Embedded Signature-based Authentication by Channel Statistics Using Watermarking Techniques[§]

Zhi-Fang Yang and Wen-Hsiang Tsai
Department of Computer and Information Science
National Chiao Tung University
Hsinchu, Taiwan 300, R. O. C.
E-mail: gis82504@cis.nctu.edu.tw; whtsai@cis.nctu.edu.tw

***Abstract*—This study examines the feasibility of using channel statistics to perform embedded signature-based authentication. Doing so is an attempt to view a reference watermark embedded with signature as side information. The reference watermark is extracted to determine channel statistics, based on the Bayes theorem, and used to extract the embedded signature. The reliability of the extracted signature and the uncertainty of the channel status are also measured. Experimental results demonstrate that the reliability and uncertainty measures based on channel statistics are meaningful, and that the embedded signature can survive high-quality JPEG compression and manipulation such as negation.**

I. INTRODUCTION

In the last decade, copyright marking has attracted extensive interest for tackling unauthorized copying and distribution of digital multimedia data [2, 5]. Approaches to watermarking-based authentication fall into two categories — fragile watermarking and robust watermarking [1]. The former is usually applied at the pixel-level and is good for low-level authentication. The latter type of watermark is typically a signature of the host, so that high-level authentication can be easily performed. The reliability of embedding many signature data in hosts is therefore the key to successful authentication. Research into the latter category is still in its infancy [1]. Generally, more data in the signature imply more accuracy. However, more data hidden in the host imply lower robustness. A compromise is to determine the capacity of the host, but the problem remains unsolved because of an inability to model arbitrary manipulations [1, 4].

This study uses a reference watermark embedded, with the signature, in the host, to compute channel statistics. Channel statistics are used to tackle issues concerning reliability of the extracted signature at the receiver site. Doing so is an attempt to provide a mechanism for detecting the reliability of the transmitted signature. This perspective goes beyond using side information to increase the visual quality in the embedding stage or utilize cover data to improve fidelity at the receiver site. The aim is to use channel statistics to realize transmission conditions.

The rest of this paper is organized as follows. Section 2 explains the view of the authentication problem as communications with side information. Section 3 describes the proposed approach. Section 4 offers some experimental results. Conclusions and areas for future research are finally made in Section 5.

II. VIEWING AUTHENTICATION AS COMMUNICATIONS WITH SIDE INFORMATION

According to Fig. 1, a communication system encodes a message using an encoder, sends it through a channel, and decodes it using a decoder. The side information is used at both the encoder and the decoder.

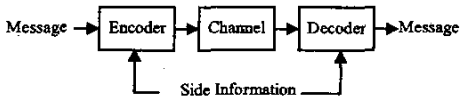

Figure 1. A communication model with side information.

Fig. 2 depicts the proposed approach. The signature is embedded into the host, with a reference watermark. After being transmitted through the channel, the embedded reference watermark is extracted first and used to compute channel statistics. The signature is then detected to authenticate the transmitted host data.

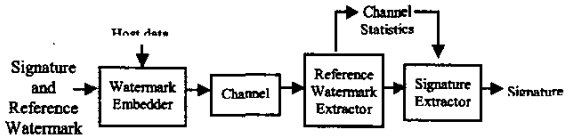

Figure 2. The model of the proposed approach.

In this work, the side information, a reference watermark, is obtained at both the watermark embedder at the sender site

[§] This work was supported by the Ministry of Education under the Project of Excellency No. 89-E-FA04-1-4.

and the watermark extractor at the receiver site. The channel statistics are given to the signature extractor using the side information. The above mapping naturally leads to an example of communication models with side information, Fig. 3.

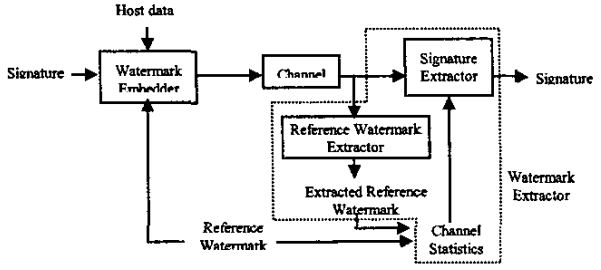

Figure 3. The proposed approach viewed as communications with side information.

