Analysis of a Zero Location based Authentication Scheme for Biomedical Images

Antionette W.-T. Goh
Information Security Research Laboratory
Swinburne University of Technology
Malaysia
agoh@swinburne.edu.my

M. L. Dennis Wong
Information Security Research Laboratory
Swinburne University of Technology
Malaysia
dwong@swinburne.edu.my

Raphael C.-W. Phan
Security and Cryptography Laboratory
EPFL
Switzerland
raphael.phan@epfl.ch

ABSTRACT

In this paper, a fragile watermarking scheme proposed for authentication of biomedical images in the Z-transform domain is analysed. The studied scheme takes advantage of the zero locations in the Z-domain which are sensitive to any tampering made on a watermarked host image. However, we are able to refute the designers' claims by carrying out at least two attacks on the scheme. Our attacks are based on the ability to determine the embedded check-bits in the watermarked image and thereafter, exploit the locations of the check-bits to alter the watermarked image while still resulting in successful authentication in the end.

Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information Systems**]: Security and Protection—*authentication, unauthorized access*

General Terms

Security

Keywords

Watermarking, integrity, zero locations, Z-transform, biomedical images

1. INTRODUCTION

Authentication of digital biomedical images is an arising concern as these images carry potentially sensitive data. The move to convert old analog biomedical images to their digital equivalents has been largely applauded due to reasons of convenient storage and retrieval, especially in hospital databases. However, such a move attracts undesirable tampering of biomedical images which could result in forensic and insurance frauds. Furthermore, it is essential that the digital equivalent does not differ perceptually from its original to ensure correct diagnosis. Therefore, there has to be a measure of integrity and authenticity to protect the virtue of the images.

Several watermarking methods have been proposed over the last decade with the purpose of content authentication for biomedical images solely. Most watermarking schemes proposed for image authentication are fragile types.

Fragile watermarking algorithms are usually strict tamper detection tools. They are built on the basis of inserting a watermark in such a way that any attempt to alter the host image will also result in altering (destroying) the watermark itself. As such, any manipulation of the image immediately causes the content itself to lose integrity. Also with fragile watermarking schemes, one can locate the regions of distorsions on the image that have been tampered with.

Due to the strict nature of biomedical images, fragile watermarking schemes were readily adopted to provide content authentication. Examples of work based on fragile watermarks can be found in [1, 2, 3].

Ho et al. [5] recently proposed a novel fragile watermarking scheme for authentication of biomedical images based on the zero locations in the Z-transform domain [6]. To the best of our knowledge, this is the first time that zero locations have been applied in transform domain watermarking. Although Ho et al.'s work is considered isolated in this field, we believe that the scheme is worth analyzing as it contributes to the progress of existing literature in the area of biomedical image authentication which is scarce.

In their work, Ho et al. proposed to embed the watermark bits based on a secret random sequence into the negative real roots of the Z-transform coefficients (zero locations). The argument is that the zero locations are sensitive to any image pixel changes. As such, this property provides a good solution towards any content alteration. During extraction, the scheme checks the presence of the watermark against the secret random sequence (used during embedding) to verify the authenticity of the image. The assumption taken is that the secret random sequence is unknown to an attacker and cannot be derived by the attacker, therefore he is unable to alter the embedded watermark. The authentication part locates the watermark distortions first, followed by the locating of altered regions on the image.

However, in this work, we exploit the embedding mechanism and hence are able to retrieve the embedded secret sequence, upon which the digital image could be altered

E-FORENSICS 2008, January 21-23, Adelaide, Australia

DOI 10.4108/e-forensics.2008.2674

and thus an attacker can succeed in fooling the detector. Hence, the authentication scheme is not as secure as claimed. We also suggest some countermeasures to secure the scheme against our attacks.

2. HO ET AL'S SCHEME

The proposed scheme is divided in two parts: *hiding* of check-bits and *authentication* of check-bits after extraction.

The first part consists of three stages: separation of image into blocks, applying Z-transform and embedding of check-bits.

