A SECURE WATERMARKING TECHNIQUE WITHOUT LOSS OF ROBUSTNESS

Jian Cao[1], *Haodong Li*[2], *Weiqi Luo*[2] ✉, *Jiwu Huang*[3]

[1] Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China
[2] School of Data and Computer Science and Guangdong Key Lab. of Information Security and Technology, Sun Yat-Sen University, Guangzhou, China
[3] College of Information Engineering and Shenzhen Key Laboratory of Media Security, Shenzhen University, Shenzhen, China

ABSTRACT

This paper proposes a novel watermarking technique called random direction embedding (RDE) watermarking. Unlike the conventional watermarking techniques, the watermark energy after the RDE embedding does not focus on a fixed direction, hence the proposed technique is secure against the traditional unauthorized watermark removal attack. The important merit of the RDE watermarking is that it achieves the security without loss of robustness. Compared with the existing secure watermarking techniques, such as natural watermarking (NW), the RDE watermarking achieves significant improvement in terms of robustness. On the other hand, it is also more robust than the spread spectrum watermarking techniques, which have been demonstrated to be very insecure.

Index Terms— Spread spectrum watermarking, robustness, watermarking security

1. INTRODUCTION

Watermarking security has become a major concern in the past several years. The reason is that if the same secret key is reused, the observations of several watermarked signals can probably provide sufficient information for an attacker to estimate the secret key. For example, in methods [1], [2] and [3], it has already shown that the secret carriers of the additive spread spectrum (SS) watermarking [4] and the improved spread spectrum (ISS) watermarking [5] can be estimated by using the blind source separation techniques, such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA) [6]. The attacker can design more powerful robustness attacks if he can obtain sufficient knowledge about the secret key. In [3], for instance, the authors proposed a method to remove the watermark of the SS and the ISS with low distortion by nullifying the watermarked signals projection in the estimated embedding subspace.

Many recent efforts have focused on designing secure SS watermarking techniques. One example is the circular extension of ISS (CW-ISS) [10], which randomly spreads the clusters of ISS on the whole decoding regions such that the projection of watermarked signal onto the embedding subspace after the embedding has an invariant distribution under rotations. The CW-ISS can resist against the unauthorized embedding attack since the attacker can estimate the secret carriers only up to the embedding subspace. The other example is the natural watermarking (NW) [11], which can keep the distribution of host signal's projection onto the embedding subspace invariant during embedding. The NW can resist against both the unauthorized embedding attack and the unauthorized removal attack. However, the robustness of NW is low since there are a lot of watermarked signals close to decoding boundary.

This paper proposes a novel watermarking technique called the random direction embedding (RDE) watermarking, which achieves the watermarking security without loss of robustness. The RDE watermarking uses an orthogonal matrix $\mathbf{Q}$ of size $N_v \times N_v$ as a secret key, in which N_v denotes the length of host signal $\mathbf{x}$. The RDE watermarking embeds the watermark message $m = 1$ by adding the embedding strength to the largest component of $\mathbf{x}_Q$, in which $\mathbf{x}_Q$ denotes the representation of the host signal in the space spanned by the column vectors of $\mathbf{Q}$, i.e., $\mathbf{x}_Q = \mathbf{Q}^T\mathbf{x}$; while it embeds the watermark message $m = -1$ by subtracting the embedding strength from the least component of $\mathbf{x}_Q$. The decoder decodes the watermark message as the sign of the sum of the least component and the largest component of $\mathbf{y}_Q$, in which $\mathbf{y}$ is a potentially distorted version of watermarked signal. The embedding direction of RDE watermarking depends on the host signal and thus is random, achieving the security of the watermarking. Compared with the traditional secure watermarking NW, the RDE watermarking significantly outperforms it in terms of robustness. In some cases, the bit error rate of the RDE is several orders of magnitude better than that of the NW. Compared with the very insecure water-

The work was supported by NSFC (61300208,61272191), the Fok Ying Tung Education Foundation (142003), the Shenzhen Municipal Science and Technology Innovation Council (JCYJ20130329154017293), and Shenzhen R&D Program (GJHZ20140418191518323).

marking SS, the RDE watermarking can even achieve more robustness.

