

on a secret key. Our paper presents an adaptive collusion attack to the buyer watermarking scheme by selectively manipulating the watermarked pixels. Concretely, when the traitors find two unequal watermarked pixels generated from the same original pixel, they average these two pixels so as to alleviate the watermark information. This attack not only removes the watermark so that the traitors escape from being identified, but also increases the watermarked image quality. We present a theoretical analysis on the size of traitor group and quality improvement. Our experimental result and theoretical analysis show that the attack is effective.

## REFERENCES

- [1] D. P. Mukherjee, S. Maitra, and S. T. Acton, "Spatial domain digital watermarking of multimedia objects for buyer authentication," *IEEE Trans. Multimedia*, vol. 6, no. 1, pp. 1–15, Feb. 2004.
- [2] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoan, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.
- [3] I. Pitas, "A method for signature casting on digital images," in *IEEE Int. Conf. Image Processing*, 1996, vol. 3, pp. 215–218.
- [4] R. B. Wolfgang and E. J. Delp, "A watermark for digital images," in *IEEE Int. Conf. on Image Processing*, 1996, vol. 3, pp. 219–222.
- [5] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Transparent robust image watermarking," in *IEEE Int. Conf. Image Processing*, 1996, vol. 3, pp. 211–214.
- [6] J. F. Delaigle, C. De Vleeschouwer, and B. Macq, "Digital watermarking," in *Proc. SPIE, Optical Security and Counterfeit Deterrence Techniques*, 1996, vol. 2659, pp. 99–110.
- [7] F. Ergun, J. Kilian, and R. Kumar, "A note on the limits of collusion-resistant watermarks," *Advances in Cryptology—EUROCRYPT'99 Lecture Notes in Computer Science*, vol. 1592, pp. 140–149, 1999.
- [8] T. K. Das and S. Maitra, "A robust pixel oriented watermarking scheme in spatial domain," *International Conference on Information and Communications Security (ICICS'02) Lecture Notes in Computer Science*, vol. 2513, pp. 184–196, 2002.
- [9] W. Trappe, M. Wu, Z. J. Wang, and K. J. Ray Liu, "Anti-collusion fingerprinting for multimedia," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 1069–1087, 2003.
- [10] Y. Wu and R. Deng, "Adaptive collusion attack to a block oriented watermarking scheme," *International Conference on Information Communication Security (ICICS) Lecture Notes in Computer Science*, vol. 2836, pp. 238–248, 2003.
- [11] Y. Wu, "Linear combination collusion attack and its application on an anti-collusion fingerprinting," in *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Philadelphia, PA, Mar. 2005, pp. II.13–II.16.
- [12] M. Wu, W. Trappe, Z. Wang, and K. J. R. Liu, "Collusion resistant fingerprinting for multimedia," *IEEE Signal Process. Mag., Special Issue on Digital Rights Management*, pp. 15–27, 2004.
- [13] H. Zhao, M. Wu, Z. J. Wang, and K. J. R. Liu, "Nonlinear collusion attacks on independent fingerprints for multimedia," *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. V-664–V-667, 2003.

## Cryptanalysis of Chu's DCT Based Watermarking Scheme

Tanmoy Kanti Das, Subhamoy Maitra, and Jianying Zhou

**Abstract**—In 2003, Chu proposed an oblivious watermarking algorithm by modifying the CKLS scheme proposed by Cox, Kilian, Leighton, and Shamoan in 1997, known as the CKLS scheme. In this correspondence, we report that the modification presented by Chu is susceptible to a suitably modified attack devised by Das and Maitra in 2004. In fact, the experimental results show that Chu's scheme is even weaker than the CKLS scheme in terms of our attack.

**Index Terms**—Cryptanalysis, digital watermarking, single copy attacks, subsampling.

## I. INTRODUCTION

As the quest for robust digital watermarking schemes becomes more and more intense, researchers are designing new watermarking schemes or extending the existing ones keeping in mind the changing needs of the user. However, during design and extension of watermarking schemes, most of the times security of the watermarking schemes gets neglected. There exist a number of image processing based benchmarks which robust watermarking techniques should pass. However, these benchmarks never take into account the individual watermarking techniques to discover any design flaws that may be unique. Thus, there is a need to analyze each individual scheme in detail to identify its weaknesses.

