# A successful attack and revision of a chaotic system based fragile watermarking scheme for image tamper detection

Marco Botta
Department of Computer Science
Università degli Studi di Torino
10149, Torino, Italy
marco.botta@unito.it

Davide Cavagnino
Department of Computer Science
Università degli Studi di Torino
10149, Torino, Italy
davide.cavagnino@unito.it

Victor Pomponiu
Department of Computer Science
Università degli Studi di Torino
10149, Torino, Italy
victor.pomponiu@ieee.org

**Abstract.** In this paper we show that a recently proposed fragile watermarking scheme by Rawat et al. does not detect and localize tampering, therefore cannot be used for authentication applications. The problem lies in that the scheme embeds an authentication code into the LSBs of pixels without taking into consideration the image content. To overcome this issue the authentication should be combined with the first seven bits of the image pixels, and in this paper a revision in this sense is proposed.

**Keywords:** fragile watermarking, content authentication, chaotic maps, Arnold cat map, logistic map, security.

## 1. Introduction

These days witnessed the predominance of digital images thanks to the development of affordable digital cameras and high-speed Internet. Nevertheless, concerns with respect to the origin and integrity of digital images have raised and received increasing attention since their content can be easily manipulated and edited.

The study of *fragile image watermarking* aims at addressing these issues by answering questions about the authenticity of digital images, localization of the tampered areas and, in some cases, the capacity to recover them. In order to achieve these goals, a fragile watermark (which cannot survive to any content alterations) is embedded into the image.

In the last years numerous image authentication techniques have been devised in pixel domain [1, 2] and transform domain, e.g., the Karhunen-Loève transform, [3, 4]. Soft computing techniques [5] have been extensively used to improve the efficiency of the watermarking schemes [6-8].

Security of the watermarking schemes [9-12] is another important feature resulting from applications where there exist adversaries willing to bypass watermarking properties such as copyright and integrity protection.

The outline of this paper is as follows: in the next section, we briefly review several concepts of the chaos theory. In section 3, we describe the unsecure fragile watermarking scheme proposed by Rawat et al. [8] while in section 4 we present our attack and other remarks on the scheme. Section 5 concludes this paper.

## 2. Background

Prior to describing the fragile watermarking algorithm introduced by Rawat et al. [8], we firstly present its main feature, the chaotic maps.

### *2.1. Chaotic maps*

Chaotic maps, such as the Arnold cat map and the logistic map, are widely used for encryption and data hiding applications since they provide a high sensitivity to initial conditions [5].

The Arnold cat map is a two-dimensional invertible map which simply illustrates the principles of chaos. For instance, if the Arnold cat map is applied on an image $I$ of size $m \times n$ then its pixel positions are randomized by the following relation:

$$\begin{bmatrix} p_i(x+1) \\ p_i(y+1) \end{bmatrix} = \begin{bmatrix} 1 & \alpha \\ \beta & \alpha\beta+1 \end{bmatrix} \cdot \begin{bmatrix} p_i(x) \\ p_i(y) \end{bmatrix} \bmod n = \Delta \begin{bmatrix} p_i(x) \\ p_i(y) \end{bmatrix} \bmod n \quad (1)$$

where $0 \le i \le n-1$, $p_i(x)$ and $p_i(y)$ denote the coordinates $(x, y)$ of the pixel $p_i$, mod is the modulo operator, $\alpha$ and $\beta$ are two positive integers that characterize the phase space, and $\det(\Delta) = 1$.

Due to the restriction imposed to the parameters $\alpha$ and $\beta$, the Arnold cat map becomes periodic, i.e., if the pixel $p_i$ at location $(x, y)$ returns to its original position after applying $T$ times the Arnold map, then the chaotic map has period $T$. It is worth to point out that the period of the map is closely related to parameters $\alpha$ and $\beta$, and to the size of the image.

Another instance of the chaotic maps is the logistic map, which is obtained by the following relation:

$$p_i(x+1) = \mu p_i(x)(1 - p_i(x)) \quad (2)$$

where $0 < \mu \le 4$. If $3.5699456 < \mu \le 4$, then the logistic map becomes chaotic. In this state, the sequences generated have a high sensitivity to the initial conditions.

Rawat et al. [8] algorithm makes use of the Arnold cat map to shuffle the pixel positions of the host image, and of the logistic map to encrypt the watermark sequence.

## 3. A chaotic system based fragile watermarking scheme

The fragile watermarking scheme proposed by Rawat et al. [8] can be summarized as follows:

E1. By employing the Arnold cat map $k$ times, shuffle the host image $I_h$, of size $m \times n$, to obtain the image $I_s$.

E2. Split each pixel of $I_s$ into 8-bits planes.

E3. By means of a logistic map create a chaotic sequence $C$, of the same size as $I_h$. Further, the values of $C$ are rounded off to obtain an integer chaotic sequence.