A host image is separated into K non-overlapping blocks of a fixed size of 8 x 8 pixels. A 1-D Z-transform is then carried out on each row of pixels of each block. The Z-transform of a sequence, *x(n)*, which is causal (i.e. $x_n = 0$ where $n = 0$) is defined as

$$X(z) = \sum_{n=0}^{\infty} x(n) z^{-n}, \tag{1}$$

where z is a complex variable, whose form depends on the sequence itself. Equation (1) is then used when processing the row-by-row Z-transform in each sub-block. We normalize the first pixel of each line to unity and in the case that the first pixel of a given line is zero, we then bypass this particular step. Such normalization is customary as it facilitates the computation of the Z-transform and its inverse.

Only the zeros are considered whereas all poles are equivalent to zero or a null value in the system. Ho et al. chose to use only the absolute value or amplitude of the negative real roots of the zeros for embedding of check-bits to avoid the phase change caused by the complex number computation [5].

After the check-bits are generated using a random seed which is kept secret, the check-bits are stored in an output key file. In their experiment to prove their claims, 8 check-bits were embedded into each block, i.e. one check bit per row of the block. Due to the implementation of the proposed embedding algorithm in [5], some roots may have been relocated after crossing the threshold determined by the mean of a biomedical image.

From their experimental biomedical image *Brain*, the histogram distribution is observed to have a mean of -1. This mean value is chosen as the mid-point to decide whether to allocate a binary check bit (i.e. watermark) of 0 or 1. For example, if the check bit is 1 and the current location of the real root is at a negative ϵ distance away from the -1 threshold, it has to be moved to the right of the threshold to encode a bit 1. If the check bit is 0 and the current location of the real root is at a positive ϵ distance away from the -1 threshold, it has to be moved to the left of the threshold to encode a bit 0. Despite causing some minor changes to the image, Ho et al. attest that the distortions are not perceptually visible and thus, can be considered negligible. Proofs of their experimental results are shown in [5]. However, it is interesting to note that there is no mention of their tested range of ϵ for their experiments.

After the check-bits are encoded into all blocks, the zeros are then transformed back to the pixel sequence using the inverse Z-transform equation. To recover a causal sequence, $x'(n)$, whose Z-transform is $X'(z)$, that is,

$$x'(n) = Z^{-1}[X'(z)]. \tag{2}$$

Upon completion of the inverse Z-transform, another image pixel sequence, $x'(n)$, is produced. One may note that there may be some different values when compared with the original image pixel sequence, $x(n)$. However, the difference is not perceptually significant even with different values of pixels imposed on the image. Before we can obtain the watermarked block, some operations like re-arrangement and scaling of pixel values [5] have to be carried out. The whole embedding process is repeated for every single block, resulting in a watermarked object, when all the blocks are recombined later.

For authentication at the receiving end, the embedded information is extracted and checked to see whether the check-bits have been changed or not. The negative real roots of the image pixels are compared with the check-bits sent through the output key file. If the region has been unaltered, the authentication is successful. However, if the region is found to be altered, the authentication returns a "0" indicating that the particular region has failed the authentication process.

3. ATTACKING THE ZERO LOCATIONS

The security of Ho et al.'s scheme relies on a secret random seed which is used to generate the random check-bit sequence. It is assumed that the check-bits are not disclosed to anyone and the secret seed is unknown too. Therefore any modification made by a malicious attacker would alter the check-bits and in turn would fail the authentication process.

Under the assumption of Kerckhoffs' principle [7], it is customary to assume that every detail and the implementation of a security scheme is available except for the secret. As the proposed scheme is a fragile watermarking scheme, the objective from the attackers' point of view is to alter the image without getting caught by the detector in order to have a successful attack. In Ho et al.'s scheme, the embedder only encodes the check-bits into the real negative root, it is this act that allows possible exploits to compromise the security of the proposed algorithm. It turned out that one does not need the secret seed in order to mount a successful attack.

