Journal of Theoretical and Applied Information Technology
20th April 2013. Vol. 50 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN IMPROVED ALGORITHM ON ASYMMETRIC TEXT WATERMARKING BASED ON SVM

YUE FUQIANG[1] , CONG JICHENG[2]

[1.] School of Automotive and Electronic Engineering, Xichang College, Xichang, China

[2.] Huanghuai University, Zhumadian, China

ABSTRACT

The development of digital technology and the Internet has facilitated the spread of various forms of digital works. At the same time, the characteristics that digital works is extremely vulnerable to reproduce ideally will likely be used by pirates. Digital watermark technique is one of effective means of copyright protection recently. On the basis of a detailed analysis of natural language watermarking, an asymmetric text digital watermarking algorithm based on Naturallanguage is proposed. First, digital watermarking is produced using asymmetric algorithm. Second, the index sequence of paragraphs and sentences are permuted using DES and Queue Scrambling. Third, watermarking is embedded by sentence transformation, and during the embedding, Multi-degree relativity is introduced. Last, analysis and provident show that the schemes have higher security, they can prevent from forging the watermarks and trace piracy; copyright certification processes have higher efficiency. The watermarked documents have good robustness.

Keywords: *Natural language; Asymmetric; Queue Scrambling; Multi-degree relativity; Text; Digital Watermarking*

1 INTRODUCTION

In today's rapid development of the information society, the network has become the "fifth media" after traditional media, all kinds of digital information in the network requires a lot of issuance, and transfermation. Network transmission of information brings convenience to people at the same time, the digital file or works copyright protection issues also face greater challenges. Digital watermarking technology has been widely considered to be one of the effective means to realize copyright protection. Digital watermarking technology according to the carrier types can be divided into image digital watermarking, video watermarking, digital audio watermarking, digital watermarking and relational database digital watermarking technology. So far, the digital watermarking research mainly concentrates on the image, video and audio, image digital watermarking, study is the most mature technology, but less research on text digital watermarking. Because of the redundancy of text of less information, it leads to a very limited hiding information, so the text watermarking research is mainly done by changing the text format to embed digital watermark, such as Brassily and Maxemchunk et al proposed by modifying the document word spacing (or row spacing, the character feature) to embed watermark information, this kind of algorithm is completely dependent on the text format, the watermark information is embedded in the text itself, so the robustness is not strong, weak robustness. In order to solve this problem, Purdue professor Atallah of the University of the natural language text watermarking technology, this technology does not change the text meaning under the premise, through the adjustment of sentence structure of the watermark is embedded into the original text in an information hiding technology. This paper on the basis of the natural language watermarking algorithm is put forward an improved asymmetric digital watermarking algorithm, which introduces the asymmetric encryption, transformation of sentence pattern, redundant embedding method and mechanism, through the analysis and the proof shows that the algorithm has higher robustness and better robustness.

2 ASYMMETRIC DIGITAL TEXT WATERMARKING ALGORITHM BASED ON NATURAL LANGUAGE

2.1 Digital watermarking system model

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A complete digital watermarking system consists of two basic modules: watermark embedding and watermark detection and extraction module. Watermark embedding module is responsible for the watermark signal is added the original data, as shown in figure 1. Watermark detection and extraction module is used to judge whether a data containing the specified watermark or the watermark is extracted, as shown in figure 2. The dotted line indicates in certain circumstances (such as blind watermarking system), the original data is not required.

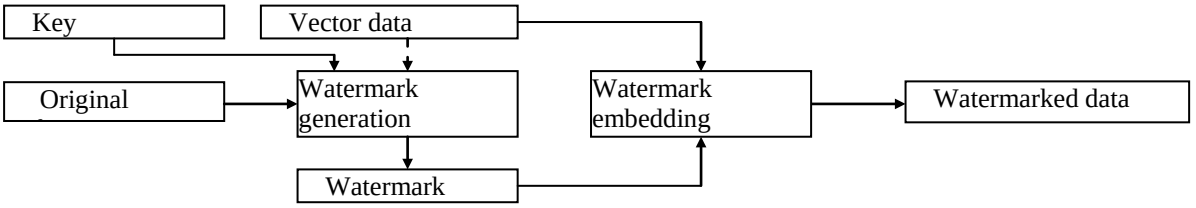

Fig 1 Watermark Embedding Process

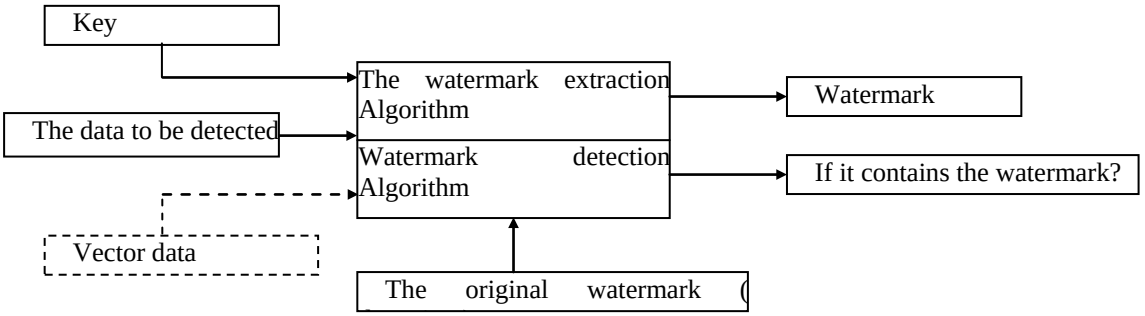

Fig 2 Watermark Detection And Extraction Process

Digital watermarking technology is actually based on the watermark embedding information media analysis, pretreatment, information embedding point selection, embedded design, embedded modulation control of several key technologies of reasonable optimization, seeking to satisfy imperceptibility, security and reliability, robustness and so under the restrained conditions of quasi optimization design problems. As an important part of the watermark information -- key, each design scheme is an important feature of the host.

