication mode can be adopted to reduce a sender's computation and the network bandwidth consumption during group communication. In such a group communication setting, an efficient tracing scheme needs to be developed such that leaked images can be traced back to the infringing individual source.

In this paper, we study the problem of tracing illegal distribution of medical images. Given a leaked image, we use watermarking techniques [6] to trace back to the authorized entity that initiated illegal distribution of image. We identify a special requirement on watermarking medical image for tracing purposes: high image fidelity, and robustness to frequency selective operation in medical images while withstanding conventional attacks [7]. We formulate the tracing problem in a group communication, aiming to take advantage of the efficiency of multicast communication, while enabling tracing with the use of watermarks. We propose a solution that is not only suitable for many-to-many communication, but also scalable in user's storage and efficient in update communication and computation cost. We confirm the feasibility of our solution in medical environments by analysis and simulation of various modality of medical images.

The structure of this paper follows. Section 2 presents related work and a review on watermarking techniques. In Section 3, we identify requirements that need to be met for watermarking medical images for tracing purpose, and formulate the problem of tracing in a multicast environment. In Section 4, we present a solution to the medical image tracing problem as an explicit algorithm. We elaborate the implementation consideration and parameter selection, and analyze estimation attacks on the proposed model. In Section 5, we evaluate our watermarking scheme in terms of robustness and imperceptibility by extensive simulation. Finally, we summarize our results, and present open problems in Section 6.

2. Background on watermarking

Watermarking [6] is a technique of embedding identification codes, called watermarks, into cover media or host media (images, text). Watermarking is used for the protection of intellectual property, data integrity, and data authentication [7]. Under the assumption of a unique watermark per user, watermarking can be used as *fingerprinting*. A watermarking system consists of two components: a watermark embedder, and a watermark detector, as illustrated in Fig. 1.

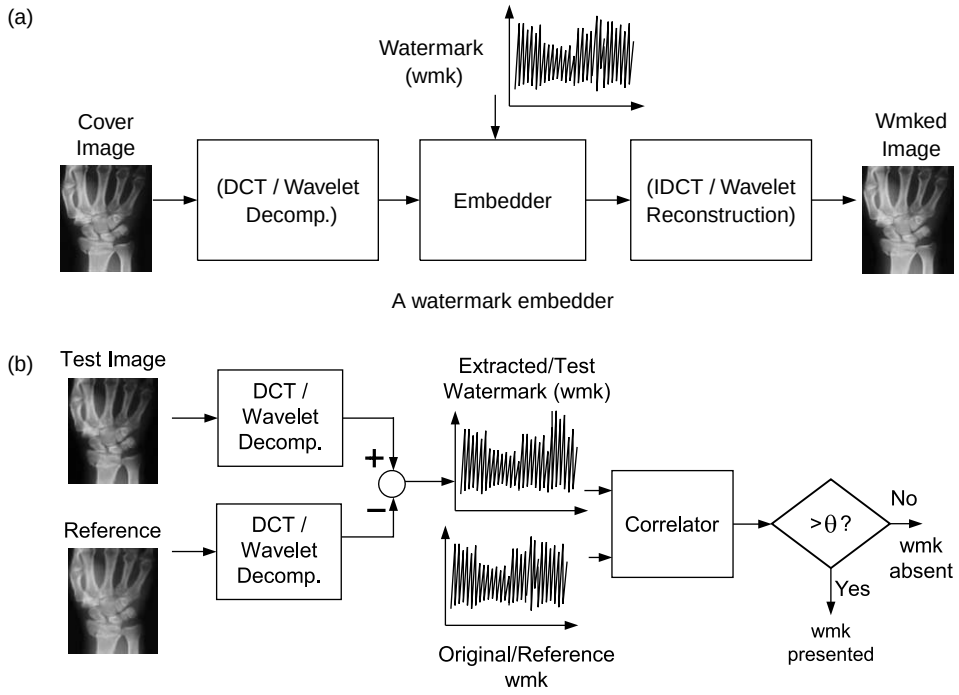

Fig. 1. A transform domain watermarking system: a watermark embedder and a watermark detector.

Watermark embedding is performed in either the spatial domain or the transform domain [6]. At the detection module, original images as well as the watermarks are assumed to be available. To examine the presence of the watermark, the detector extracts a test watermark and calculates the correlation between the reference and the test watermark. If the correlation is above a pre-specified threshold θ, the reference watermark is determined to be present. In what follows, we first review the related work on watermarking for medical images.

2.1. Related work on watermarking for medical images

The applications of watermarking for integrity protection and authentication of medical images has been investigated extensively in [3–5,8]. Cao et al. [3] presented a HIPAA compliant image security architecture that provides confidentiality, integrity and authenticity of images in an open transmission channel. Fridrich et al. [8] proposed 'invertible watermarks' for integrity protection, and these invertible watermarks can be completely erased once the integrity of an image has been verified. Coatrieux et al. [4] studied the applications of watermarking in medical imaging, including data hiding, integrity control, authenticity, and discussed the requirements for each application. In order to reduce storage and bandwidth consumption, Acharya et al. [5] applied watermarking for interleaving patient information with medical images.

At present, to our knowledge, no study on using watermarking to trace illegal disclosure of medical images is available. Watermarks used for image authenticity and integrity protection need to be sensitive to any small modifications to the images, and cannot be directly applied for tracing, because the watermarks used for tracing purposes must be highly robust to image distortion.

2.2. Performance metrics of watermarking schemes used for tracing

A watermarking scheme for tracing purposes is usually evaluated based on two critical performance metrics: (i) imperceptibility, i.e. watermarks should be invisible, and (ii) robustness to incidental distortion, such as signal processing, and attacks intended to remove the watermarks. For high imperceptibility, a watermark of small magnitude is inserted, in order not to interfere with the original image. On the other hand, a watermark of low magnitude is susceptible to noise and can be easily eliminated, making it hard to satisfy the robustness requirement. Hence, it is challenging to achieve both high imperceptibility and high robustness. We now describe the imperceptibility and robustness measures used in this paper.

2.2.1. Imperceptibility measures

Peak Signal to Noise Ratio (PSNR) is a widely used measure of fidelity (similarity between the original and the distorted image), since it is easy to compute and analytically tractable. For a b-bit image, PSNR is defined as

$$\mathrm{PSNR} = 10 \log\left(\frac{\sum_k (2^b - 1)^2}{\sum_k (x_k - x'_k)^2}\right), \tag{1}$$

where x_k and x'_k are the values of the kth pixel in original image and watermarked image, respectively. However, it is known that PSNR does not consider human visual sensitivities. In order to better evaluate image quality using objective measures, new image quality indices, Q Index (QI) and Moran Coefficient (MC) based indices, have been proposed in [9] and [10], respectively. The Q Index (QI) is defined as [9]

$$Q = \frac{4\sigma_{xx'} \sum_k x_k \sum_k x'_k}{(\sigma_x^2 + \sigma_{x'}^2)\left[\left(\sum_k x_k\right)^2 + \left(\sum_k x'_k\right)^2\right]}, \tag{2}$$

where σ_x and σ'_x are variances of x and x', respectively, and $\sigma_{xx'}$ is the cross correlation of x and x'. The QI is a product of correlation, luminance similarity and contrast, and has been shown to conform with human observer inspection very well. MC based indices are proposed in [11], based on the observation that MC can be an index of image smoothness and sharpness. Chen et al. [10] used Mean Moran Error (MME) and Mean Squared Moran Error (MSME) as quality indices, and more recently, used the Kolmogorov–Smirnov test to determine whether the discrepancy between original and altered images is due to chance, by studying Moran statistics of the images [11]. Though QI and MC based indices capture the local statistics and yield better compliance with human evaluation than PSNR, they do not lead to analytically tractable modes. We use PSNR in our analytical study due to its analytical tractability, but use all three measures: PSNR, QI and MSME, in our simulation study to evaluate the performance of our approach.[2]

We note that in addition to objective measures, visual inspection of an image by medical experts is another approach to assess image quality. However, the subjective evaluation varies from one expert to another and is time-consuming. In this paper, we will use objective measurements only. Apart from image quality, we also need to evaluate the robustness of watermarks.

2.2.2. Robustness measure

Correlation is a measure of statistical similarity of sequences. To evaluate the robustness against certain distortion or an attack, a watermarked image is first subjected to that distortion or attack and a test watermark is then extracted from the altered image. Correlation between the test watermark and a reference watermark is then calculated. Normalized correlation, denoted as *sim* [7],

[2] MME can be positive (smoother) and negative (sharper), it is not easy to compare two manipulated images that yield the same absolute value of MME, but opposite signs.

is used in this paper and is defined as

$$\mathrm{sim}(w, w^*) = \frac{w \cdot w^*}{\sqrt{w^* \cdot w^*}}, \tag{3}$$

where w and w^* are the reference and extracted watermark sequences, respectively.

2.3. Candidate watermarking techniques for tracing purposes

Cox et al. [6] provided a comprehensive study on watermarking techniques. There are two major classes of watermarking techniques: Spread Spectrum (SS) [7] watermarking and Quantization Index Modulation (QIM) [12]. QIM watermarking has been proven to allow embedding and reliably decoding more information bits, i.e. higher *watermarking capacity*, than SS watermarking when original images are absent at the detector [12]. However, the capacities of SS and QIM are identical when original images are available. For tracing purposes, the availability of original images is a reasonable assumption [13]. At the same time, it is known [14] that SS is more robust than QIM under the distortions that attenuate or remove the watermarks. Since tracing requires high robustness, we focus on the SS-based watermarking schemes.