3.1 Attack I: Modulation of Complex Conjugate Roots

We take a block of the image and perform steps as in the embedding algorithm, i.e. take each row and perform a Z-transform. In Ho et al.'s setting, we would have 3 pairs of complex conjugate roots and 1 real root for each check-bit embedded. We then applied a chosen scaling factor, β to the 3 pairs of complex conjugate roots but not to the real roots. After which, the inverse Z-transform is applied.

This would give rise to a perceptually very different sub-block that would still pass the authentication process.

The reason for this is that the Z-transform could be viewed as a generalisation of the Discrete Time Fourier transform (DTFT) [9], as the DTFT is found by evaluating $X(z)$ on the unit circle.

$$X(\omega) = Z(e^{j\omega}) \tag{3}$$

Assuming a stable system, the magnitude response of a discrete time system could be found by calculating the ratio of the factorial distances between the poles and the zeros to the unit circle at angle, θ [10]:

$$\frac{A(\theta)}{A_0} = \frac{\prod_k (\text{distance from } k^{th} \text{ pole to the point at } \theta)}{\prod_p (\text{distance from } p^{th} \text{ zero to the point at } \theta)} \tag{4}$$

The real negative root coincides with a digital frequency of π, which is normally interpreted as the Nyquist Frequency. The Nyquist Frequency (by definition, half of the sampling frequency) is the highest frequency component one could obtain from a Discrete Time Fourier Analysis such as the DTFT. Therefore when the negative real root is modulated, in reality the high-frequency feature of that row is altered. High frequency components relate to sharp edges in the spatial domain. Therefore, the proposed scheme has a good perceptual quality from the spatial domain.

On the other hand, if one is to alter the other complex conjugate roots, s/he would be altering the low pass or band pass characteristics of an image as well. Thus, changes are made to the other roots leaving the modulated negative real root untouched. However, the changes must be symmetrical for the conjugate pairs to avoid possible suspicions. A simple scaling factor applied to each conjugate pairs would satisfy this requirement. Experimental results are shown in Section 4.

The steps of the attack based on a block of the image are summarised in Table 1.

3.2 Attack II: Retrieval of Check-bits

Attack I does not give the attacker much control to create modifications on the original image as the attack is carried out in the Z-transform domain. We demonstrate in Attack II a method that allows one to replace the image block with a chosen pattern. This is possible owing to the fact that in the proposed scheme, the random check-bits could be retrieved from the embedded image. Once the secret check-bits are known, one could then make any modification and then re-insert the retrieved check-bits into the altered image. As mentioned in previous sections, the embedder hides the $\{0, 1\}$ check-bit patterns by modulating the negative real root to be either greater than -1 or less than -1. We could retrieve the check-bits simply by performing the Z-transform procedures and judge the position of the negative real root to determine the value of the embedded check bit.

The steps of the attack are enumerated in Table 2.

4. EXPERIMENTAL RESULTS

In our efforts to verify Ho et al.'s scheme, we used 512 x 512 grayscale single frame biomedical image samples provided by S. Barré [8]. The proposed scheme was implemented on Matlab version 7.0.4.

To demonstrate the attacks, for simplicity but without loss of generality, we used just 8 lines of 8 pixels to test our attacks due to the fact that the blocks are not dependent on each other. Two 512 x 512 grayscale biomedical images (OT-MONO2-8-a7 and OT-MONO2-8-hip) were used as shown in Fig.1(a) and Fig. 2(a) . The ϵ used was 0.0005. After the embedding process was carried out, distorsions were observed on the watermarked images, especially on the edges of the digital images. More specifically, when there is more than one "0" value pixel in the row of an 8 x 8 block, distorsions occur also within the image itself as seen in Figs. 1(b) and 2(b). Such distorsions would result in fallacious diagnosis which is undesirable in any biomedical image watermarking scheme.

The histogram distribution of negative real roots of both images had a mean of -1 as shown in Fig. 3. This showed that our implementation of the embedding process was correct.