In this correspondence, we concentrate on Chu's [3] watermarking strategy, which is an extension of the basic CKLS [4] scheme. We show that 1) Chu's scheme [3] is not secured and 2) it is even less secured than the original CKLS scheme [4] with respect to a suitably modified cryptanalysis proposed in [5].

Given an image  $I$ , by  $\mathcal{N}(I)$  we denote the set of images which are visually indistinguishable from  $I$ . A typical invisible watermarking scheme adds a signal  $s^{(i)}$  to the original image  $I$  in such a manner that the watermarked image  $I^{(i)} = I + s^{(i)}$  remains in  $\mathcal{N}(I)$ . The signal  $s^{(i)}$ , known as watermark, helps in forensic tracking of the buyer. Invisible watermarking schemes are divided into two groups depending on the requirement of original image during watermark verification. Oblivious schemes unlike nonoblivious schemes do not require the original image during watermark verification. Generally in nonoblivious schemes original image  $I$  is subtracted from the watermarked image  $I^{(i)}$  to retrieve a signal  $s^{\#} = I^{(i)} - I$ . On the other hand, in oblivious schemes, the signal  $s^{\#}$  is recovered using some other information related to original image, but not the image itself. Buyer  $i$  is suspected if correlation between recovered signal  $s^{\#}$  and embedded signal  $s^{(i)}$  is significant. It is also important that the probability of wrongly implicating an honest buyer should be extremely small.

Note that, most of the present watermarking schemes are correlation based, i.e., during verification process, the correlation between embedded signal  $s^{(i)}$  and recovered signal  $s^{\#}$  is used as the measure of confidence. Thus it is easy to see that an attacker would try to remove this correlation to evade detection. To achieve his goal, given a watermarked image  $I^{(i)}$ , the attacker attempts to construct an image  $I^{(i)\#}$

Manuscript received October 12, 2004; revised July 26, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Wenjun (Kevin) Zeng.

T. K. Das and J. Zhou are with the Infocomm Security Department, Institute for Infocomm Research, Singapore 119613 (e-mail: tkdas@i2r.a-star.edu.sg, jyzhou@i2r.a-star.edu.sg).

S. Maitra is with Applied Statistics Unit, Indian Statistical Institute, Calcutta 700 108, India (e-mail: subho@isical.ac.in).

Digital Object Identifier 10.1109/TMM.2006.870718

in such a manner that  $I^{W^\#} \in \mathcal{N}(I)$  and  $s^\# = I^{W^\#} - I$  is uncorrelated to  $s^{(i)}$ . Thus it is not possible to identify the malicious buyer  $i$  any more.

## II. CHU'S SCHEME: AN OVERVIEW

Chu's [3] scheme is an extension of the CKLS [4] scheme. However, it no longer requires the original image during watermark verification unlike CKLS scheme, i.e., the extended scheme is oblivious. To convert the existing scheme to oblivious one, author first subsamples the original image  $I$  to generate  $m$  different subimages. During subsampling, one should assign each pixel from a block of size  $2 \times 2$  to different subimages. As example, one can create four subimages by subsampling  $I$  as follows:

$$\begin{aligned} I_1[i, j] &= I[2i, 2j], & I_2[i, j] &= I[2i - 1, 2j], \\ I_3[i, j] &= I[2i, 2j + 1], & I_4[i, j] &= I[2i + 1, 2j + 1], \end{aligned}$$

where,  $i = 0, \dots, (N_1/2) - 1, j = 0, \dots, (N_2/2) - 1, N_1, N_2$  are the height and width of the original image  $I$ . Each of these subimages are subjected to Discrete Cosine Transform (DCT) before watermark insertion. The watermark  $W$  is a sequence of real numbers, selected from the standard normal distribution with mean 0 and standard deviation 1. DCT coefficients from corresponding location of the subimages are modified in different fashion to embed the watermark. Typically, a sample  $W_k$  of watermark sequence  $W$  is embedded in a pair of subimages using basic CKLS [4] method. Let the  $k$ th and  $l$ th subimages be employed to embed  $W_k$ , which can be accomplished in the following manner:

$$I'_{kD}[i] = I_{kD}[i](1 + \alpha W_k) \quad (1)$$

$$I'_{lD}[i] = I_{lD}[i](1 - \alpha W_k) \quad (2)$$

Here  $I_{kD}, I_{lD}$  and  $I'_{kD}, I'_{lD}$  indicate DCT domain representations of the subimages  $I_k, I_l$  and corresponding watermarked subimages for  $k \neq l$ . We consider the two dimensional DCT coefficient matrices as single dimensional arrays and use only one index  $i$ , instead of two indices  $i, j$ . Here  $\alpha$  is known as watermark strength [4] and used to control the distortion introduced due to watermarking. An "watermark insertion order sequence" determines how a pair of coefficient is chosen from the available coefficients of different subimages for a particular location  $i$ . As the number of possible order sequences is very large, it is not possible for an attacker, without any knowledge of the proper order sequence, to extract the watermark in a reasonable amount of time. Thus order sequence acts as a part of the secret key in watermark insertion process. Also  $\text{If} \left| \frac{(I_{kD}[i] - I_{lD}[i])}{(I_{kD}[i] + I_{lD}[i])} \right| \geq 3\alpha$ , one should not use the pair  $(I_{kD}[i], I_{lD}[i])$  to embed the watermark [3].

Extraction of watermark is possible by comparing DCT coefficient of different subimages as same watermark order sequence is available during extraction. Let  $(I'_{iD}[i], I'_{jD}[i])$  indicates a pair of coefficients selected according to the order sequence. If  $\left| \frac{(I'_{iD}[i] - I'_{jD}[i])}{(I'_{iD}[i] + I'_{jD}[i])} \right| \geq 3\alpha$ , then no watermark had been inserted in this pair, otherwise the extracted watermark  $W'[i] = (1/\alpha) \left[ \frac{(I'_{iD}[i] - I'_{jD}[i])}{(I'_{iD}[i] + I'_{jD}[i])} \right]$ . Like the CKLS scheme, the correlation between the embedded and extracted watermark is used as the measure of confidence.

Chu's watermarking scheme divides the original image into several subimages. Then it watermarks each of them, so that information from different subimages can be compared to recover the secret watermark signal. Basically, each of the subimages are watermarked using CKLS scheme employing either (1) or (2). We apply our attack reported in [5] on each of the four subimages first. Then we select four subimages with certain properties to construct the attacked image. Experimental results

show that our attack is successful as the correlation between attacked and watermarked images are negligible.

## III. PROPOSED CRYPTANALYTIC ATTACK

We start by presenting the security paradigm of a digital watermarking scheme from cryptographic viewpoint. It is well accepted in the field of cryptography and information hiding [1] that the algorithm will be known to the attacker but not the secret information (key or watermark). Hence, a watermarking strategy is considered to be secured, if a cryptanalyst, who has access to the algorithmic principle of the strategy but has no access to the secret key, should not be able to erase the watermark [1]. This principle had been introduced by Kerckhoffs [6] in as early as 1883. There are many instances to justify that "Security by Obscurity" (the assumption that opponent will stay ignorant about the system being used) can't work. One of the recent examples is the Secure Digital Music Initiative (SDMI) challenge [2]. In the challenge, the algorithmic principles of watermarking strategies were kept secret. In spite of that, authors of [2] were able to successfully attack one of the schemes. Thus while analyzing a watermarking scheme, it is assumed that the scheme is known to the attacker but the keys are unknown. In light of this, we consider that (i) the watermarking algorithm and the number of subimages are **known** and (ii) the order sequence (secret information) is **not known** to the attacker.

It has been assumed by Chu that different subimages generated during the watermark embedding are almost similar (see [3, equation 3]) and the success of the decoder depends on how well this approximation works. This poses a constraint on the number of subimages one can generate. In fact, subimages will be quite similar only when the number of subimages generated is four. In all other cases this approximation will be coarse enough to degrade the performance to a large extent. Thus one can safely assume that for all practical purposes, number of subimages generated is four. Indeed, all the algorithms and experimental results reported in [3] are on the basis that the number of subimages is four.

### A. Basic Statistical Tools and Fitting DCT Polynomial

We like to summarize the results from [5, Sec. II-A and II-B] for the tools used for cryptanalysis.