2.3.1. Basic spread spectrum (SS) watermarking

Secure SS watermarking proposed in [7] still remains the most widely used watermarking scheme [14]. Cox et al. [7] viewed the channel between a watermark embedder and a detector as a communication channel, and extended the idea of spread spectrum communication to watermarking: a watermark, as a narrow band signal, is transmitted in a wide band signal (cover image) such that the signal energy presented in any single frequency is imperceptible. They discovered that watermarks should be embedded into perceptually dominant spectral components of the cover, making them imperceptible and robust to distortions. Furthermore, they chose watermarks that are sequences of independent zero mean unit variance Gaussian random variables, and thus orthogonal to each other, in order to avoid *false positives*, which is the probability of falsely detecting the existence of an absent watermark.

Watermarks of length L are embedded into the L largest (in magnitude) Discrete Cosine Transform (DCT) coefficients excluding the DC component. Let w_j denote the jth component of a watermark for $j=1,\ldots,L$, and X_j, X'_j denote the coefficients of the jth frequency of the original and the watermarked images, respectively. X'_j can be generated from w_j and X_j using Additive Embedding (AE) or Multiplicative Embedding (ME) as shown below

$$X'_j = X_j + \beta w_j, \quad w_j \sim N(0,1),\ j = 1, \ldots, L, \tag{4}$$

$$X'_j = X_j(1 + \beta w_j),\ j = 1, \ldots, L, \tag{5}$$

where β is a scaling factor. The smaller the β, the less distorted the watermarked images will be. The AE in (4) is image independent, and hence is suitable for the cases where watermarks need to be forcibly embedded regardless of the magnitudes of X_js. The ME in (5) better exploits the embedding capacity of frequency components and is suitable when the magnitudes of X_js vary widely [7].

The basic SS scheme does not take into account the human perceptual sensitivity as a factor to enhance the embedding capacity of an image. We now describe the image adaptive schemes that incorporate human perceptual models into watermark embedding.

2.3.2. Image adaptive schemes

Podilchuk et al. [13] introduced human visual models into basic SS to improve the robustness of SS without introducing perceptual degradation. In the Image Adaptive (IA) schemes [13], the strength and embedding positions of watermarks are adaptively chosen based on the perceptual sensitivity of the cover image in terms of Just Noticeable Differences (JND), which is image dependent.[3] There are two types of IA schemes: IA-DCT and IA-W. The IA-DCT is based on 8×8 block DCT transform and employs Watson's DCT sensitivity model [15]. The IA-W is based on wavelet transform and uses Watson's wavelet visual model [16]. Let JND_j be the calculated JND values. The embedding in IA schemes is described as:

$$X'_j = \begin{cases} X_j + \beta \cdot \mathrm{JND}_j \cdot w_j, & \text{if } X_j > \mathrm{JND}_j \\ X_j, & \text{otherwise.} \end{cases} \tag{6}$$

Note that instead of fixed length watermarks used in SS, the IA schemes explore the potential embedding capacity of every frequency component in the host image and thus in general, have a longer watermark sequence and hence more resilient to various processing procedures than SS [13].

3. Problem statement

The problem we study in this paper is tracing the source of an unauthorized release of medical images to enhance patients' privacy. On detecting a leaked image, we employ watermarking techniques to enable tracing back to the entity who initiates the illegal distributed of image intentionally or accidentally. We first identify the special requirements on watermarking techniques used for tracing medical images. Since group communication is common in a clinical environment for high quality health care, we then describe the challenge of tracing medical images in a group communication mode. Finally, we present the explicit formulation of our problem.

[3] JND is a concept in visual modeling defining a level of distortion that is perceptual in 50% of experimental trials [6] and is used to derive perceptually based quantizers in compression applications [13].

3.1. Requirements on using watermarking for tracing medical images

3.1.1. High image fidelity requirement

In medical imaging, distortion introduced by watermarks or image compression should be kept to a minimum so that diagnostic value is not compromised. The high fidelity requirement renders watermarking in medical images a harder task than multimedia applications where a much higher perceptual distortion is tolerated. Based on studies involving suitability of lossy compression for medical images, a minimum PSNR of 40–50 db is required [17], while 20–30 db PSNR is acceptable for multimedia applications. More recent studies in [11,18] show that no perceptual degradation for diagnosis purpose is found in Computerized Tomography (CT) body images with JPEG compression ratio 10:1. Hence, we use JPEG 10:1 compressed images as the baseline to compare with watermarked images in our simulation study. The stringent requirement of high PSNR limits the resilience of watermarks to various attacks.

3.1.2. Robustness to image processing

More importantly, watermarks should survive the standard image processing like low pass filter (LPF) and high pass filter (HPF) that improve diagnostic quality of medical images. A LPF removes noise and hence improves visual quality. A HPF functioning as an edge detector enhances the information content. LPF and HPF introduce interference into images, which can destroy watermarks, and thus are interpreted as attacks on watermark. The watermarks must also survive other common image processing operations such as: JPEG compression, cropping, and intentional attacks, such as averaging attack, where several watermarked copies are averaged in order to remove or attenuate a watermark.

In summary, watermarked medical images should retain high fidelity for diagnostic accuracy, and display robustness against filtering operations, in addition to conventional attacks [7], for tracing purposes.

3.2. Challenge on tracing in a multicast communication

Group communication is inherent in most of the clinical treatments. Let us consider the following treatment scenario, as shown in Fig. 2. A patient visits his/her primary doctor. If the condition is complicated, the doctor forms a team of physicians with different specialties to collaborate on the patient's case. The primary doctor retrieves the patient's related Electronic Health Record (EHR), specifically images, from a database, and delivers the EHR to other physicians for review. Each physician writes his/her comments and broadcasts them to the entire group. Upon the receipt of all the reports, the primary doctor will decide a treatment for the patient.

In the above scenario where there are multiple senders and multiple recipients, many-to-many multicast communication mode can be employed to transmit images, and hence save network bandwidth consumption and senders' computation power.

However, tracing medical images while retaining the efficiency of multicast communication is challenging to attain. In a multicast communication, multiple recipients receive an *identical* copy of an image. At the same time, a unique code for each user has to be embedded into the image to enable tracing images back to the source of leakage. Hence there will be *distinct* fingerprinted copies for different users, and the problem of multicasting different fingerprinted images is a challenge. This issue is addressed in Section 4.

The only relevant study on multicast of fingerprinted copies for tracing purposes has been in the watermarked

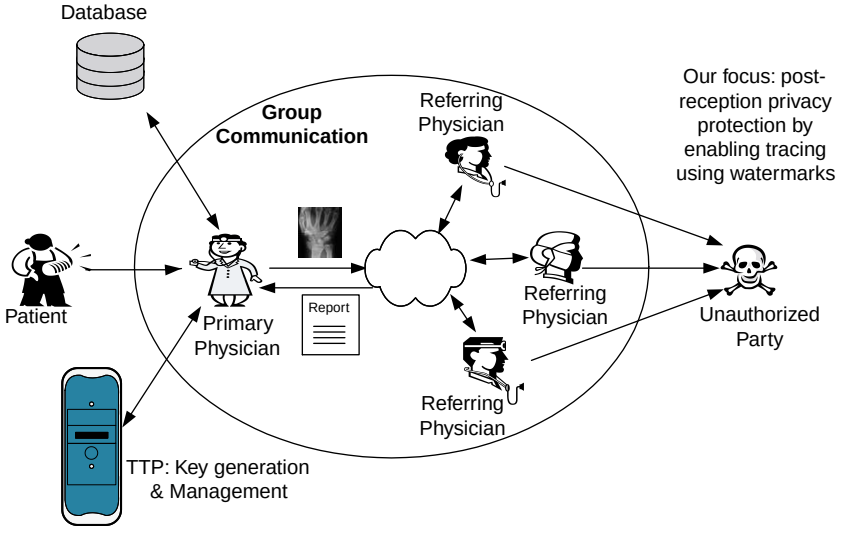

Fig. 2. A group communication scenario in a clinical environment and our study focus.

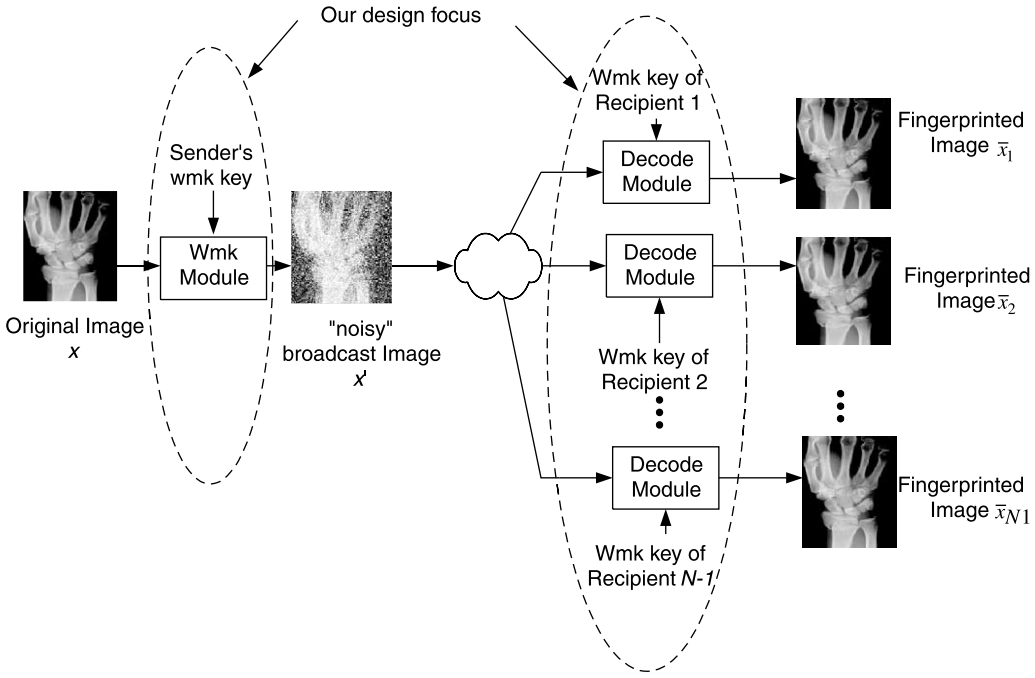

Fig. 3. A tracing model in a multicast environment with N users.

video [19–23]. Compared to videos, static images have less capacity for watermark insertion and hence it is harder to retain watermarks in images after processing and intentional distortion.