Two assumptions are made without a loss of generality. The first is that manipulations of the channel cannot distinguish between the embedded signature and the embedded reference watermark; the second is that the manipulations on the host are locally consistent. Thus, local channel statistics can be analyzed using both a designed reference watermark and the corresponding extracted reference watermark.

III. PROPOSED APPROACH

A. Watermark Design

A binary bit stream is used as the reference watermark. It is used to detect channel statistics based on the two assumptions concerning manipulations stated above. The signature is generated by thresholding the host image according to a predefined threshold. Such a design avoids the difficulty of summarizing information contained in the host data. Other kinds of signatures are also suitable.

The watermark is wholly composed of the signature and the reference watermark. It cannot be so large as to destroy the transparency of the image. Moreover, the number of reference watermark bits cannot be too sparse with respect to the number of signature bits, so the computed channel statistics are reliable in the local area. The signature and the reference watermark are randomly permuted using a permutation key to increase security and distribute the reference watermark bits as uniformly as possible.

B. Watermark Embedding

The embedding method is basically the quantization-based method developed by Kundur and Hatzinakos [3]: the host image is transformed using a wavelet transform, and the watermark is embedded bit by bit in the randomly chosen detail coefficients. Two modifications are made to improve robustness: one is not to embed watermark bits at the highest resolution; the other is to set the quantization parameter as a constant.

Thus, for each watermark bit $w(i)$, a wavelet coefficient, $f_{k,l}(m, n)$, is randomly selected, where $k = h$, v, and d denote "horizontal', "vertical", and "diagonal" detail coefficients, respectively; $l = 1, 2, \ldots, L$ specifies a resolution, and (m, n) represents a spatial location. Each wavelet coefficient can be chosen only once. The embedding rules are as follows:

$$f_{k,l}(m,n) := \begin{cases} f_{k,l}(m,n)+\Delta & \text{if } f_{k,l}(m,n) \le 0 \text{ and } Q(f_{k,l}(m,n)) \ne w(i) \\ f_{k,l}(m,n)-\Delta & \text{if } f_{k,l}(m,n) \ge 0 \text{ and } Q(f_{k,l}(m,n)) \ne w(i) \\ f_{k,l}(m,n) & \text{if } Q(f_{k,l}(m,n)) = w(i) \end{cases}, \quad (1)$$

where := is the assignment operator, Δ is the quantization parameter, and $Q(.)$ is a quantization function defined as follows:

$$Q(f_{k,l}(m,n)) = \begin{cases} 0 & \text{if } \left\lfloor \frac{f_{k,l}(m,n)}{\Delta} \right\rfloor \text{ is an even number} \\ 1 & \text{otherwise} \end{cases}. \quad (2)$$

C. Watermark Extraction and Authentication

The wavelet coefficient selection key is used to locate the positions of the embedded wavelet coefficients, and the permutation key is used to separate the embedded signature bits from the embedded reference watermark bits. Then, the reference watermark bits in an $N \times N$ neighborhood of each signature bit, b, are used to determine conditional probabilities, $p_{i|j}$, (where i is the value of the extracted reference watermark bit and j is the value of the corresponding original reference watermark bit) according to the following equation:

$$p_{i|j} = \frac{\sum_{1 \le p,q \le N} (1-e_{p,q} \oplus i)(1-r_{p,q} \oplus j)}{\sum_{1 \le p,q \le N} (1-r_{p,q} \oplus j)} \quad \text{for } i,j \in \{0,1\}, \quad (3)$$

where $\oplus$ represents the XOR operator, $e_{p,q}$ an extracted reference watermark bit at position (p, q) in the $N \times N$ neighborhood, and $r_{p,q}$ the original reference watermark bit embedded at the same position at the sender site.