Consider that the data set  $x_1, \dots, x_t$  represents the original image where the watermark will be added. Now we consider another data set  $v_1, \dots, v_t$  taken from a distribution with mean  $\bar{v}$  and standard deviation  $\sigma_v$ , which will basically work as the watermark. Hence the watermarked image can be viewed as a data set  $x_1(1 + \alpha v_1), \dots, x_t(1 + \alpha v_t)$ , where,  $\alpha$  is a very small value which does not disturb the visual quality of the image. Further, consider another watermarked image  $x_1(1 - \alpha \mu_1), \dots, x_t(1 + \alpha \mu_t)$ , where the watermark  $\mu_1, \dots, \mu_t$  is selected from a distribution with mean  $\bar{\mu}$  and standard deviation  $\sigma_\mu$ .

Now analyze the data  $x_1(1 + \alpha v_1), \dots, x_t(1 + \alpha v_t)$ , with respect to the data  $x_1(1 + \alpha \mu_1), \dots, x_t(1 + \alpha \mu_t)$ . We need this to understand what is the distribution of watermarking data when considered with respect to the data available from another watermarked image. Thus we need to calculate the parameters for the data set

$$\frac{x_k(1 + \alpha v_k) - x_k(1 + \alpha \mu_k)}{\alpha v_k(1 + \alpha v_k)} = \frac{v_k - \mu_k}{(1 + \alpha v_k)}, \text{ for } k=1, \dots, t. \quad (3)$$

From the basic assumption of uncorrelatedness of two different watermark signal, we always assume that the data sets  $v_1, \dots, v_t$  and  $\mu_1, \dots, \mu_t$  are uncorrelated, i.e.,  $\sum_{k=1}^t \mu_k v_k = 0$ .

When we consider that  $\mu_k, v_k$  are chosen from standard normal distribution [4], [5], then  $\bar{\mu} = \bar{v} = 0$  and  $\sigma_\mu = \sigma_v = 1$ . This also gives that as  $(1/t) \sum_{k=1}^t \mu_k^2 = (1/t) \sum_{k=1}^t v_k^2 = 1$ .

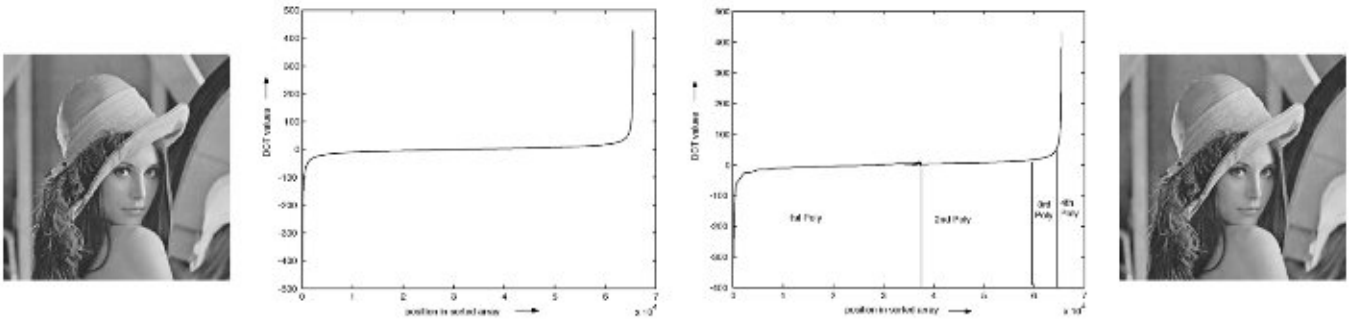


Fig. 1. Original image, sorted DCT data, DCT polynomials and the recovered image from the DCT polynomials.