3.3. Problem of tracing medical image in a multicast environment

To ensure patient's privacy protection after the reception of an image in a group communication, as illustrated in Fig. 2, we study the problem of tracing medical images in a multicast environment, to which we refer as the *tracing problem*. We assume that each user in the group is assigned a *watermark key* for watermarks embedding and decoding. Specifically, we address the following problem: given an original image x, how to design a watermarking module with the sender's watermark key, and a decode module with each recipient's watermark key, so that a single copy of the image, x', can be broadcast to the rest of the group, and the decoded copy $\bar{x}_i$ carries a unique fingerprint assigned to user i, who is the intended recipient. Fig. 3 shows the modules of our design focus in this paper.

3.3.1. Image requirements

As discussed in Section 3.1, a fingerprinted image $\bar{x}_i$ should satisfy the following design requirements:

- *Req1*: High image fidelity to meet diagnostic requirements;
- *Req2*: Robustness to filtering operations while withstanding conventional attacks [7].

3.3.2. Efficiency requirements

To minimize the operational overhead incurred by the tracing model, we aim at a lightweight model that provides the following:

- *Req3*: Scalability in senders' and recipients' storage;
- *Req4*: Low cost in key update communication when a membership change occurs;
- *Req5*: Computational efficiency in both the watermarking and decoding modules.

We assume that there is a Trusted Third Party (TTP) who is in charge of the watermark key generation, distribution, and update, and tracing on detection of a leaked copy. We also assume that the TTP can access the original image when necessary.

4. An algorithm for tracing medical images in a group communication environment

In this section, we will first present our solution to the tracing problem in a clinical group communication environment as an explicit algorithm. For the algorithm to be employed in practice, we will address system parameter selection and system implementation issues. We also propose a computationally efficient version of our algorithm.

4.1. Algorithm description

To enable tracing in a multicast environment, the main idea of our approach is to broadcast a 'noisy' image with

Table 1
Notation

N	Number of users in the group
L	Length of a watermark sequence
$U \times V$	Size of an image
i	Index of users, $i = 1,\ldots,N$
j	Index of a watermark sequence, $j = 1,\ldots,L$
s, r	Generic index for a sender and a recipient, respectively
w_i	The watermark vector assigned to user i, $w_i = \{w_{i,1}, w_{i,2},\ldots,w_{i,L},\}$, each component $w_{i,j}$ is an independent normal distribution according to $N(0, \sigma_w^2)$; $\sigma_w^2 = 1$ unless otherwise mentioned
ν	An additional Gaussian white noise vector $N(0, \sigma_\nu^2)$, where σ_ν^2 is the minimum variance that ensures an initial watermarked copy from the sender is unusable
Y_i	Watermark key held by user i, $Y_i = \sum_{k=1}^{N} w_k + \nu - w_i$
x	Coefficients of an original image in spatial domain
X	Coefficients of an original image in watermarking domain w log, x and X are considered as the vectors of length $U \times V$
x'	Coefficients of a broadcast image from a sender
$\bar{x}$	Coefficients of a fingerprinted image for a user
β	Watermarking scaling factor
α	Watermarking weight vector $\alpha = \beta \cdot \text{JND}$, where $\alpha = \{\alpha_1,\ldots,\alpha_2,\ldots,\alpha_L\}$
T	Orthonormal transform, such as DCT, wavelet
θ	Threshold on correlation for watermark detection
δ	Minimum PSNR difference between a usable and an unusable images
Δ	Minimum PSNR requirement to maintain diagnostic image quality
C	Set of colluding members, and its cardinality is denoted as c, $c = \lvert C \rvert$

a strong watermark that prevents the image from being used directly to diagnose a patient. To achieve diagnostic value of the image, a recipient has to eliminate part of the watermark using a pre-assigned watermark key. This denoising process ensures that the fingerprint of the recipient will be imprinted into the image. Therefore, the usable image obtained by each recipient is a unique fingerprinted copy.

For clarity of presentation, we first list the notations used in Table 1.

Our algorithm for tracing in a multicast environment is described as follows. Box 1.

4.2. Algorithm properties

Our multicast fingerprinting algorithm has the following advantages.

4.2.1. Many-to-many multicast

Our algorithm enables fingerprinting in many-to-many, specifically, any-to-many multicast mode, by employing *only one* watermark key per user per group. This is a major advantage of the proposed fingerprinting model, over models proposed in [19–23], which are suitable for one-to-many scenarios. Extending those schemes to a

Box 1. Tracing algorithm in group communication

Initialization: The TTP generates a watermark sequence w_i for user i, and calculates the watermark key as $Y_i = \sum_{k=1,k\neq i}^{N} w_k + \nu$, and securely distributes Y_i to user i.*Watermarking module*: To send an image x to the entire group, sender s, who holds watermark key Y_s

1. Transforms the image to watermark domain as $X = T(x)$, where T denotes transform function.
2. Adds its scaled watermark key onto the image $X' = X + \alpha Y_s$, where α is watermark scaling vector.
3. Inversely transforms $x' = T^{-1}(X + \alpha Y_s)$ and broadcasts x' and α.

Decode module: On the reception of x' and α, recipient r, who holds $Y_r = \sum_{k=1,k\neq r}^{N} w_k + \nu$

1. Performs the transformation $X' = T(x')$.
2. Subtracts its weighted watermark key from x' to obtain $X + \alpha(Y_s - Y_r) = X + \alpha(w_r - w_s)$.
3. Performs the inverse transform and obtains the fingerprinted copy as $\bar{x}_r = x + \alpha T^{-1}(w_r - w_s)$.

Tracing: The TTP extracts a watermark from the test image, and computes correlations between the candidate watermark w_i and the extracted watermark in the watermark domain, if the correlation is larger than a threshold θ, the TTP concludes that watermark w_i is present in the test image.

many-to-many scenario, where there are N senders, increases the key storage requirement for each user by N-fold. Any-to-many is the desired communication mode for referral group applications.

4.2.2. Recipient and sender identification

The fingerprinted copy $\bar{x}_r = x + \alpha T^{-1}(w_r - w_s)$ bears not only the signature of the recipient w_r, but also that of the sender w_s. The TTP can identify the original sender as well as the recipient by calculating the correlation between the test copy and all candidate fingerprints. If the absolute values of both the largest positive and the smallest negative correlations are greater than the threshold θ, the positive indicates the recipient, and the negative correlation indicates the sender.

4.2.3. No image quality degradation due to relay

In case group members did not receive the broadcast images, the sender can either resend the images himself or designate a member m to relay the images to them. If relaying takes place, and m has already decoded its copy by subtracting Y_m from the broadcast copy to obtain $X + \alpha(Y_s - Y_m) = X + \alpha(w_m - w_s)$, m, now as a sender, can add Y_m back to the decoded copy to get $X + \alpha(w_m - w_s + Y_m) = X + \alpha Y_s$ before forwarding the image. This is equivalent to forwarding the original received copy $X + \alpha Y_s$. In case of a transmission chain sequence, relay nodes will not leave their fingerprints on the image and only the watermarks of the original sender and the last recipient in the transmission remain in the image. Since at any given time, only two fingerprints are present in the decoded image, the quality of the image at recipient side is controllable.

4.2.4. Combination of watermarking and encryption

In the proposed scheme, a broadcast image from the sender is of low perceptual quality and not easily recognizable. Therefore, certain degree of image secrecy is provided as a side product of our watermarking algorithm, which potentially eliminates the necessity of separate encryption. The proposed fingerprinting scheme can be regarded as a combination of watermarking and encryption, the partial watermark removal by the recipient as the combination of fingerprinting and decryption, and the watermarking keys as encryption keys. Special care, however, has to be taken for member eviction when using watermarking only, for both tracing and confidentiality. Since there is no separate set of encryption keys used for access control, the only way to disable the evicted member from viewing the image is to renew all the watermark keys of the remaining members, which incurs update communication cost of $\mathcal{O}(N)$. We note that we can indeed add and manage a separate set of keys for encrypting multicast data that is watermarked, with additional overhead of $\mathcal{O}(\log N)$. However, we do not address it in this paper.