The remainder of the paper is organized as follows. In section 2, we set up the notations used in this paper. In section 3, we present the RDE watermarking technique and derive the analytical expression for its embedding distortion by means of the watermark-to-content ratio. In section 4, we compare the performance of RDE, SS, and NW in terms of robustness and security. Finally, the conclusions are drawn in Section 5.

2. NOTATIONS

In this paper, we use an upper case and bold letter for a matrix, a lower case and bold letter for a vector, and a lower case and italic letter for a scalar variable. We assume that the host signal $\mathbf{x} \in \mathcal{R}^{N_v}$ is Gaussian-distributed with mean vector $\mathbf{0}$ and covariance matrix $\sigma_x^2\mathbf{I}_{N_v}$, i.e., $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \sigma_x^2\mathbf{I}_{N_v})$, where $\mathbf{I}_{N_v}$ denotes the identity matrix of size $N_v \times N_v$.

The watermark embedding is to add the watermark signal $\mathbf{w}$ to the host signal $\mathbf{x}$, resulting in a watermarked signal $\mathbf{s}$, i.e.,

$$\mathbf{s} = \mathbf{x} + \mathbf{w}. \tag{1}$$

We measure the embedding distortion with the watermark-to-content ratio (WCR):

$$\text{WCR}_{[\text{dB}]} = 10\log_{10}\left(\frac{\sigma_w^2}{\sigma_x^2}\right), \tag{2}$$

where σ_w^2 denotes the variance of the watermark signal, and σ_x^2 denotes the variance of the host signal. The robustness attacks are modeled as additive noise, resulting in an attacked signal $\mathbf{y}$, i.e.,

$$\mathbf{y} = \mathbf{s} + \mathbf{n}, \tag{3}$$

where the noise $\mathbf{n}$ is Gaussian-distributed with mean vector $\mathbf{0}$ and covariance matrix $\sigma_n^2\mathbf{I}_{N_v}$, i.e., $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2\mathbf{I}_{N_v})$. As it did in [5], we assess the strength of robustness attacks by means of the signal-to-noise ratio (SNR):

$$\text{SNR}_{[\text{dB}]} = 10\log_{10}\left(\frac{\sigma_x^2}{\sigma_n^2}\right). \tag{4}$$

The performance of decoding is measured by means of the bit error probability P_e:

$$P_e = Pr\{\hat{m} \neq m\}, \tag{5}$$

where $m \in \{-1, 1\}$ denotes the watermark message, and $\hat{m}$ denotes the estimation of m.

3. RANDOM DIRECTION EMBEDDING WATERMARKING

This section presents the RDE watermarking, which uses an orthogonal matrix $\mathbf{Q}$ of size $N_v \times N_v$ as a secret key (N_v denotes the length of host signal $\mathbf{x}$), and randomly selects a vector from N_v column vectors of the orthogonal matrix $\mathbf{Q}$ as embedding direction, such that the decoder can correctly decode the watermark.

3.1. Embedding

The RDE watermarking embeds the watermark message $m = 1$ by adding the embedding strength to the largest component of $\mathbf{x}_Q$, in which $\mathbf{x}_Q$ denotes the representation of the host signal in the space spanned by the column vectors of the orthogonal matrix $\mathbf{Q}$, i.e., $\mathbf{x}_Q = \mathbf{Q}^T\mathbf{x}$; the RDE watermarking embeds the watermark message $m = -1$ by subtracting the embedding strength from the least component of $\mathbf{x}_Q$. Specifically, for a given host signal, let $\mathbf{q}_{\text{max}}$ denote the column vector of the orthogonal matrix $\mathbf{Q}$ with the largest component, i.e., $\forall i \neq \text{max}, \mathbf{x}^T\mathbf{q}_{\text{max}} > \mathbf{x}^T\mathbf{q}_i$, let $\mathbf{q}_{\text{min}}$ denote the column vector of the orthogonal matrix $\mathbf{Q}$ with the least component, i.e., $\forall i \neq \text{min}, \mathbf{x}^T\mathbf{q}_{\text{min}} < \mathbf{x}^T\mathbf{q}_i$. Then, the watermark message $m = 1$ is embedded as follows:

$$\mathbf{s}^T\mathbf{q}_i = \begin{cases} \mathbf{x}^T\mathbf{q}_i + d, & \text{if} \quad i = \text{max} \\ \mathbf{x}^T\mathbf{q}_i, & \text{otherwise;} \end{cases} \tag{6}$$

and the watermark message $m = -1$ is embedded as follows:

$$\mathbf{s}^T\mathbf{q}_i = \begin{cases} \mathbf{x}^T\mathbf{q}_i - d, & \text{if} \quad i = \text{min} \\ \mathbf{x}^T\mathbf{q}_i, & \text{otherwise,} \end{cases} \tag{7}$$

in which $d > 0$ denotes the embedding strength, whose value depends on the expected embedding distortion. Finally, the watermarked signal is given as follows:

$$\mathbf{s} = \sum_{i=1}^{N_v}(\mathbf{s}^T\mathbf{q}_i)\mathbf{q}_i, \tag{8}$$

where $\mathbf{s}^T\mathbf{q}_i$ is given by (6) if the RDE watermarking try to embed the watermark message $m = 1$, and $\mathbf{s}^T\mathbf{q}_i$ is given by (7) if the watermark message to be embedded is $m = -1$.

3.2. Decoding

Let $\mathbf{s}_Q$ be the watermarked signal representation in the space spanned by the column vectors of the orthogonal matrix $\mathbf{Q}$, i.e., $\mathbf{s}_Q = \mathbf{Q}^T\mathbf{s}$. It is easy to show that the sum of the largest component and the least component of $\mathbf{s}_Q$ satisfies the following expression:

$$\text{max}(\mathbf{s}_Q)+\text{min}(\mathbf{s}_Q) = \begin{cases} \text{max}(\mathbf{x}_Q) + \text{min}(\mathbf{x}_Q) + d, & m = +1 \\ \text{max}(\mathbf{x}_Q) + \text{min}(\mathbf{x}_Q) - d, & m = -1 \end{cases} \tag{9}$$

The expectation of the largest component and the least component of $\mathbf{x}_Q$ is equal to zero if the distribution of host signal is symmetric. In this case, the maximum posterior estimation of the watermark message is given as follows:

$$\hat{m} = \text{sign}\left(\text{max}(\mathbf{s}_Q) + \text{min}(\mathbf{s}_Q)\right). \tag{10}$$

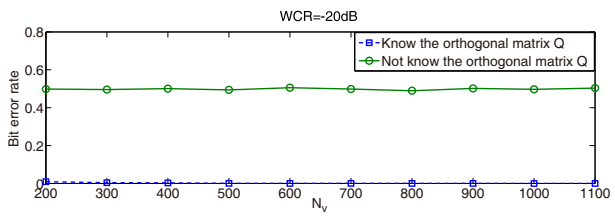

Fig. 1. The bit error probabilities for decoding the watermark with known secret key (i.e., the orthogonal matrix $\mathbf{Q}$) and unknown secret key.

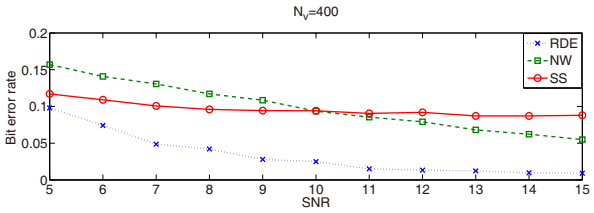

Fig. 2. The comparison of robustness for the RDE, the NW, and the SS in which the WCR of the RDE watermarking and the SS watermarking is set as the same as that of the NW watermarking.