P_0 and P_1 are set to 0.5 without a loss of generality. The estimated bit value, b', of the extracted signature bit, b, is then computed using the Bayes theorem:

$$b' = \begin{cases} \arg\max_{j \in \{0,1\}} \frac{p_{b|j}P_j}{\sum_{0 \le k \le 1} p_{b|k}P_k} & \text{for } \max_{j \in \{0,1\}} \frac{p_{b|j}P_j}{\sum_{0 \le k \le 1} p_{b|k}P_k} > T \\ \textit{unreliable} & \text{otherwise} \end{cases}, \quad (4)$$

where T is a predefined threshold with a value of 0.8, established by experiment. The extracted signature bit, b', is classified into the "*reliable*" class if one of the *a posteriori* probabilities pass the test; otherwise,bit b' is classified into the "*unreliable*" class and marked as "*unreliable*". Then, a reliability measure, R, of the extracted signature is defined as follows:

$$R = \frac{\sum_{1 \le i \le L_s} (1 - b_i' \oplus b_i')}{L_s} \quad \text{for } b_i' \text{ is in the "}reliable\text{" class}, \quad (5)$$

where L_s is the length of the signature stream. Furthermore, an entropy measure H is computed as follows:

$$H = \frac{\sum_{1 \le i \le L_s} \sum_{0 \le j \le 1} (-p_{i,j} \log_2 p_{i,j})}{L_s}, \quad (6)$$

where the joint probabilities, $p_{i,j}$, are computed as follows:

$$p_{i,j} = \frac{\sum_{1 \le p,q \le N} (1 - e_{p,q} \oplus i)(1 - r_{p,q} \oplus j)}{\sum_{1 \le p,q \le N} ((1 - r_{p,q} \oplus 0) + (1 - r_{p,q} \oplus 1))} \text{ for } i, j \in \{0,1\} \quad (7)$$

with $e_{p,q}$ and $r_{p,q}$ indicate an extracted reference watermark bit and the corresponding correct bit at position (p, q) in the tested $N \times N$ neighborhood, respectively. The entropy measure, H, is used to represent the uncertainty status of the channel. Finally, the watermarked image is compared to the extracted signature visually, by typical matching techniques, or by ad hoc methods, to classify various manipulations.

IV. RESULTS AND DISCUSSION

Is the channel statistics-based reliability measure R reliable? Two experiments were performed to answer the question. One compared the reliability measure R with the hit ratio. The hit ratio A is computed as follows:

$$A = \frac{\sum_{1 \le i \le L_s} (1 - b_i' \oplus b_i^\circ)}{\sum_{1 \le i \le L_s} b_i'} \quad \text{for } b_i' \text{ is in the "}reliable\text{" class}, \quad (8)$$

where b_i' is a reliable bit and the ith bit of the extracted signature, and b_i° is the ith bit of the original signature stream. This hit ratio A is the percentage of reliable signature bits that are correct. If reliability R is non-zero at a large hit ratio, then the signature bits detected as reliable really include a high fraction of correct bits. The other experiment compared the reliability, R, with the uncertainty, H. When a reliability measure, R, is non-zero at a low entropy value, the detected signature bits recognized as reliable are quite certain.

The 512x512 image, Airplane, is utilized. The signature is generated using a threshold value of 185, and then reduced to a size of 128x128. The Haar wavelet transform is used and resolution level is three. The length of the reference watermark is set to a quarter of the signature size. After the permutations, the watermark is embedded into the transformed images using a quantization parameter, 2^3. The PSNRs calculated from [3] are all between 40 and 50. At the receiver site, the channel statistics are analyzed in a 21×21 neighborhood around each signature bit. Finally, the computed signatures are visually compared to the transmitted images.

A series of JPEG images, with lost data percentages ranging from 3% to 98%, are used. These images undergo various degrees of distortion, as therefore do the embedded signatures. Fig. 4 shows some examples of signatures extracted from these JPEG images.

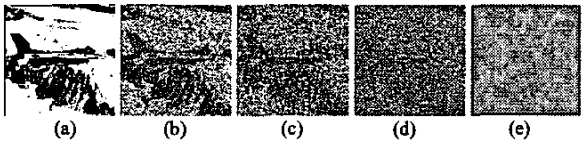

Figure 4. Examples of extracted signatures from JPEG images. The *reliable* extracted signature bits are represented by black and white pixels, which indicate bit values 0 and 1, respectively. The *unreliable* signature bits are represented by gray pixels: (a) original; (b) $R = 0.52$ and $H = 1.24$; (c) $R = 0.43$ and $H = 1.72$; (d) $R = 0.24$ and $H = 1.78$; (e) $R = 0.03$ and $H = 1.87$.