4.3. Selection of watermarking schemes

In order to employ our algorithm, we need to choose an underlying watermarking scheme and select an embedding rule. For scheme selection, we mainly consider the following two criteria: (i) *image adaptiveness*, which indicates the ability of watermarks to exploit image features such as local contrast and illuminance [6]; in order to better explore the image embedding capacity for high perceptual quality and robustness; (ii) image secrecy: the weight α cannot reveal information on original images since recipients need to know α to decode fingerprinted copies from a broadcast image in our algorithm.

4.3.1. Selection of schemes

We described the basic SS, IA-DCT and IA-W schemes in Section 2.3. The drawback of the basic SS is that it does not exploit a perceptual model for image adaptiveness. For IA schemes described in Eq. (6), the weights $\alpha_j = \beta \cdot \mathrm{JND}_j$ are used to shape watermarks for better perceptual quality with the same level of tamper resistance. For the IA-DCT scheme, which employs Watson's DCT visual model [15], each JND_j value is a function of the frequency coefficient X_j, and will potentially disclose the information on image x. We calculated the JND_js on several images and noticed that once JND_js are given, almost 80% of all frequency coefficients can be correctly derived. This observation rules out the possibility of using IA-DCT as the underlying watermarking scheme for image secrecy protection.

For IA-W, JND_j is a constant for a particular orientation and level [16], and hence reveals no information on the original image. However, $\alpha_j = \beta \cdot \mathrm{JND}_j$ does reveal 1-bit information on the original image. The principle for IA schemes is that a watermark is embedded only when the magnitude of a frequency coefficient is larger than the JND for that frequency, i.e. when $X_j > \mathrm{JND}_j$, and hence every component of α is either $\beta \cdot \mathrm{JND}_j$ or zero. One can infer whether X_j is larger or smaller than JND_j from α_j. Thus, in order to meet the criterion of image secrecy, we will employ the fixed position embedding instead.

We embed a watermark on a chosen position regardless of the magnitude. Hence, there is no need to transmit α_js but β only. The choice of the embedding position is a tradeoff between robustness and image fidelity. Although the fixed position embedding is not as image adaptive as the embedding strategy in the original IA schemes, there are advantages to its use: (i) there will be no information leakage due to the transmission of α_js, (ii) it facilitates efficient watermarking without computation of transform and its inverse as described in Section 4.5, and (iii) it helps to survive frequency selective processing, while the original IA schemes fail HPF [24]. Furthermore, we can compensate for the loss of fidelity when using fixed position embedding by adjusting the scaling factor β. We also notice that although the embedding position is publicly known, it is difficult to remove the watermark without significantly

corrupting image quality, since the precise magnitude of each inserted watermark is known only to the TTP [7].

Thus, we conclude that among the basic SS, IA-DCT, and IA-W, the weight vector α used in IA-W provides the best trade-off between the image adaptiveness and secrecy, and hence we select IA-W as the underlying watermarking scheme. Next, we need to choose between additive and multiplicative embedding.

4.3.2. Choice of embedding rule

Using additive embedding, a broadcast image is $x' = x + T^{-1}(\alpha Y_s)$. As we discussed above, fixed position embedding α in IA-W reveals zero knowledge on the original image. However, if multiplicative embedding is used, the broadcast image becomes $x' = T^{-1}(X(1+\alpha Y_s)) = x + T^{-1}(aX \cdot Y_s)$, and the weight vector becomes αX, which is a scaled version of the original image in the transform domain. By broadcasting the weight vector αX, the image secrecy is completely compromised. Hence only additive embedding can be used in our algorithm.

We now show how to choose system parameters: the watermarking scaling factor β and the additional noise sequence ν.

4.4. Specification of system parameters

In this section, we address the following two issues: (i) specification of the watermark scaling factor β to ensure high PSNR of fingerprinted images, and (ii) quantization of additional noise sequence ν to guarantee low PSNR of broadcast images, so that recipients will be forced to decode the images and leave fingerprints in the process.

4.4.1. Specification of watermark scaling factor β

The parameter β is computed from the requirement that the PSNR of a fingerprinted image must be at least Δ to maintain diagnostic value of the image

$$E\left[10\log\left(\frac{U\times V\times(2^b-1)^2}{\sum_k(\bar{x}_k-x_k)^2}\right)\right] = 10\log\left(\frac{U\times V\times(2^b-1)^2}{\sum_j \alpha_j^2 E(w_{r,j}-w_{s,j})^2}\right) = 10\log\left(\frac{U\times V\times(2^b-1)^2}{\sum_j \beta^2 \mathrm{JND}_j^2(2\sigma_w^2)}\right) \geq \Delta.$$

Therefore,

$$\beta^2 \leq \frac{U\times V\times(2^b-1)^2}{10^{\Delta/10}\sum_{j=1}^{L} 2\sigma_w^2 \mathrm{JND}_j^2}. \tag{7}$$

4.4.2. Quantization of additional noise sequence ν

As additional noise, ν is used to keep the PSNR of the transmitted image low. The added noise ν is a sequence of independent zero mean Gaussian variables with variance σ_ν^2. Given the expected minimum PSNR difference between a transmitted image x' and a fingerprinted image $\bar{x}$ as δ, we can determine the minimum value of σ_ν^2 required to achieve the minimum PSNR difference as follows

$$\begin{aligned}\delta &\leq E\left[10\log\left(\frac{U\times V\times(2^b-1)^2}{\sum_k(\bar{x}_k-x_k)^2}\right) - 10\log\left(\frac{U\times V\times(2^b-1)^2}{\sum_k(x'_k-x_k)^2}\right)\right]\\ &= 10\log_{10}E\left(\frac{\left(\sum_k x'_k-x_k\right)^2)}{\sum_k(\bar{x}_k-x_k)^2}\right)\\ &= 10\log_{10}\left(\frac{\sum_j \alpha_j^2 E\left(\sum_{i=1}^{N} w_{i,j}-w_{s,j}\right)^2}{\sum_j \alpha_j^2 E(w_{r,j}-w_{s,j})^2}\right),\end{aligned} \tag{8}$$

$$\delta = 10\log_{10}\left(\frac{(N-1)\sigma_w^2+\sigma_\nu^2}{2\sigma_w^2}\right). \tag{9}$$

Eq. (8) holds due to the fact that an orthonormal transformation preserves the total variance [25]. From (9), we have:

$$\sigma_\nu^2 \geq 10^{\delta/10}2\sigma_w^2 - (N-1)\sigma_w^2. \tag{10}$$

Note that the solved σ_ν^2 is the minimum variance of ν to guarantee the PSNR difference between a transmitted image and a decoded image. If the value calculated from the right hand side of (10) is less than zero, it implies that even without adding ν, the transmitted image $x' = x + \alpha T^{-1}(Y_s)$ is not directly viewable.

4.5. A computationally efficient transform domain watermarking

In this subsection, we present a computationally efficient transform domain watermarking scheme that is applicable when all images have fixed size, or the size of the image is known before the current watermark keys are generated.

A broadcast image from sender s is $x' = x + T^{-1}(\alpha Y_s), = x + \beta \cdot T^{-1}(\mathrm{JND}\cdot Y_s)$. If the inverse transform of the scaled watermark key, $T^{-1}(\mathrm{JND}\cdot Y_i)$, is pre-computed for each user i, the computation of both the transform and the inverse transform can be saved by the sender and receivers. The sender only needs to decide β based on (7), add $T^{-1}(\mathrm{JND}\cdot Y_s)$ to the image, and broadcasts both x' and β. Recipients subtract their inverse transformed watermark keys $\beta T^{-1}(\mathrm{JND}\cdot Y_r)$ from x' directly, and there is no computation of transform involved. This approach needs only $U\times V$ additions and $U\times V$ subtractions, compared to other transform-domain watermarking that needs at least additional $U\times V\times \log_2 U$ multiplications by each user. To enable this pre-computation of inverse transform approach, we need to adopt fixed position embedding, and know the image size in advance. In practice, medical images of different modality come in certain sizes, thus, we can pre-compute the inverse transform of the watermark keys

for several common sizes. If none of these sizes fits the image to be watermarked, the original algorithm can still be used and the transform can be computed on-the-fly.

4.6. System implementation issues

In order to implement our algorithm properly, we elaborate on two system issues: (i) bit representation of broadcast images with noise added, and (ii) watermark key management. We assume that all the system issues are transparent to users and will be automatically processed.

4.6.1. Bit representation of broadcast images

We observed that individual pixel values in broadcast images will no longer be within the range [0,255] after being inserted with watermark keys, even though the original images are 8-bit images. To avoid an overflow error, we use more bits to represent pixel values in a broadcast image, and ensure accurate fingerprinting and correct watermark detection.

Let us consider watermarking of an 8-bit image x. We transform the image into the frequency domain and add the watermark key that consists of fingerprints and added noise ν. Upon performing the inverse transform on the watermarked image, many pixel values are above the 8-bit maximum of 255 and below the minimum of zero, the bit overflow occurs due to the intentional image corruption contributed by ν. Clipping the pixel values between 0 and 255 will lead to loss of information and hence improper watermark detection.