3.3. Embedding Distortion

To compute the WCR value for the RDE embedding, it firstly need to obtain the variance of watermark signal, i.e., the expectation of $\frac{1}{N_v}w^T w$, which is the weighted average of all possible values that the random variable $\frac{1}{N_v}w^T w$ can take on. Please note that the watermark signal can be expressed as follows:

$$\mathbf{w} = \mathbf{s} - \mathbf{x} = \mathbf{Q}(\mathbf{Q}^T\mathbf{s} - \mathbf{Q}^T\mathbf{x}). \tag{11}$$

Hence, from the embedding rule of the RDE embedding, i.e., (6) and (7), we can obtain

$$\frac{1}{N_v}w^T w = \frac{d^2}{N_v}. \tag{12}$$

Finally, the embedding distortion of the RDE embedding is given as follows:

$$\mathrm{WCR}_{[\mathrm{dB}]} = 10\log_{10}\left(\frac{d^2}{N_v\sigma_x^2}\right). \tag{13}$$

Giving a target WCR value, the embedding strength d of the RDE embedding has the following expression:

$$d = \sigma_x\sqrt{N_v 10^{\frac{\mathrm{WCR}}{10}}}.$$

4. SIMULATION RESULTS AND ANALYSIS

In this section, we first test whether the orthogonal matrix $\mathbf{Q}$ can act as a secret key or not. Specifically, we generate the watermark signal with the orthogonal matrix $\mathbf{Q}$, and decode the embedded watermark message from the watermark signal in two ways. One uses a random orthogonal matrix $\mathbf{U}$ of size $N_v \times N_v$ as follows:

$$\hat{m} = \mathrm{sign}\left(\max(\mathbf{s}_U) + \min(\mathbf{s}_U)\right), \tag{14}$$

and the other one uses the orthogonal matrix $\mathbf{Q}$. We independently carry out this experiment 5000 times and show the bit error rates in Fig. 1. As we expect, the orthogonal matrix $\mathbf{Q}$ does act as the secret key of the RDE embedding. If $\mathbf{Q}$ is unavailable, the detection would be failed.

4.1. Robustness Analysis

We decodes the watermark message from the watermarked signal corrupted by additive white Gaussian noise. The experiment is carried out 5000 times independently. The more robust the watermarking algorithm is, the lower the bit error rate will be. Fig. 2 depicts the bit error rate for the RDE watermarking, the NW watermarking, and the SS watermarking, in which the WCR of the three watermarking techniques are set as the same. We can observe that the RDE watermarking is the most robust one among these three watermarking techniques.

4.2. Security Analysis

We measure the security by the bit error rate that the decoder decodes the watermark message after the attacker carries out the unauthorized removal attack, i.e., firstly estimating the embedding direction used at the encoder by the PCA, and then nullifying the watermarked signal's projection onto the estimated projection vector. The more accurate estimation of the embedding direction, the higher the bit error rate is, and thus the more insecure the corresponding watermarking algorithm is. We show the experimental results in Fig. 3. It is observed that the traditional SS watermarking is very insecure. Since its bit error rate is close to that of random guess (i.e., 0.5) when the unauthorized removal attack is performed, meaning that the watermark message would be completely removed. From Fig. 3, we can also observe that the unauthorized removal attack cannot remove the watermark message of the NW and the RDE. Compared with NW, RDE achieves lower bit error rates in most cases, implying that it is more secure.