To calculate the mean of the data set  $(\mu_k - \nu_k / (1 - \alpha\nu_k))$ , for  $k = 1, \dots, t$ , we approximate this by  $(\mu_k - \nu_k) / (1 - \alpha\nu_k)$ , as for small  $\alpha$ ,  $(1 + \alpha\nu_k)^{-1}$  can be approximated as  $(1 - \alpha\nu_k)$ . Thus the approximate mean can be calculated as

$$\frac{1}{t} \sum_{k=1}^t (\mu_k - \nu_k - \alpha\mu_k\nu_k + \alpha\nu_k^2) = \alpha. \quad (4)$$

To calculate the standard deviation of the data set  $(\mu_k - \nu_k / (1 - \alpha\nu_k))$ , for  $k = 1, \dots, t$ , we approximate this in a coarser way by  $(\mu_k - \nu_k)$ , as for small  $\alpha$ ,  $(1 + \alpha\nu_k)$  can be approximated as 1. Thus the approximate variance can be calculated as

$$\frac{1}{t} \sum_{k=1}^t [(\mu_k - \nu_k)^2 - (\bar{\mu} - \bar{\nu})^2] = \frac{1}{t} \sum_{k=1}^t \mu_k^2 + \frac{1}{t} \sum_{k=1}^t \nu_k^2 = 2 \quad (5)$$

which gives the approximate standard deviation as  $\sqrt{2}$ .

We like to identify that the statement of [5, Corollary 1] has an error. It has been written that the mean and standard deviation of the data set  $\mu_k - \nu_k$ ,  $k = 1, \dots, t$  are approximately  $\alpha$ ,  $\sqrt{2}$ . Basically the data set is  $(\mu_k - \nu_k / (1 + \alpha\nu_k))$  as explained in [5, Th. 1], not  $\mu_k - \nu_k$ .

For better understanding we also like to summarize the construction of DCT polynomial which is the main tool for cryptanalysis. We sort the DCT coefficients of an image  $I$  in ascending order in an array  $A$ , storing an index vector using which we can get back to the DCT matrix again from  $A$ . Thus  $A$  is a sorted array of real numbers. We partition the values of  $A$  contiguously in  $q$  different parts  $A_1, A_2, \dots, A_q$ . Note that each of  $A_1, A_2, \dots, A_q$  is also a sorted array. Corresponding to each  $A_i$ , we fit a polynomial of degree  $d_i$  [7, Chs. 2, 3], such that the mean square error is minimized. It is clear that as we increase the degree  $d_i$  of the polynomial, the mean square error is less. However, since the data is monotonically increasing, we get very good approximation using polynomials with moderate degrees. Thus, we get a series of polynomials  $P_1, P_2, \dots, P_q$  corresponding to the arrays  $A_1, A_2, \dots, A_q$ .

Let us now present an experimental result. See the leftmost image of Fig. 1 for the original image  $I$  (Lena image of size  $256 \times 256$ ). Corresponding to the image  $I$  we get a DCT matrix and we sort it in ascending order. We keep the first 50 values (the lower most) as it is. Then we approximate the next 22000, 37436, 5000, 1000 data by four different polynomials of degree 30. The maximum 50 data are also kept as it is. The DCT data pattern (second from left) and the polynomials (third from left) are presented in Fig. 1. In these figures, we have not presented the data for the lower most (negative) 50 values and the uppermost (positive) 50 values since those values are very large and not presenting them in the figure makes the nature of the graphs clearer.

Then we recover an image as follows: i) extracting data from the DCT polynomials, ii) placing them in proper order in the DCT matrix, and then iii) using the inverse DCT transform to get back the image in

spatial domain. The resulting image is displayed in the rightmost side of Fig. 1. Note that the recovered image (top right hand side) has PSNR value as high as 41.3 dB with respect to the original image (left hand side of Fig. 1).

Note that by slightly changing the coefficients of the polynomials, we get a large pool of polynomial sets  $P_1^j, P_2^j, \dots, P_q^j$  for  $j = 1, 2, \dots$ . The DCT matrix recovered from each set of polynomials will produce an image. Depending on the modifications of the polynomial coefficients, the image quality will vary. In fact, it is clear that if we change the coefficients by a very small amount, the image quality will stay good. However, large change in the coefficients of the polynomials will indeed degrade the images.

Now the attack (watermark removal) works as follows. Consider the DCT matrix corresponding to a watermarked image. Changing the polynomial coefficients slightly, one can produce a large set of images with good visual quality (we use PSNR measure) and statistical criteria will be used to find out one or more images where the watermark has been removed.