We estimate the range of the watermarked image when using IA-W scheme as follows. From (9), it is suggested that a broadcast image X' carries a zero-mean Gaussian watermark with variance $2\sigma_j^2\sigma_w^2 10^{\delta/10}$, which is upper bounded, based on (7) by

$$\sigma_{ub}^2 = 2\frac{U\times V\times(2^b-1)^2}{10^{\Delta/10}\sum_{j=1}^{L}2\mathrm{JND}_j^2}\max(\mathrm{JND}_j^2)10^{\delta/10},$$

when $\sigma_w^2 = 1$. Therefore, the distortion caused by the watermark in X' is less than that caused by a zero mean Identical Independent Distributed (IID) Gaussian vector w_{ub} with variance σ_{ub}^2. A zero-mean IID Gaussian vector remains as an IID Gaussian with the same variance after an orthonormal transform [25]. Therefore, $T^{-1}(X+w_{ub}) = x+W_{ub}$, where W_{ub} is also a zero mean Gaussian with the same variance σ_{ub}^2. The probability of a normally distributed random variable exceeding its mean by more than three standard deviations is less than 0.003. Hence, with confidence level 99:7%, the range of x' in spatial domain is

$$\left[-\frac{U\times V\times(2^b-1)^2\max(\mathrm{JND}_j^2)10^{(\delta-\Delta)/10}}{\sum_{j=1}^{L}\mathrm{JND}_j^2}, 2^b-1 + \frac{U\times V\times(2^b-1)^2\max(\mathrm{JND}_j^2)10^{(\delta-\Delta)/10}}{\sum_{j=1}^{L}\mathrm{JND}_j^2}\right].$$

Note that the range of x' depends on the length of inserted watermark and image dependent JND values, and hence will be different for different images. For all the 8-bit images tested in our simulation, the largest calculated range is $[-2^{12}, 2^{12}]$, and simulation results confirm that the range is adequate to avoid overflow errors. For easy implementation, we use multiple 8-bits (bytes), 16-bits, to represent pixel values in broadcast images. The 16-bits are used only for transmission of the broadcast image. Upon reception of the images, the recipients remove the watermark using their keys and display the fingerprinted 8-bit image.

4.6.2. Watermark key management

Watermark keys are employed to fingerprint the images for tracing purposes, and hence their management is of special importance to ensure the correctness of our algorithm. In this section, we address watermark key management for member join, member deletion, and multiple groups. Periodic key update is also used to ensure the secrecy of watermark keys.

4.6.2.1. Handling member join. To add new members, no change of the watermark keys held by current members is needed, due to the presence of additional noise ν. For the newly joined member, a new fingerprint sequence, w_{N+1}, is generated. The watermark key assigned to the new member will be

$$Y_{N+1} = \sum_{k=1}^{N} w_k + \nu - w_{N+1} = \sum_{k=1,k\neq N+1}^{N+1} w_k + \nu',$$

where $\nu' = \nu - w_{N+1}$. The updated watermark key for current member i should be

$$Y_i' = \sum_{k=1,k\neq i}^{N+1} w_k + \nu' = \sum_{k=1,k\neq i}^{N} w_k + w_{N+1} + \nu - w_{N+1} = \sum_{k=1,k\neq i}^{N} w_k + \nu = Y_i,$$

which is exactly the current watermark key held by user i. Therefore, no renewal of watermark keys of current members is necessary while adding new users and thus communication and computation resources are saved. Based on (9), more than 100 members can be added when PSNR difference δ is at least 20 db and the original group size is less than 50.

4.6.2.2. Handling member deletion. To prevent deleted members from accessing patients' records and further discussion among the group, the group session key, which is a group encryption/decryption key to ensure the confidentiality of the communication, needs to be updated.[4] However, the revocation of a member does not affect

[4] The management of multicast group session key is out of scope of this paper. Interested readers are referred to [26].

the watermark keys for tracing purposes, therefore, the watermarking keys of remaining members will not be changed.

4.6.2.3. Handling multiple groups. When a medical professional m belongs to more than one group, and each group takes care of a different patient, the TTP assigns different watermarks to m for different groups to prevent m from forwarding an image to unintended groups without being traced. Let member m belong to two groups G_1 and G_2 and have two watermark keys

$$Y_{m1} = \sum_{i \in G_1} w_i - w_{m1} + \nu_1,$$

used in group G_1 and

$$Y_{m2} = \sum_{j \in G_2} w_j - w_{m2} + \nu_1,$$

used in group G_2. If m obtains an image from group G_1 and decodes as $X + w_{m1} - w_s$, it cannot simply add Y_{m2} and forwards it to group G_2 without being traced. Any recipient in group G_2 will decrypt the image as $X + w_{m1} - w_s + Y_{m2} - Y_r = X + w_{m1} - w_s + w_r - w_{m2}$, where both m's fingerprints remain. However, if $w_{m1} = w_{m2}$, the fingerprint of m vanishes in the decoded image by recipients in group G_2, and member m cannot be traced. Therefore, different watermark keys have to be generated for a single member to use in different groups.

4.6.2.4. Periodic update of watermark keys. Watermark keys need to be renewed periodically to prevent key search attack. Between key updates, the same watermark for each user, per group, is embedded into multiple images. By collecting all the fingerprinted images, one member may mount a watermark estimation attack. Assume that there is a sequence of images where each frequency component of images satisfies a Gaussian distribution $N(\mu,\sigma^2)$ across all images, with μ being unknown. Assuming member m has collected t images $X_{l,m}$ for $l = 1,\ldots,t$ from the same sender s marked with the same fingerprint w_m i.e. $X_{l,m} = X_l + \alpha(w_m - w_s)$, the member can derive the Maximum Likelihood (ML) estimation as

$$\hat{w}_m + \frac{1}{\alpha t}\sum_{l=1}^{t} X_{l,m} - \frac{\mu}{\alpha} + w_s,$$

and the mean square estimation error is σ^2/t, which decreases as t increases. Since μ and w_s are unknown, the member cannot have a good estimation of w_m. However, if member m uses

$$\hat{w}_m + \mu - m_s = \frac{1}{\alpha t}\sum_{l=1}^{t} X_{l,m},$$

as the estimation, the detector yields the same correlation as using $\hat{w}_m$, since a linear correlator is invariant to a constant shift. The invariance is shown as $E[(w_m - (\hat{w}_m + \mu - w_s)) \cdot w_m] = E[(w_m - \hat{w}_m) \cdot w_m] + \mu E(w_m) + E(w_m)\ E(w_s) = E[(w_m - \hat{w}_m) \cdot w_m]$, and the last two items are zero mean due to the fact that all the watermarks are zero and independent of each other. Therefore, if a large number of images embedded with the same watermark are available, a good estimation of a watermark used to defeat the correlator can be achieved.

We note that the TTP needs to keep all the seeds that generate watermarks—current and previous—of all the users in the system, in order to perform tracing once a leaked copy is detected. The storage requirement for TTP is high. Even if one assumes that key storage is not a concern, there is the problem of searching through all the candidate watermarks in order to locate the one who distributed the image illegally. We leave this issue as a topic in a forthcoming paper and it is not discussed here.

4.7. Estimation attacks by collusion

In addition to defending against geometric distortion [6], such as cropping, and signal processing techniques that destroy the watermark, we are interested in collusion attacks, due to the fact that wired and wireless networks are vulnerable to collusion (illegal collaboration) among multiple users. Collusion occurs when a set of group members collaborate and compare their different watermarked keys or fingerprinted images, to create a copy of the image that cannot be traced back to any of the colluders [26]. Such an illegally obtained or a pirated copy can be generated either by first successfully estimating and then subtracting the watermark from the fingerprinted copy, or by estimation of original images. We will present both the watermark estimation and the estimation of original images in this section.

4.7.1. Watermark estimation from watermark keys

We now present the ML estimation of a watermark, w_i, when colluders pool their watermark keys together, i.e. they have Y_i for $i \in C$. The estimation problem can be converted to the ML estimation of S, for $w_i = S - Y_i$. w log, let us assume that users $i = 1,2,\ldots,c$ collude. Then, for watermark sequence index $j = 1,\ldots,L$

$$Y_{i,j} = S_j - w_{i,j} \quad \text{for } i = 1, \ldots, c.$$

First we observe that for each j, there are c equations and $c+1$ unknowns in the above equation arrays, even all the $N-1$ recipients collude, they cannot derive S analytically. However, if the size of collusion c is large enough, colluders can obtain a fairly accurate estimation of S. The problem of estimating S can be interpreted as estimating the non-random parameter S_j for all j, given c observations $Y_{i,j}$ for $i = 1,\ldots,c$, which are S_j but corrupted by independent additive white Gaussian noise with variance $\sigma_w^2 = 1$.

Given that S_j is fixed and $w_{i,j}$ is normally distributed according to $N(0, 1)$, $Y_{i,j}$ is Gaussian distributed with mean S_j and variance 1. For the simplicity of notation, we drop subscript j in the following whenever there is no confusion. The conditional probability density function of Gaussian vector Y_i for all $i \in C$ given S is [27]:

$$P_{y|s}(Y_1, \ldots, Y_c|S) = \prod_{i=1}^{C} P_{y|s}(Y_i|S) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\sum_{i=1}^{c} \frac{(Y_i - S)^2}{2}}. \tag{11}$$

By taking the derivative of (11) and setting it equal to zero, we obtain the ML estimation of S as $\hat{S} = 1/c \sum_{i \in c} Y_i$, which is the average of all Y_i for $i \in C$. This unbiased ML estimation by collusion is essentially an average attack.