E4. Compute the binary chaotic watermark $W_c$ as:

$$W_c = W \oplus C \quad (3)$$

where $W$ represents the original watermark and $\oplus$ denotes the Boolean exclusive-or operation.

E5. Substitute the LSB of each pixel of $I_s$ with the bits of $W_c$.

E6. To obtain the watermarked image $I_w$ apply the Arnold cat map $T - k$ times, where $T$ denotes the period of the chaotic map.

A block diagram illustration of the embedding process is presented in Fig. 1.

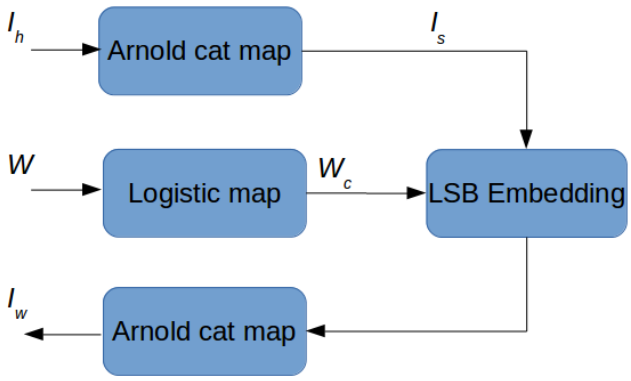


**Fig. 1.** The embedding procedure.

The process of extracting the watermark is as follows:

D1. Shuffle the watermarked image $I_w$, via the Arnold cat map $p$ times to obtain $I_{ws}$.

D2. Split each pixel of $I_{ws}$ into 8-bits planes.

D3. As done in the embedding process, generate the chaotic sequence $C$ and round off each of its values.

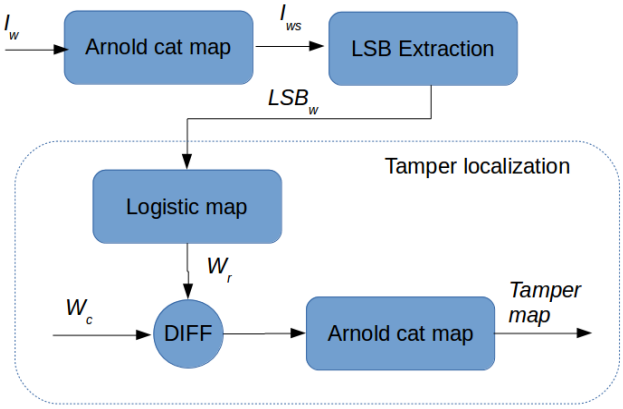


**Fig. 2.** The watermark extraction and verification procedures.

D4. To recover the watermark the XOR operation is applied between the LSBs of $I_{ws}$ and the chaotic sequence $C$.

A block diagram illustration of the extraction and verification procedures is presented in Fig. 2.

To localize the tampered regions, within the watermarked image $I_w$, perform the absolute difference between the original and the extracted watermark, followed by the Arnold cat map $T - k$ times.

## 4. The proposed attack

The security analysis adopted follows a cryptanalytic approach: the watermarking algorithm is assumed to be public while the security relies only on the Arnold cat map and the chaotic sequence which are used to watermark the media contents. The adversary, using the devised attack, will aim to tamper the integrity of the watermarked content without leaving any traces and thus circumventing the watermarking verification procedure.

Before describing our attack, we make some observations on this scheme.

Firstly, note that the Arnold cat map, which is employed in step E1, only changes the pixels *position* (i.e., the $(x, y)$ coordinates) of the host image $I_f$. For instance, the effect of applying the Arnold cat map, with $\alpha = \beta = 1, k = 5$, on a $4 \times 4$ matrix is shown in Fig. 3. Therefore the 8-bits planes of each pixel remain unchanged, even if the pixel's position is shuffled by the chaotic map.

Secondly, the algorithm does not employ any interdependency between the bit planes of the marked pixels.

| | | | |
|---|---|---|---|
| $140^{(1,1)}$ | $79^{(1,2)}$ | $100^{(1,3)}$ | $95^{(1,4)}$ |
| $158^{(2,1)}$ | $88^{(2,2)}$ | $97^{(2,3)}$ | $91^{(2,4)}$ |
| $157^{(3,1)}$ | $126^{(3,2)}$ | $96^{(3,3)}$ | $99^{(3,4)}$ |
| $111^{(4,1)}$ | $154^{(4,2)}$ | $87^{(4,3)}$ | $70^{(4,4)}$ |