5. CONCLUDING REMARKS

This paper describes a novel watermarking technique called random direction embedding watermarking. The proposed method uses an orthogonal matrix $\mathbf{Q}$ of size $N_v \times N_v$ as a secret key, in which N_v denotes the length of host signal $\mathbf{x}$. The RDE watermarking embeds the watermark message

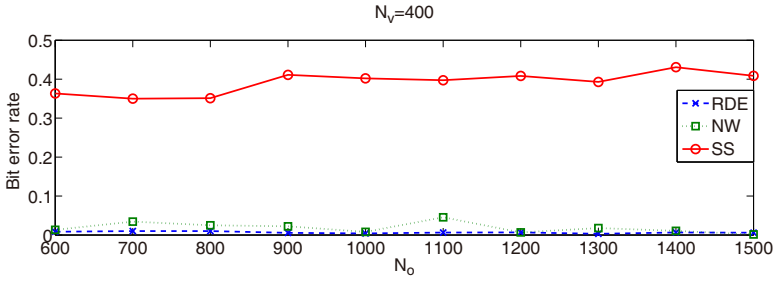

Fig. 3. The comparison of security for the NW, the RDE, and the SS, in which the WCR of the RDE watermarking and the SS watermarking is set as the same as that of the NW watermarking.

$m = 1$ by adding the embedding strength to the largest component of $\mathbf{x}_Q$, in which $\mathbf{x}_Q$ denotes the representation of the host signal in the space spanned by the column vectors of the orthogonal matrix $\mathbf{Q}$, i.e., $\mathbf{x}_Q = \mathbf{Q}^T\mathbf{x}$; the RDE watermarking embeds the watermark message $m = -1$ by subtracting the embedding strength from the least component of $\mathbf{x}_Q$. The RDE watermarking spreads the watermark energy on the whole host space, so as to provide the security against the unauthorized watermark removal attack based on the projection pursuit techniques such as PCA. It is noted that the security of the RDE watermarking is not at the cost of low robustness. The experimental results show that the RDE watermarking gains significant improvement in terms of robustness compared with the existing secure watermarking NW, and it is even more robust than the traditional insecure spread spectrum watermarking. In the future, we will try to provide more theoretical analysis to show the security of the proposed method, and further analyze its robustness against some other attacks including filtering, compressions, scaling, etc.

6. REFERENCES

[1] L. Pérez-Freire and F. Pérez-González, "Spread-spectrum watermarking security," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 1, pp. 2–24, 2009.

[2] F. Cayre and P. Bas, "Kerckhoffs-based embedding security classes for WOA data hiding," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 1–15, Mar. 2008.

[3] F. Cayre, C. Fontaine, and T. Furon, "Watermarking security: theory and practice," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3976–3987, Oct. 2005.

[4] I. J. Cox, J. Kilian, F. 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] H. S. Malvar and D. A. F. Florencio, "Improved spread spectrum: a new modulation technique for robust watermarking," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 898–905, Apr. 2003.

[6] A. Hyv̈arinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. on Neural Networks* vol.10 no.3 pp. 626-634, 1999.

[7] L. Pérez-Freire, F. Pérez-González, T. Furon, P. Comesana, "Security of lattice based data hiding against the known message attack," *IEEE Trans. on Information Forensics and Security* vol.1, no. 4 pp.421–439, 2006

[8] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1423–1443, May. 2001.

[9] L. Pérez-Freire, P. Comesaña, and F. Pérez-González, "Information-theoretic analysis of security in side-informed data hiding," in *Proc. 7th Information Hiding Workshop (IH2005)*. Lectures Notes in Computer Science, Springer-Verlag, vol. 3727, Barcelona, Spain, Jun. 2005, pp. 131–145.

[10] P. Bas and F. Cayre, "Achieving subspace or key security for WOA using natural or circular watermarking," in *Proc. 8th ACM Workshop on Multimedia and Security (MM&Sec'2006)*, Geneva, Switzerland, Sept. 2006, pp. 80–88.

[11] P. Bas and F. Cayre, "Natural watermarking: a secure spread spectrum technique for WOA," in *Proc. 8th Information Hiding Workshop (IH2006)*, Lecture Notes in Computer Science, Springer-Verlag, vol. 4437, Alexandria, VA, USA, Jul. 2006, pp. 1–14.

[12] I. Cox, M. Miller, and A. McKellips, "Watermarking as communications with side information," *Proc. IEEE*, vol. 87, pp. 1127–1141, Jul. 1999.