### B. Exact Attack on Chu's Scheme

Each of the four subimages generated by Chu's watermark embedding algorithm is watermarked using basic CKLS scheme [4]. Naturally, we use the exact algorithm as [5, Algorithm Attack 1] on each of the subimages. We generate four containers  $AT_1, \dots, AT_4$  containing sufficient number of attacked subimages. Thus considering that each subimage is watermarked using CKLS scheme, one can not recover the watermark from these subimages after the attack. This is the first step in our attack.

Now we add some extra steps. We like to guarantee that the inter-correlations between the watermarking signal of subimages with respect to the original image (introduced by Chu's watermarking scheme) are also removed. Thus before merging the attacked subimages to get the attacked image, we need to ensure that the inter-correlations cease to exist.

Based on this assumption and the analysis presented in [5], we can consider that subimages in four different containers  $AT_1, \dots, AT_4$  as the four different watermarked version generated from the same subimage. First we consider the containers  $AT_1, AT_2$ . Let  $I_{d,1}$  (respectively  $I_{d,2}$ ) be the DCT version of an attacked subimage from container 1 (respectively container 2). Let  $s_d^{(k,2)}[i] = (I_{d,1}[i] - I_{d,2}[i]) / \alpha I_{d,2}[i]$ . We will consider these two images if we find that the mean and standard deviation of  $s_d^{(k,2)}$  are close to  $\alpha$  and  $\sqrt{2}$  respectively (for detailed reasoning see Section III-A). Then we also need to consider the subimages from containers  $AT_3$  and  $AT_4$  successively. Ultimately we need four subimages  $I_{d,1}, I_{d,2}, I_{d,3}, I_{d,4}$  from four containers  $AT_1, AT_2, AT_3, AT_4$ , such that the mean and standard deviation of  $s_d^{(k,l)}$  are close to  $\alpha$  and  $\sqrt{2}$  where  $s_d^{(k,l)}[i] = (I_{d,k}[i] - I_{d,l}[i]) / \alpha I_{d,l}[i]$  for any  $k, l$ ,  $1 \leq k \neq l \leq 4$ . This guarantees that the inter-correlation between the watermarking signal of subimages with respect to the

TABLE I  
CORRELATION AND PSNR VALUES OF ATTACKED IMAGES

Image	CKLS Scheme [4]		Chu's Scheme [3]	
	Corr	PSNR(w,a)	Corr	PSNR(w,a)
Lena	0.178	33.88	0.113	31.78
Peppers	0.182	31.93	0.092	35.32
Pentagon	0.169	32.70	0.158	37.30
Goldhill	0.172	31.83	0.124	36.11
Watch	0.210	32.38	0.184	36.02
Elastic	0.232	32.07	0.168	31.53
Fishboat	0.180	32.43	0.137	34.02
F16	0.174	32.93	0.114	35.14
Bridge	0.220	31.30	0.191	34.30
Couple	0.187	33.12	0.158	34.87

original image (introduced by Chu's watermarking scheme) becomes close to zero (similar to the idea explained in Section III-A).

Once we get four such subimages, we merge them to get the attacked image. We like to mention that one does not need to search  $\prod_{i=1}^4 |AT_i|$  many images where the container  $AT_i$  contains  $|AT_i|$  many images. In fact, for all our attacks we just took the first available subimage from each of the containers and found desired property.

#### IV. EXPERIMENTAL RESULTS

Let us now present the experimental results with the same experimental set-up as used in [3]. We use ten different gray scale images of size  $256 \times 256$  for experimental purpose which are available from [8]. Also the values of  $\alpha$ ,  $N$  are 0.1 and 1000 respectively. As  $N = 1000$ , we deal with 1000 DCT coefficients. After sorting them in ascending order, initially we do not disturb the top 50 coefficients. The rest 950 data are approximated using a polynomial (see [5, Sec. II-B]) of degree three. In case of the Lena image, the four polynomials corresponding to four subimages are as follows.