We also note that $E[\hat{S} - S]^2 = 1/c$, and hence the variance achieves the lower bound of Cramer–Rao inequality [27] for non-random parameter estimation. Therefore, the ML estimation is an efficient estimation. Then the estimation of w_i for $i = 1,\ldots,c$ is:

$$\hat{w}_i = \hat{S} - Y_i = \frac{1}{c} \sum_{k \neq i, k \in C} Y_k + \left(\frac{1}{c} - 1\right) Y_i. \tag{12}$$

If one colluder, i, tries to remove the watermark using estimated $\hat{w}_i$, then the expected normalized correlation between the extracted watermark $[(w_i - \hat{w}_i)]$ and w_i is

$$1/L \sum_{j=1}^{L} E[(w_{i,j} - \hat{w}_{i,j}) w_{i,j}] = 1/c.$$

As the collusion size c increases, the correlation decreases and it becomes harder to detect the watermark reliably.

4.7.2. Image estimation by collusion

In addition to combining their watermark keys to mount watermark estimation, colluders can compare their fingerprinted images to obtain the estimated copy of the original image. Now we study the ML estimation of the original image X, given X_i for $i \in C$, where $X_i = X + \alpha(w_i - w_s)$.

Each X_i can be regarded as a copy of X corrupted by noise $\alpha(w_i - w_s)$, and hence, at each frequency, X_i for $i \in C$ is Gaussian distributed with mean X and variance $2\alpha^2$. Following a similar analysis as in the watermark estimation, the ML estimation of X is the average $\hat{X} = \sum_{k \in C} X_i/c = X + \alpha(\sum_{k \in C} w_k/c - w_s)$, and the mean square error of the estimation is $2\alpha^2/c$.

If members collude to remove a watermark from images by using watermark estimation, the resulting image will be the same as the one generated by image estimation, and is given by:

$$\begin{aligned}
\hat{X} &= X + \alpha(Y_s - Y_r) - \hat{w}_r \\
&= X + \alpha\left[(Y_s - Y_r) - \frac{1}{c}\left(\sum_{k \in C} Y_k - Y_r\right)\right] \\
&= X + \alpha\left[S - w_s - \frac{1}{c}\sum_{k \in C}(S - w_k)\right] \\
&= X + \alpha\left(\sum_{k \in C} \frac{w_k}{c} - w_s\right).
\end{aligned} \tag{13}$$

The result in (13) is predictable, since both estimations are essentially averaging attacks.

In Section 5, we will evaluate our proposed algorithm by simulation study.

5. Performance evaluation by simulation

To assess the feasibility of the proposed fingerprint scheme in medical group communication environment, we need to ensure that (i) a broadcast image from the sender is of poor perceptual quality and useless on its own, (ii) a fingerprinted copy obtained by each designated recipient has high quality for diagnostic purposes, (iii) the watermarks in all fingerprinted images are robust to various image processing, such as LPF, HPF, JPEG compression, cropping, and collusion attack.

5.1. Simulation setup

In our simulation study, we assume a group of $N = 7$ medical professionals. We examine three 490×490 fluoroscopic images, three 512×512 Computerized Tomography (CT) body images, three 512×512 Optical Tomography (OT) images, two 256×256 Magnetic Resonance (MR) images and $20{,}640 \times 480$ ultrasound images. Some of the images are downloaded from the Internet [28], and the images represent a good collection of different modalities and a mix of 8-, 12-bit images 16-bit images.

We use a four level IA Wavelet scheme [13] as the underlying watermarking scheme. We insert the watermark into low–high and high–low band of level two and four wavelet decompositions. The low–low band represents approximate information of the image and should be avoided to preserve visual quality. The coefficients of high–high bands are of small magnitude and hence of low embedding capacity, therefore, we do not use them for watermarking. The length of a watermark L depends on the image size, for example, $L = 8407$ for 640×480 images and $L = 34{,}816$ for 512×512 images. Watermarks are generated using uniform random number generators, and watermark keys for common image sizes are computed and transformed to the wavelet domain.

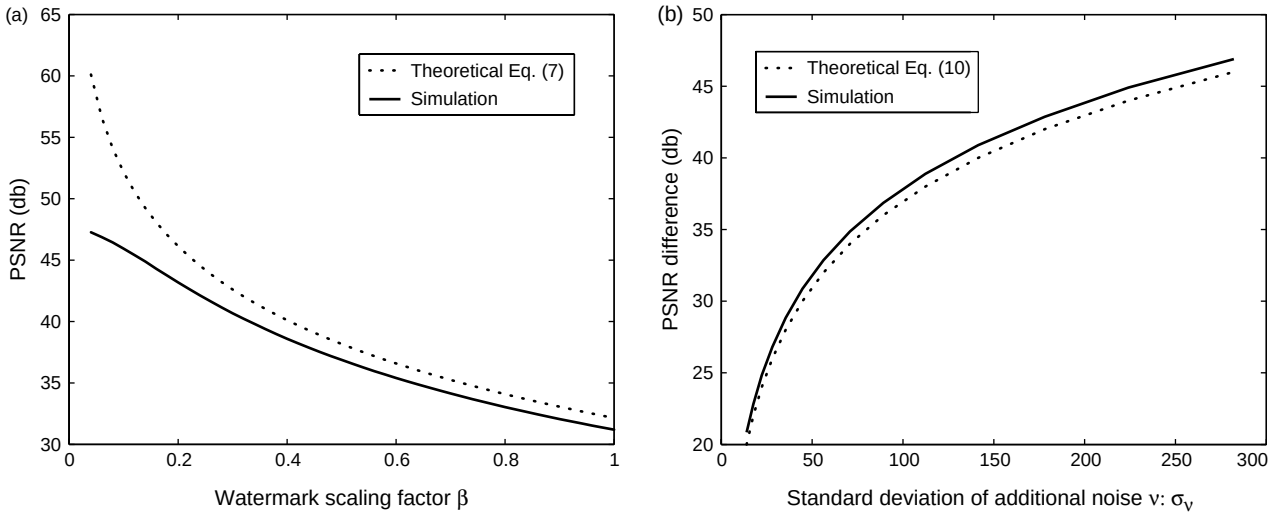

Fig. 4. Comparison of the theoretical values and simulation results for the watermark scaling factor β (a) and standard deviation of additional noise σ_ν (b).

5.2. Comparison of the theoretical and simulation values of system parameters

To utilize our watermarking scheme properly, we need to specify (i) the scaling factor β, to produce high perceptual quality images at the recipients' end, and (ii) the standard deviation of the additional noise ν, denoted as σ_ν, to ensure that broadcast images are of no diagnostic value. Fig. 4 illustrates the theoretical values of β in (7) and σ_ν in (10) compared with the simulation results for 8-bit ultrasound images. We notice that the theoretical values based on (7) and (10) provide a good guideline for choosing the values of β and σ_ν. Since (7) is an upper bound on β for the average value of PSNR, the resulting image may have slightly lower PSNR. If a higher PSNR is required, we can choose a smaller value of β than the computed one. Similarly, (10) only provides a lower bound on σ_ν, we can select larger σ_ν to decrease PSNR of broadcast images.

Fig. 5 presents two visual examples of the proposed watermark schemes. Fig. 5 (a1), (b1), and (c1) represent the original, the broadcast and the fingerprinted fluoroscopic images, and those ultrasound images are shown in (a2), (b2), and (c2), respectively. From Fig. 5, we note that the original and the watermarked images are indistinguishable, while the broadcast images are noisy and not usable directly. The broadcasted ultrasound image presents very little structural information, and no text information is revealed.

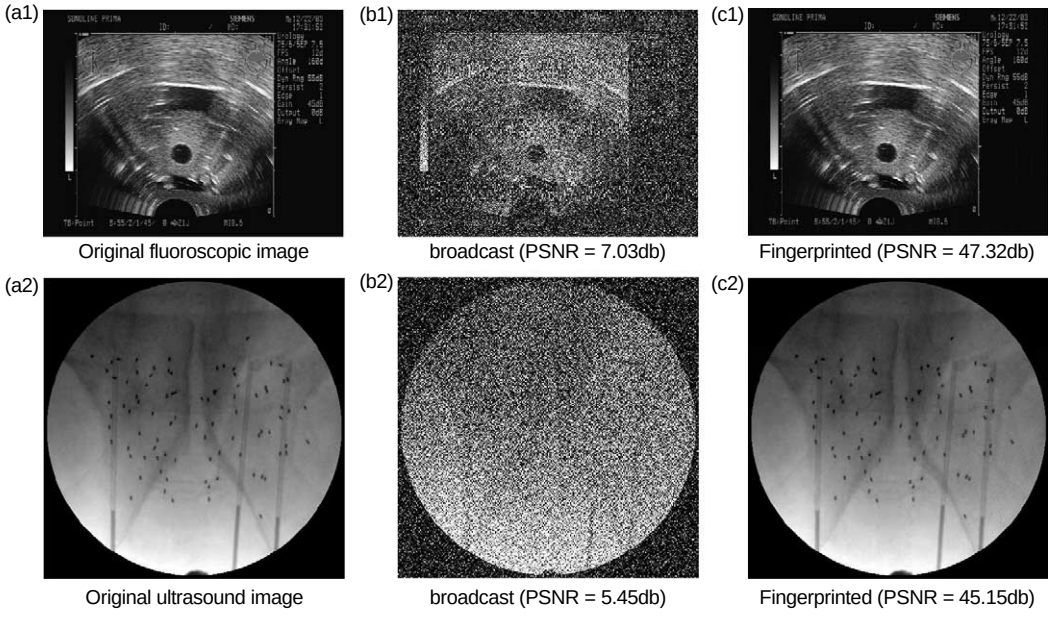

Fig. 5. Comparison of the original (a1), the broadcast (b1) and the fingerprinted (c1) fluoroscopic images and comparison of those ultrasound images (a2, b2, c2). There is no noticeable visual difference between the original (column one) and the fingerprinted (column three) images, and the broadcast images (column two) are blurred and hence useless.