Fig. 5 shows that higher hit ratios are associated with higher reliability measures. In particular, the hit ratios exceed 90% when the reliability measures exceed 0.5, and under such circumstances, the percentage of lost data in the JPEG images can reach 60%. If the lost data percentages rise greatly, then the reliability measures approach zero, meaning that the distortions caused by JPEG compression so severely damaged the images that very few or no extracted signature bits can be considered to be reliable. Hit ratios are smaller than 0.5 or are zero under such circumstances. Notably, values of the reliable signature bits equal randomly chosen values if the corresponding hit ratio is 50%.

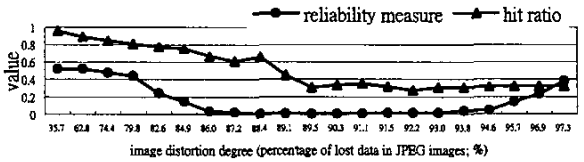

Figure 5. Values of reliability, R, and hit ratio, A, across a series of JPEG images.

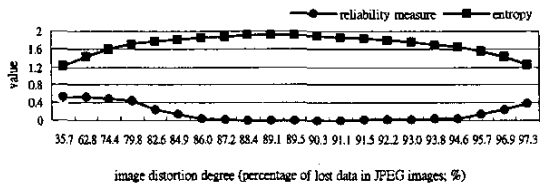

Figure 6. Values of reliability, R, and entropy, H, across a series of JPEG images.

Fig. 6 compares the entropy values of signatures. A higher reliability is associated with lower entropy. That is, the uncertainty in the information is quite low for computed reliable signature bits. Therefore, using the reliability measure based on channel statistics is meaningful. When the reliability measures are about zero, the corresponding entropy values are still rather high. (The maximum entropy value is two in the experiments.) That is, the uncertainty in the information is high under such circumstances. The validity of using the channel statistics-based measure is again confirmed. Moreover, the shape of the entropy value curve is rather symmetrical; that is, entropy values are low at both smaller and larger degrees of distortion. Therefore, the channel statistics are highly certain in these two situations.

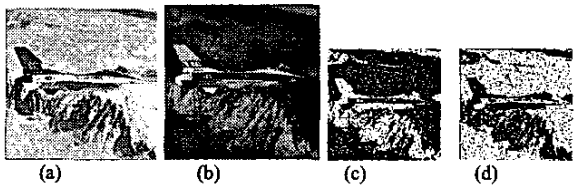

Figure 7. Example of negation attack: (a) the original host image Airplane; (b) negation of the watermarked image; (c) directly extracted signature without applying the proposed mechanism based on channel statistics; (d) signature extracted by the proposed approach.

Fig. 7 presents some experimental results pertaining to authentication. Fig. 7(a) shows the original image — Airplane. After the watermark is embedded, the host image is manipulated using the negation operation. As shown in Fig. 7(b), the airplane seems to be imaged at night. In Fig. 7(c), the extracted signature is obtained by directly retrieving the bit values stored in the signature positions, and the airplane can still be seen to fly at night. However, after the proposed approach based on channel-statistics, is applied to detect the embedded signature, the result is correct: the airplane actually flies under the sun.

V. CONCLUSIONS

Authentication based on embedded signatures is a growing field of research. This study develops a novel approach based on channel statistics, which provide useful information for authentication, including a measure of the reliability of the extracted signature and of the uncertainty of the channel status. The experimental results concerning bit ratios and uncertainties showed that such a method based on channel statistics can be used to describe reasonably the channel status. Future research may be aimed at developing more robust embedding methods, deriving the optimal size of the neighborhood used to authenticate the signature, and generating more representative signatures.

REFERENCES

[1] F. Bartolini, A. Tefas, M. Barni, and I. Pitas, "Image authentication techniques for surveillance applications," *Proceedings of the IEEE*, Vol. 89, No. 10, pp. 1403-1418, Oct. 2001.

[2] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proceedings of the IEEE*, Vol. 87, No. 7, pp. 1079-1107, July. 1999.

[3] D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," *Proceedings of the IEEE*, Vol. 87, No. 7, pp. 1167-1180, July. 1999.

[4] P. Moulin, "The role of information theory in watermarking and its application to image," *Signal Processing*, Vol. 81, No. 6, pp. 1121-1139, June. 2001.

[5] F. A. P. Petitcolas, R.J. Anderson, and M.G. Kuhn, "Information hiding — a survey," *Proceedings of the IEEE*, Vol. 87, No. 7, pp. 1062-1078, July 1999.