- 1)  $0.000\ 000\ 268\ 425\ 5 \quad x^3 \quad -$   
 $0.000\ 261\ 015\ 871\ 0 \quad x^2 - 0.095\ 909\ 888\ 208\ 8 \quad x \quad +$   
 $22.375\ 208\ 818\ 712\ 0;$
- 2)  $0.000\ 000\ 289\ 733\ 2 \quad x^3 \quad -$   
 $0.000\ 287\ 328\ 283\ 9 \quad x^2 - 0.101\ 747\ 344\ 104\ 2 \quad x \quad -$   
 $21.431\ 624\ 657\ 947\ 3;$
- 3)  $0.000\ 000\ 255\ 841\ 2 \quad x^3 \quad -$   
 $0.000\ 249\ 169\ 240\ 4 \quad x^2 - 0.091\ 527\ 454\ 947\ 9 \quad x \quad +$   
 $22.980\ 877\ 408\ 378\ 8;$
- 4)  $0.000\ 000\ 280\ 743\ 0 \quad x^3 \quad -$   
 $0.000\ 281\ 135\ 160\ 8 \quad x^2 - 0.105\ 813\ 131\ 097\ 0 \quad x \quad +$   
 $20.995\ 772\ 766\ 498\ 5.$

For each of the polynomials, we vary the coefficients of  $x^2$  and  $x$  in the range of  $\pm 2\%$  and  $\pm 3\%$  respectively at a step of 0.2%. Finally we change the top 50 DCT coefficients randomly in the range of  $\pm 2\%$ . In the next step we consider only those subimages whose PSNR values are greater than 30 dB to maintain perceptual quality. From the four containers we choose the right ones (in fact, in our experiments we found the first ones satisfying the requirements) for merging to generate the attacked image.

Experimental results are presented in Table I. Here PSNR(w,a) represents the PSNR value of the attacked image with respect to the watermarked image. From the result in Table I, one can find that 1) the correlations (between embedded and extracted watermark signals) for the attacked images in Chu's scheme are lower than the correlations for the attacked images in CKLS scheme and 2) further the PSNR(w, a) in case of Chu's scheme are higher than that of the CKLS scheme. Thus it can be concluded that with respect to our attack, Chu's modified scheme is weaker than the basic CKLS scheme.

Now we present a justification why the attack is stronger against the Chu's scheme [3] than the basic CKLS scheme [4]. This is because of two stages in the attack which could be mounted due to the fact that the Chu's scheme is oblivious (extraction of watermark does not rely on the original image), unlike CKLS scheme which is nonoblivious (extraction of the watermark does rely on the original image).

- 1) First, we try to get attacked subimages such that the correlation between
  - a) "information in the attacked subimages with respect to the original subimages";
  - b) "information in the watermarked subimages with respect to the original subimages"
 becomes close to zero. This is similar to the basic attack on the CKLS scheme as mentioned in [5].
- 2) Next, we try to remove the correlation among the "information in four subimages with respect to the original image" that constitute the attacked image. Note that during the Chu's watermarking process [3], a negative correlation is induced between two subimages as the value in one of them is increased and the value in the other one is decreased. In this step that correlation is also removed. This is the reason why we get a stronger attack here.

#### ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments that improved both the editorial and technical quality of this paper.

#### REFERENCES

- [1] R. J. Anderson and F. A. P. Petitcolas, "On the limits of steganography," *IEEE J. Select. Areas Commun., Special Issue on Copyright and Privacy Protection*, vol. 16, no. 4, pp. 474–481, May 1998.
- [2] J. Boeuf and J. P. Stern, "An analysis of one of the SDMI candidates," 4th Int. Workshop on Information Hiding, IHW 2001 pp. 395–410, vol. 2137, Lecture Notes in Computer Science, 2001.
- [3] W. C. Chu, "DCT-based image watermarking using sub-sampling," *IEEE Trans. Multimedia*, vol. 5, no. 1, pp. 34–38, Mar. 2003.
- [4] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.
- [5] T. K. Das, S., and S. Maitra, "Cryptanalysis of correlation based watermarking schemes using single watermarked copy," *IEEE Signal Process. Lett.*, vol. 11, no. 4, pp. 446–449, Apr. 2004.
- [6] A. Kerckhoffs, "La Cryptographie Militaire," *J. Sci. Militaires*, ser. 9th, vol. IX, pp. 5–38, Jan. 1883, see also pp. 161–191, Feb. 1883.
- [7] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*. Cambridge, U.K.: Cambridge Univ. Press, 1992.
- [8] [Online]. Available: [http://www.petitcolas.net/fabien/watermarking/image\\_database/](http://www.petitcolas.net/fabien/watermarking/image_database/)