Table 2
A comparison of image quality of JPEG images with compression ratio 10:1 and the fingerprinted images by a recipient in terms of PSNR, Q Index (QI) and Mean Moran Square Error (MSME)

Images	Processing	PSNR	QI	MSME
Abdomens	JPEG 10:1	34.32	0.8800	0.3735
	wmk $\beta=0.1$	54.62	0.9205	0.02557
Brain	JPEG 10:1	84.91	0.7278	1.6435
	wmk $\beta=0.8$	86.34	0.7821	0.01368
Orthopaedic	JPEG 10:1	73.94	0.7670	0.02425
	wmk $\beta=0.8$	86.22	0.8116	0.004312

The recipient is randomly selected from six receivers. Three 512×512 CT images, including a 8-bit abdomen image, a 16-bit brain and a 16-bit orthopaedic image, were tested. 'wmk' stands for watermarking.

5.3. Quality evaluation of fingerprinted images

Fingerprinted images obtained by recipients have to preserve high fidelity to ensure diagnostic accuracy. In [11,18], it is reported that no visual differences from original CT body images are found in JPEG images with 10:1 compression ratio. We use the 10:1 compressed JPEG images as baseline and compare them with our watermarked images in Table 2, in terms of three image quality indices: PSNR, Q Index (QI) [9], and Mean Squared Moran Error (MSME) [10].

The results in Table 2 demonstrate that our watermarked images with $\beta=0.1$ for the 8-bit image and $\beta=0.8$ for 16-bit images present better similarity with the original images than the JPEG 10:1 compressed images consistently, in terms of all three image quality indices. Note that the smaller MSME, the better fidelity.

In the following simulation study, we set $\beta=0.1$ for 8-bit images and $\beta=0.8$ for 16-bit images as in Table 2, to ensure that the perceptual degradation by fingerprints is less than that of 10:1 JPEG compression. We also set a theoretical $\delta=45$ db. Table 3 presents the PSNR, QI and MSME of a broadcast image and decoded images by all receivers for each modality.

From each row of Table 3, it shows that there is almost no quality difference across the fingerprinted copies obtained by different recipients, and all the fingerprinted images retain high fidelity with the original image.

5.4. Robustness tests against image processing and averaging attacks

Apart from meeting perceptual quality requirements on broadcast images and fingerprinted images, we also need to ensure that fingerprinted images are resilient to common processing operations relevant to medical images. We consider the following processing.

- LPF: image is operated upon by a 3×3 smoothing window. LPF is used for image smoothing and noise reduction.
- HPF: image is operated upon by a standard 3×3 high pass kernel. HPF is used for edge detection.
- JPEG compression: compressed images are written in JPEG format using the imwrite function in Matlab, with quality factor being 75.
- Cropping: the central quarter of the watermarked image is cropped to mimic a cropping attack. For the purpose of watermark detection under the cropping attack, the missing part is replaced from the original cover image.

Table 4 presents the robustness test results based on all 31 images obtained by one randomly selected recipient. We generate 20 different watermark keys for the recipient, and the average is presented. The percentage indicates the surviving watermarks. The threshold for sim defined in (3) is set at $\theta=5$, which is well above the similarity between two independent Gaussian random sequences [7]. No false negative is found.

Table 3
PSNR (db), QI, and MSME of broadcast images and fingerprinted images when $\beta=0.1$ for 8-bit images and $\beta=0.8$ for 16-bit images and $\delta=45$ db

Modality		Sender	*R*1	*R*2	*R*3	*R*4	*R*5	*R*6
490×490	PSNR	11.6471	52.0892	52.0954	52.1007	52.0826	52.0542	52.0739
fluoroscopic	QI	0.0606	0.9521	0.9535	0.9540	0.9529	0.9504	0.9511
(8-bit)	MSME	4.3882	0.00063	0.00064	0.00061	0.00062	0.00063	0.00061
512×512 CT	PSNR	11.6077	54.6486	54.6612	54.6558	54.6468	54.6623	54.6401
(8-bit)	QI	0.1339	0.9220	0.9264	0.9171	0.9219	0.9212	0.9227
	MSME	12.4491	0.00281	0.00285	0.00350	0.00242	0.00354	0.00218
512×512 OT	PSNR	11.5877	52.5543	52.5492	52.5364	52.5327	52.5579	52.5348
(8-bit)	QI	0.1077	0.9548	0.9576	0.9599	0.9609	0.9560	0.9643
	MSME	1.0819	0.00037	0.00039	0.00040	0.00038	0.00039	0.00039
256×256 MR	PSNR	41.7143	86.6365	86.7207	86.7048	86.7299	86.6904	86.8000
(16-bit)	QI	0.0139	0.9353	0.9346	0.9354	0.9347	0.9337	0.9360
	MSME	7.0342	0.1097	0.1085	0.1134	0.1093	0.1082	0.1051
640×480	PSNR	11.6519	50.9463	50.9627	50.9384	50.9594	50.9596	50.9322
ultrasound (8-	QI	0.1053	0.7675	0.7685	0.7680	0.7676	0.7682	0.7679
bit)	MSME	4.85870	0.001103	0.001104	0.001138	0.001124	0.001120	0.001092

For each modality, one test image is randomly chosen from available images. '*R*' denotes recipient.

Table 4
Robustness against different distortions

Modality	LPF (%)	HPF (%)	JPEG (75; %)	Cropping (%)
Fluoroscopic	100	100	100	100
CT	100	100	100	100
MR	100	100	100	100
OT	100	100	100	100
Ultrasound	100	100	100	100

Percentage of surviving watermarks is presented when $\beta = 0.1$ for 8-bit images and $\beta = 0.8$ for 16-bit images with PSNR > 50 db.

The results in Table 4 suggest that the watermarks display a high immunity to common signal and geometric processing procedures, without sacrificing the image quality: PSNR > 50 db (see Table 3).

As explained earlier, collusion resilience is of special importance in a multiuser environment. We have analyzed and shown in Section 4.7 that if several members collude, averaging is the maximum likelihood estimation of both watermark keys and original images. We now present the results of robustness tests against the averaging attack and its combination with other distortion in Table 5. We consider two scenarios: (i) there are two colluders, which is the smallest possible collusion size; (ii) there are five colluders, which is the largest collusion size where we can test for *false negative* in colluder detection, i.e. the probability of falsely detecting a honest user as a colluder.

From Column two of Table 5, we notice that the presented watermarking scheme is not susceptible to an averaging attack, and that *all* the colluding users can be correctly identified with zero false negatives. On the other hand, from columns three, four and five, it follows that the combination of an averaging attack and other distortion can be an effective way to remove watermarks. If the manipulated images, after averaging plus other distortions, are still useful, a robust watermarking scheme against the combined averaging attacks is desired. But this goal appears to be difficult to achieve with the same degree of imperceptibility as in the case of a simple averaging attack.

Table 5
Robustness against averaging attack of two and five colluders

Modality	Avg	LPF + Avg	HPF + Avg	JPEG (75) + Crop + Avg
Fluoro-scopic	2/2, 5/5	2/2, 5/5	0.3/2, 0.3/5	2/2, 5/5
CT	2/2, 5/5	1.3/2, 3.3/5	1.3/2, 3.3/5	1.3/2, 0/5
MR	2/2, 5/5	2/2, 4/5	2/2, 0/5	1/2, 2.5/5
OT	2/2, 5/5	2/2, 2/5	0.7/2, 0.3/5	2/2, 1.6/5
Ultrasound	2/2, 5/5	2/2, 4.8/5	2/2, 4.8/5	2/2, 2.4/5

The denominator shows the number of colluders, and the numerator shows the average number of identified colluders. The false negative in colluder detection is zero. Avg denotes averaging attack.

6. Conclusions and open problems

Privacy protection of patients' medical records, including images, have become a pressing social issue after the HIPAA mandate took effect in April 2003. Privacy assurance of medical images at retrieval and during storage, represents the mainstream of research efforts to meet HIPAA requirements. However, privacy can be violated after the images are accessed by authorized parties. At present, to our knowledge, the problem of ensuring patient's privacy after the reception of medical images has not been addressed anywhere. This is an important problem since we try to protect privacy and also allow organizers to trace and deal with legal challenges.

In this paper, we identified and studied the problem of tracing the source of an unauthorized release of medical images in multicast environments to enhance patients' privacy. We proposed a multicast fingerprinting solution based on IA-W watermarking, in which a broadcast image needs to be decoded by watermark key holders before the image can be used to diagnose an illness and fingerprints are left on the images during the decoding. Our simulation results confirm that our fingerprinted images preserve higher perceptual quality than 10:1 JPEG compressed images, in terms of three image quality indices: PSNR, QI and MSME, while withstanding LPF, HPF, JPEG, cropping, and averaging attack. Our solution is scalable in user storage and watermark key update communication cost. Each user, who can be a sender or a recipient in many-to-many multicast, only needs to store one watermark key, which takes less storage than an $U \times V$ 16-bit image. No watermark key renewal is required for member revocation[5] and for a large number of member additions. Using our computationally efficient algorithm, described in Section 4.5, when an image to be watermarked is of common size, a sender and each recipient will only need to compute $U \times V$ additions or subtractions without performing multiplications, greatly reducing computational expense. Therefore, all the five design requirements set in Section 3.3 are achieved.