However, in-depth analysis showed that the areas which suffered distorsion did not have any negative real roots. An example of a row of pixels with more than one "0" value pixel is transformed in the Z-domain; $\{118, 86, 47, 31, 18, 0, 0, 0, 6\}$. This is confirmed in Fig. 4 where there are only 4 pairs of complex conjugate roots with no presence of the negative real root.

Thus, the embedding process at those particular blocks could not be carried out. This observation highlights the lack of hindsight of the designers of the proposed scheme who assumed that there would only be one negative real root per row of an 8 x 8 block of pixels.

4.1 Attack I: Modulation of Complex Conjugate Roots

Fig. 5(a) shows the original block of 8 lines of 8 pixels. We then embedded the secret check-bits into the block as displayed in Fig. 5(b). The content of the watermarked image was then modified using a simple scaling factor of 0.250. As illustrated in Fig. 5(c), the modified areas were still considered authentic and therefore, passed the image authentication process.

4.2 Attack II: Retrieval of Check-bits

First, the check-bits are extracted from the received watermarked image based on steps 2 to 4 of Table 2. Then, the image (without the watermark) is manipulated as shown in Fig. 6(b). Thereafter, the check-bits are reinserted onto the altered image shown in Fig. 6(c), followed by a subsequent cheating of the authentication process. The altered image is still considered authentic as the watermark has not changed.

4.3 Summary and Proposed Countermeasures

In the first attack, we successfully modified the watermarked image without altering the embedded watermark. Our second attack creates a new watermarked image that the authentication detector believes as authentic. Using the method of only altering the image without affecting the check-bits, the watermark would remain at its original position and therefore, the image authentication process would not detect any falsification. Furthermore, the integrity of the watermarked image in this scheme is solely based on fragile watermarks (i.e. the embedded check-bits) which are independent of the image content.

A simple solution to our attacks is to make the fragile watermark dependent on the image content. We also propose to make all the zero locations and the position for embedding the check-bits initialized by a separate random number generator with a second secret key.

5. CONCLUSION

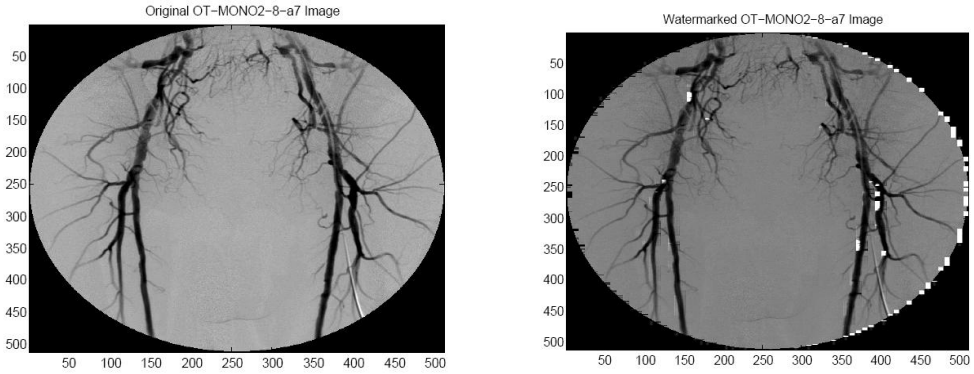
Watermarking the zeros location in the Z-domain gives strong sentivity against tampering. Thus, it could be used for authentication of biomedical images as suggested in [5]. However,the security aspect of a watermarking scheme is often overlooked by its designers, such that the exact figure of merit could well be the potential security exploit as reviewed in this paper. We presented two simple attacks and a verification of the implementation of Ho et al.'s scheme. Attack I is easy to implement but we have no control of the image after the attack. With a little extra work, Attack II, on the other hand, allows one to replace an image block with a

#	Steps
1	Obtain watermarked image
2	Perform Z-transform to find position of roots
3	Apply a scaling factor, β to the complex conjugate roots only
4	Perform inverse Z-transform
5	Authenticate tampered image (Result: Authentication PASSED)