Arnold cat map

| | | | |
|---|---|---|---|
| $140^{(1,1)}_{(1,1)}$ | $88^{(1,2)}_{(2,2)}$ | $96^{(1,3)}_{(3,3)}$ | $70^{(1,4)}_{(4,4)}$ |
| $126^{(2,1)}_{(3,2)}$ | $87^{(2,2)}_{(4,3)}$ | $95^{(2,3)}_{(1,4)}$ | $158^{(2,4)}_{(2,1)}$ |
| $100^{(3,1)}_{(1,3)}$ | $91^{(3,2)}_{(2,4)}$ | $157^{(3,3)}_{(3,1)}$ | $154^{(3,4)}_{(4,2)}$ |
| $99^{(4,1)}_{(3,4)}$ | $111^{(4,2)}_{(4,1)}$ | $79^{(4,3)}_{(1,2)}$ | $97^{(4,4)}_{(2,3)}$ |

**Fig. 3.** Illustration of the Arnold cat map ($\alpha = \beta = 1, k = 5$) on a $4 \times 4$ matrix. In the parenthesis are given the pixels positions before (black color) and after (red color) employing the chaotic map.

The key observation of the attack is to compare steps (E1−E5) and (D1−D4) to reveal the fundamental flaw of the algorithm. In step E5, only the LSBs of the shuffled pixels are *changed independently* with those of the chaotic watermark, without considering the image content [9, 11]. In step D4, in order to assess the

integrity of the suspicious image, the LSBs are extracted from the shuffled pixels. Therefore, we can tamper the watermarked image, while preserving the integrity of the watermark, using the following mechanism:

A1. In a matrix $L$, store the LSBs of all the pixels of the watermarked image $I_w$.

A2. Alter the pixels of the watermarked image $I_w$ as desired.

A3. In order to reinsert the watermark, replace the LSB of each pixel with those stored in the matrix $L$.

In other words the attack may be restated as: freely alter all the watermarked image bit-planes apart the one of the LSBs.

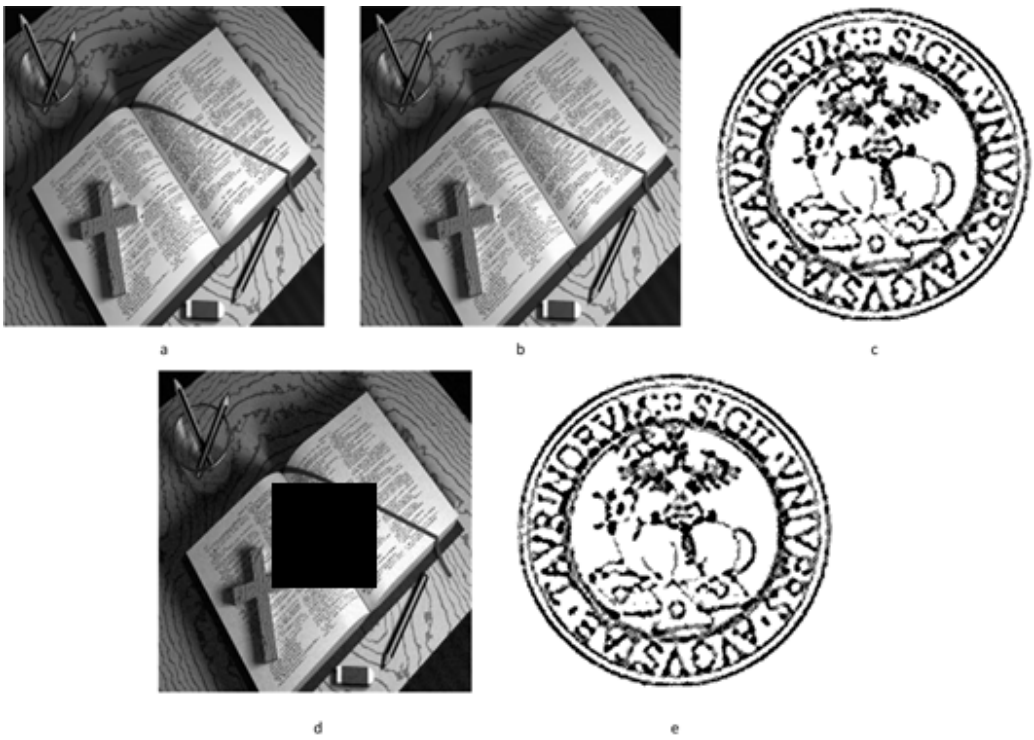

**Fig. 4.** (a) Host image (b) Watermarked image (c) Watermark image (d) Watermarked tampered image (e) Recovered watermark (identical to (c) even if image tampered).

We have verified the attack experimentally: an 8 bpp gray-scale image of size $256 \times 256$ pixels, taken from the OPTIMOL image collection [3], was chosen as the host image. As in [8], the watermark was a binary logo image of size $256 \times 256$ pixels. Furthermore, we set up the parameters of the Arnold cat map and logistic map to the values specified in [8], respectively $\alpha = \beta = 1$, $k = 75$, $\mu = 3.854$ and $p(0) = 0.654$.