Storage requirements of the TTP and fast tracing are the issues not discussed in this paper. With periodic updates of watermarks, the number of watermarks stored in the system increases linearly as time progresses. This makes tracing a harder task due to the number of comparisons that need to be performed to identify a violator. Appropriate watermark key assignment may help solve this problem.

As image compression becomes widely accepted as an important tool for efficient transmission and storage, there is a need to investigate how much JPEG compression can be allowed in the fingerprinted images, so that diagnostic results are not affected. The consideration of compression mechanisms in watermark design, may yield better

[5] Encryption keys need to be renewed to control the access to the data being communicated.

perceptual fidelity results, under the same level of image compression.

7. Summary

The HIPPA mandate enforced on April 14, 2003, requires health care providers to protect the privacy of patients' health information. Medical images are one form of private health information, accessed by authorized health care professional. Current research and mechanisms on ensuring security and privacy of medical images focus on securing storage and transmission before reception. However, an authorized recipient can breach a patient's privacy by releasing medical images to unauthorized parties, without ever being held accountable for it. To our knowledge, thus far, there has not been any progress toward protecting patients' privacy after the data has been accessed by an authorized recipient. Research in this direction has two fold advantages. First, it enhances patients' privacy by serving as a deterrence against illegal distribution of patients' data. Second, for legal and law enforcement purpose, it allows health care institution to identify the recipient who illegally distributed the patient's record.

In the paper, we employ watermarking techniques to address the problem of enhancing patients' privacy by enabling tracing of the source of unauthorized image leakage unauthorized parties. Watermarking medical images for tracing purposes needs to satisfy stringent fidelity requirements to ensure diagnostic quality and high robustness for tracing purposes. We also note that often a group of medical professionals need to collaborate to discuss a patient's case, making group communication as a common practice in providing medical service. Hence we formulate and study the problem of tracing in a multiuser clinical environment in this paper.

We propose an efficient watermarking scheme suitable for many-to-many multicast scenario, where a broadcast copy is of diagnostic value only after being decoded by the authorized watermark key holders. During the process of decoding, two fingerprints that correspond to the original sender and the recipient who performs the decoding, are imprinted onto the image. Our scheme is scalable in user storage and watermark key update communication, as it requires only one watermark key to be stored by each user, and no watermark key update is required for member join and member revocation. To facilitate the use of our scheme in practice, we considered system implementation issues and completed an analytical performance evaluation. Simulation results conducted on 31 images of five modalities confirm that our fingerprinted images are of higher quality when compared to 10:1 JPEG compressed images,[6] in terms of three image quality indices: peak signal to noise ratio (PSNR), quality index (QI) and mean squared Moran error (MSME), while withstanding various image processing including low pass filter (LPF), high pass filter (HPF), JPEG compression, cropping, and averaging attack.

In this paper, we introduced an important new research problem on patient's privacy, which has been overlooked, and a method to resolve that problem. We further identified additional challenges and open problems which need to be addressed if the HIPPA privacy mandate is to be truly realized.

[6] JPEG CT body images with compression ratio 10:1 are considered acceptable for diagnostic purpose [18].

Acknowledgements

The authors thank Dr K.S. Chuang for providing source code for computing mean Moran error (MME) mean squared Moran error (MSME).

References

[1] United States Department of Health and Human Services. HIPAA: medical privacy—national standards to protect the privacy of personal health information. Available at: http://www.hhs.gov/ocr/hipaa/
[2] HEMA. DICOM: digital imaging and communication in medicine. Available at: http://medical.nema.org/
[3] Cao F, Huang HK, Zhou XQ. Medical image security in a HIPAA mandated PACS environment. Comput Med Imaging Graphics 2003; 27:185–96.
[4] Coatrieux G, Maitre H, Sankur B, Rolland Y, Collorec R. Relevance of watermarking in medical imaging. In: IEEE EMBS Conf on Inform Tech Appl in Biomedicine, Arlington, VA; Nov 2000.
[5] Acharya UR, Bhat PS, Jumar S, Min LC. Transmission and storage of medical images with patient information. Comput Bio Med 2003;33: 303–10.
[6] Cox IJ, Miller ML, Bloom JA. Digital watermarking. San Francisco, CA: Morgan Kaufmann Publishers; 1968.
[7] Cox IJ, Kilian J, Leighton FT, Shamoon T. Secure spread spectrum watermarking for multimedia. IEEE Trans Image Process 1997;6(12): 1673–8.
[8] Fridrich J, Goljan M, Du R. Invertible authentication. In: Proceedings of SPIE, security and watermarking of multimedia contents, San Jose, CA; Jan 2001.
[9] Wang Z, Bovik AC. A universal image quality index. IEEE Signal Process Lett 2002;9(3):81–4.
[10] Chen TJ, Chuang KS, Wu J, Chen SC, Hwang IM, Liu ML. A novel image quality index using Moran I statistics. Phys Med Biol 2003;48: 131–7.
[11] Chen TJ, Chuang KS, Chiang YC, Chang JH, Liu RS. A statistical method for evaluation quality of medical images: a case study in bit discarding and image compression. Comput Med Imaging Graphics 2004;28:167–75.
[12] Chen B, Wornell GW. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. IEEE Trans Inf Theory 2001;47(4):1423–43.
[13] Podilchuk CI, Zeng W. Image-adaptive watermarking using visual models. IEEE J Select Areas Commun 1998;16(4):525–39.
[14] Eggers JJ, Buml R, Tzschoppe R, Girod B. Scalar Costa scheme for information embedding. IEEE Trans Signal Process 2003;51(4): 1003–19.
[15] Watson AB. DCT quantization matrices visually optimized for individual images. In: Proceedings of SPIE human vision, visual processing and digital display IV, vol. 1913; 1993. p. 202–16.

[16] Watson AB, Yang GY, Solomon JA, Villasenor J. Visual thresholds for wavelet quantization error. In: Proceedings of SPIE human vision and electronic imaging, vol. 2657; 1996. p. 381–92.
[17] Chen K, Ranmabadran TV. Near-lossless compression of medical images through entropy-coded dpcm. IEEE Trans Med Imaging 1994; 3(3):538–48.
[18] Kalyanpur A, Neklesa VP, Taylor CR, Daftary AR, Brink JA. Evaluation of JPEG and wavelet compression of body CT images for direct digital teleradiologic transmission. Radiology 2000.
[19] Judge PQ, Ammar MH. Whim: watermarking multicast video with a hierarchy of intermediaries. In: Proceedings of the 10th international workshop on network and operation system support for digital audio and video, Chapel Hill, NC; June 2000.
[20] Brown I, Perkins C, Crowcroft J. Watercasting: distributed watermarking of multicast media. In: Proceedings of networked group communication, Pisa, Italy; Nov 1999.
[21] Wu T, Wu S. Selective encryption and watermarking of MPEG video. Technical report, NC state University; 2000.
[22] Parviainen R, Parnes P. Large scale distributed watermarking of multicast media through encryption. In: Proceedings of the international federation for information processing, communication and multimedia security joint working conference IFIP TC6 and TC11; 2001. p. 149–58.
[23] Chu H, Qiao L, Nahrstedt K. Secure multicast protocol with copyright protection. ACM SIGCOMM Comput Commun Rev 2002;32(2): 42–60.
[24] Li M, Narayanan S, Poovendran R. Tracing medical images using multi-band watermarks. In: Proceedings of IEEE EMBC'04, San Francisco, CA; Sept 2004.
[25] Percival DB, Walden AT. Wavelet methods for time series analysis. Cambridge: Cambridge University Press; 2000 p. 262–263.
[26] Poovendran R, Baras JS. An information—theoretic approach for design and analysis of rootedtree-based multicast key management schemes. IEEE Trans Inf Theory 2001;47(7):2824–34.
[27] Van Trees HL. Detection, estimation, and modulation theory, part 1. New York: Wiley; 1968.
[28] Barre S. Available at: http://www.barre.nom.fr/medical/samples/

Mingyan Li is a PhD candidate in the Department of Electrical Engineering at University of Washington, Seattle. Her research interest is in the area of multiuser security with applications to multicast, medical security systems, and sensor networks. She is a recipient of the departmental teaching assistant award in 2002 and the departmental Society of Women in Engineering (SWE) women graduate engineer of year 2003.

Radha Poovendran is an assistant professor in the department of electrical engineering at the University of Washington, Seattle. He received all his degrees in Electrical Engineering. His research interests are in the area of applied cryptography for multiuser environments. For his research contributions in security, he has been a recipient of the 1999 Rising Star Award from the National Security Agency, 2001 Faculty Early Career Award from the National Science Foundation, 2002 Young Investigator Award from the Army Research Office, and the 2004 Young Investigator Award from the Office of Naval Research. In 2002, he was also a recipient of the Outstanding Teaching Award as well as the Outstanding Research Advisor Award of the electrical engineering department. He is a nominee of the 2005 Presidential Early Career Award for Scientists and Engineers.

Sreeram Narayanan completed his Bachelor of Engineering (BE) in Electronics and Communication from Maulana Azad College of Technology (now Maulana Azad National Institute of Technology) Bhopal, India in 1999. He completed his Master of Science in Electrical Engineering from University of Washington in 2001. In 2004 he earned Doctor of Philosophy in Electrical Engineering from University of Washington. He is currently employed as senior fellow at the Department of Radiation Oncology at University of Washington. His research interests include computer vision, image processing, 3D reconstruction and radiation treatment planning and dosimetry.