Table 1: Attack I: Modulation of Complex Conjugate Roots

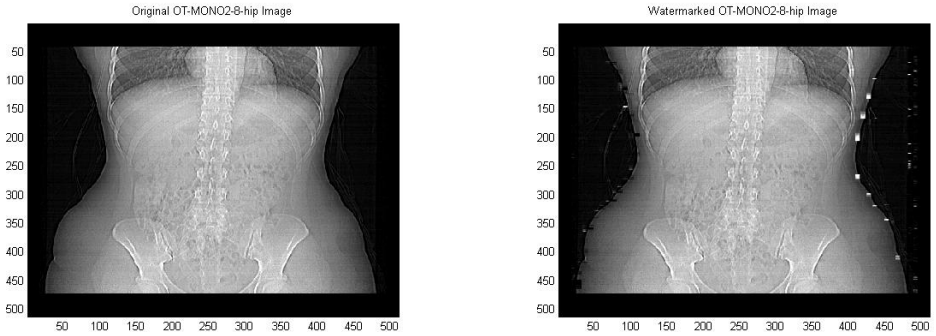
#	Steps
1	Obtain watermarked image
2	Perform Z-transform to obtain negative real roots
3	Check position of embedded check-bits
4	Extract check-bits
5	Modify image (without watermark) until perceptually different from image in Step 1
6	Embed retrieved check-bits into their original locations
7	Authenticate tampered image (Result: Authentication PASSED)