The result after applying [8] on the host image is given in Fig. 4b, while Fig. 4d shows the tampered image. The tampered image is obtained by placing a dark quadrant of size $50 \times 50$ pixels. All the LSBs of the pixels within the quadrant are left unchanged while the remaining bits are set to 0. Anyway, the watermark can still be extracted, as reported in Fig. 4e. As an extreme example, we tampered the watermarked image by setting to 0 all bits leaving

unaltered only the LSB of every pixel (see Fig. 5a). Also in this case, the watermark can be extracted as shown in Fig. 5b.

a b

**Fig. 5.** (a) Watermarked tampered image (b) Recovered watermark.

## 5. Proposed revision and discussion

To solve the presented security issue, a part of the authentication should contain information derived from the first seven bit-planes of the image pixels. For example, for each pixel $P_i$ the bit to be substituted to its LSB could be derived by xor-ing the watermark bit in $W_c$ with the bits resulting from a Message Authentication Code (e.g. HMAC) computed from a key $k$ (which may be used also for the Arnold cat map) and a concatenation of:

- the 7 MSBs of $p_i$ (i.e. all the pixel's bits, except the one that will be substituted);
- the coordinates $x$ and $y$ of $p_i$.

During the embedding process the new LSB of the pixel $p_i$ may be expressed as:

$$\text{LSB} = parity\left(W_c(i) \middle| MAC_k\left((p_i \text{ AND } 0\text{xFE}) \mid x \mid y\right)\right), \quad (4)$$

where 0xFE is a binary mask expressed in hexadecimal form, the symbol | means concatenation, $x$ and $y$ denote the pixel position, and $parity(B)$ is a function that returns 0 if the number of ones in the binary representation of $B$ is even, and 1 otherwise.
In this way the authentication bit in the LSB of every pixel depends on the pixel value, on the pixel position and on a secret key, thwarting the presented attack.

## 6. Conclusion

In this paper we proved that the fragile watermarking scheme, based on two chaotic maps, proposed by Rawat et al. [8] is not secure, and therefore cannot be used to assess the authenticity and integrity of digital images.

Moreover, we have proposed a solution to the problem, resulting in a more secure fragile watermarking algorithm.

## References


[1] Bravo-Solorio S, Nandi AK. Secure fragile watermarking method for image authentication with improved tampering localization and self-recovery capabilities. Signal Processing 2011; 91(4): 728-39.

[2] Lin P-Y, Lee J-S, Chang C-C. Protecting the content integrity of digital imagery with fidelity preservation. ACM Trans. on Multimedia Computing, Communications and Applications, 2011; 7(3): 1-20.

[3] Botta M, Cavagnino D, Pomponiu V. KL-F: Karhunen-Loève Based Fragile Watermarking. In: NSS 2011: 5th International Conference on Network and System Security, 2011. p. 65-72.

[4] Ho ATS, Zhu XZ, Shen J, Marziliano P. Fragile watermarking based on encoding of the zeroes of the z-transform. IEEE Transactions on Information Forensics and Security, 2008; 3(3): 567-9.

[5] Hassanien A-E, Abraham A, Kacprzyk J, Peters JF. Computational Intelligence in Multimedia Porcessing: Foundation and Trends. Studies in Computational Intelligence, 2008; 96: 3-49.

[6] Liu P-P, Zhu Z-L, Wang H-X, Yan T-Y. A Novel Image Fragile Watermarking Algorithm Based on Chaotic Map. In: Proc. of the 2008 Congress on Image and Signal Processing, vol. 5, 2008. p. 631-4.

[7] Pan J-S, Abraham A. Special issues in bio-inspired information hiding. Soft Computing, 2009; 13(4): 319-416.

[8] Rawat S, Raman B. A chaotic system based fragile watermarking scheme for image tamper detection. AEU-International Journal of Electronics and Communications, 2011, 65: 840-847.

[9] Fridrich J. Security of fragile authentication watermarks with localization. Proc. of SPIE - The International Society for Opt. Engineering 2002; 4675: 691-00.

[10] Wong P, Memon N. Secret and public key image watermarking schemes for image authentication and ownership verification. IEEE Transactions on Image Processing 2001; 10(10): 1593-01.

[11] Barreto PSLM, Kim HY, Rijmen V. Toward secure public key blockwise fragile authentication watermarking. IEE Proceedings on Vision, Image and Signal Processing 2002; 149(2): 57-62.

[12] Cayre F, Fontaine C, Furon T. Watermarking security: Theory and practice. IEEE Trans. Signal Processing 2005; 53(10): 3976-87.

[13] Li L-J, Wang G, Li F-F. OPTIMOL: automatic Object Picture collecTion via Incremental MOdel Learning. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2008. p. 1-8.