Table 2: Attack II: Retrieval of Check-bits



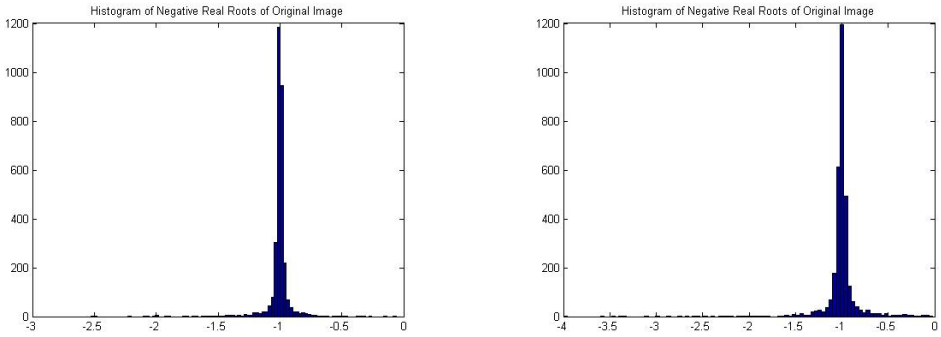
(a) Original Image (b) After Embedding

Figure 1: Experimental results of OT-MONO2-8-a7



(a) Original Image (b) After Embedding

Figure 2: Experimental results of OT-MONO2-8-hip



(a) OT-MONO2-8-a7 (b) OT-MONO2-8-hip

Figure 3: Histogram distribution of negative real roots

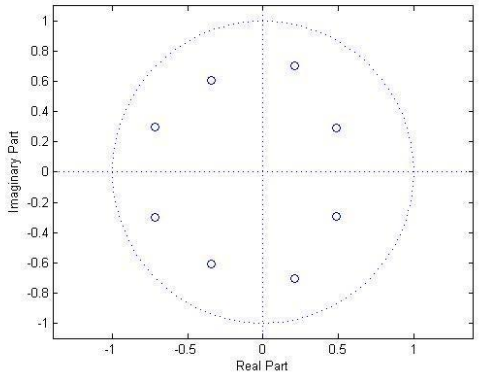
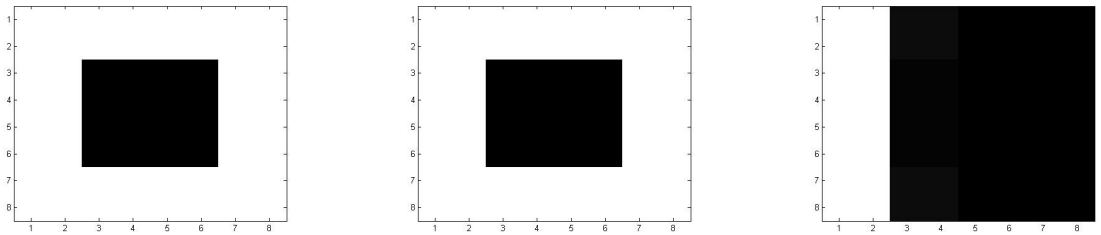
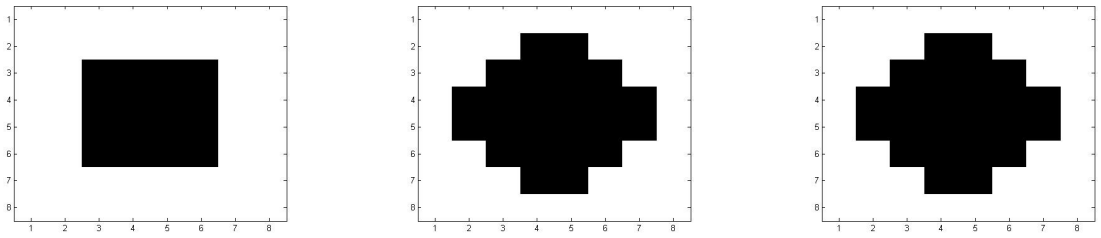


Figure 4: Transforming a Row of 8 x 8 Pixels in the Z-Domain



(a) Original Block (b) After Embedding (c) After Attacking Authentication:PASSED

Figure 5: An Illustration of Attack I



(a) Received Watermarked Block

(b) Modified Image (Without Watermark)

(c) After Re-embedding Authentication:PASSED

Figure 6: An Illustration of Attack II

chosen pattern without being detected. We conclude by reiterating the importance of security analysis while designing watermarking schemes [4], especially ones such as the Ho et al. scheme designed for forensic analysis of biomedical images, where integrity is concerned.

6. REFERENCES

[1] R. Acharya, U. C. Niranjan, S. S. Lyengar, N. Kannathal, and C. M. Lim. Simultaneous storage of patient information with medical images in the frequency domain. *Computer Methods and Programs in Biomedicine*, 76:13–19, 2004.

[2] G. Coatrieux, H. Maitre, and B. Sankur. Strict integrity control of biomedical images. In *Proc. SPIE Vol. 4314: Security and Watermarking of Multimedia Contents III*, 2001.

[3] G. Coatrieux, H. Maitre, B. Sankur, Y. Rolland, and R. Collorec. Relevance of watermarking in medical imaging. In *Proc 3rd IEEE EMBS Int. Conf. on Information Technology Applications in Biomedicine, 3rd Workshop of the Int. Telemedical Information Society*, Nov. 2000.

[4] T. Furon. A survey of watermarking security. In *Proc Digital Watermarking, 4th International Workshop, IWDW 2005, Siena, Italy*, Sept. 2005.

[5] A. T. S. Ho, X. Zhu, and J. Shen. Authentication of biomedical images based on zero location watermarking. In *Proc. 8th Int. Conf. on Control, Automation, Robotics, and Vision*, 2004.

[6] G. James. *Advanced Modern Engineering Mathematics*. Addison Wesley Publishing Company, second edition, 2000.

[7] A. Kerckhoffs. La cryptographie militaire. *Journal des Sciences Militaires*, 8:5–38, 1883.

[8] S. Barré. Medical imaging: Samples [online]. 2003. Available from: http://www.barre.nom.fr/medical/samples/.

[9] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. PWS Publishing, second edition, 1999.

[10] R. Strum and D. Kirk. *First Principles of Discrete Systems and Digital Signal Processing*. Addison-Wesley, first edition, 1989.