2.2 Embedded watermark generation

The specific steps are as follows:

1. Public watermark generation

(1) The original transformed into binary form of ASCII code, as primitive watermark information, length l.

(2) The public key $key_p = HASH(T)$, where T is the original, $HASH(\cdot)$ said a single hash function. With key_p as a pseudo-random number generator produces seeds, the length of the $n = l$ pseudo random sequence, and mapped into bipolar two value sequence

$$p_1 = \{p_1(i) | p_1(i) \in \{-1,1\},\ 1 \le i \le n\} \quad (1)$$

key_p As the public key watermarking detection, have four advantages: first, it is the original text of the Hash value,

The public key and the original text related to prevent forgery, public key, different works by different public key watermarking open detection; second, due to the one-way hash function, the attacker to the public gets a fake original text in the calculation is not feasible; third, the public key is only a pseudo random number generator seed transmission of public information, required less; fourth, user authorized given original, can use public key authentication is not original, which can prevent embedding person cheating.

(3) The original watermarking information sequences b_1 and p_1 sequences are multiplied to obtain public watermarking sequence

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$w_p = \{w_p(i) \mid w_p(i) = b_1(i) \cdot p_1(i), 1 \le i \le n\} \quad (2)$$

2. The secret watermark generation

(1) the value of two pseudo random sequence p_1 into $M \times N$ matrix A, where $M \times N = n$ using the private key key_s_1 on queue scrambling matrix B, then B dimensionality reduction for one-dimensional sequence, the two values are the pseudo random sequence p_2

(2) A primitive watermark information organized into a two-dimensional matrix of b_1 $M \times N$ R, $M \times N = n$ key_s_2, using the private key to queue scrambling, scrambling the watermark information obtained after Rt.

(3) The scrambled watermark Rt dimensionality reduction, and mapping for bipolar two value sequence of length $l = M \times N$, get a sequence of one-dimensional b_2

$$b_2 = \{b_2(i) \mid 1 \le i \le l\} \quad (3)$$

(4) The scrambled watermark information sequence pseudo random sequence of b_2 and p_2 are multiplied to obtain secret after the watermark sequence spread spectrum

$$w_s = \{w_s(i) \mid w_s(i) = b_2(i) \cdot p_2(i), 1 \le i \le n\} \quad (4)$$

From the above process, it can be seen, the pseudo random sequence p_2 is composed of p_1 queue scrambling to get, because the queue scrambling key space is infinite, without knowing the key cases, launched by p_2 p_1 in the calculation is very difficult, and can effectively resist the brute-force attack. And the secret watermark before expanding by private key $key_s\_2(L_2, I_2, J_2, count_2)$ queue scrambling encryption, it reinforces the security of system, to further ensure by public watermarking does not infer secret watermark. In addition, can also provide the same works of different versions of select different scrambling parameter, generates a different secret watermark, and these different versions of the same work by the same key detection.

2.3 Watermark embedding

In the watermark embedding, combined with encryption, the carrier of watermark embedded position sequence preprocessing.

Definition 1 the $T = \{p_1, p_2, ..., p_x\}$ where p_i is the text of article i $|pi|$ for p_i in the number of sentences.

Any segment of a definition sentence index sequence for $\{s_1, s_2, ..., s_y\}$, $|si|$ for s_i length of sentences, i.e. the number of Chinese characters in the sentence, denoted as $bi = |si|$ $\{b_{i1}, b_{i2}, ..., b_{ik}\}$ b_i binary representation, which is the lowest.

Definition 3 the T section and sentence combination, recorded as: $T = \{p_1, p_2, ..., p_y\} = \{\{s_{1+1}, ..., s_{1+y1}\}, ..., \{s_{x+1}, ..., s_{x+yx}\}\}$ p_i which is the text of article i s_{i+y} is the II Duan Wenben y sentences.

Definition of 4 order segment (or sentence) of the original primer sequence for $\{1,2,...,n\}$ (n is a text in the middle number or every paragraph sentence number), according to section number contained in the sentence (or sentence contains the numbers of Chinese characters) in ascending order of the new sequence is the paragraph (or sentence).

Definition 5 w_s as the embedded watermark information is recorded as: $w_s = \{w_1, w_2, ..., w_n\}$ where w_i is the first i bits of watermark.

This is to hide the watermark embedded sentences in sequence, the attacker does not know the key of K, is unable to obtain the watermark embedded sentence sequence, so it is not correctly extracted watermark. On the other hand, scrambling the redundant embedding position interspersed throughout the text and not concentrated in a text block, thereby enhancing the security of watermark. Note segment and the sentence initial index sequence through the DES encryption of the sequence is not necessarily the initial sequence of Full Permutation, so this paper uses the data structure of the hash table to encrypted sequence processing.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

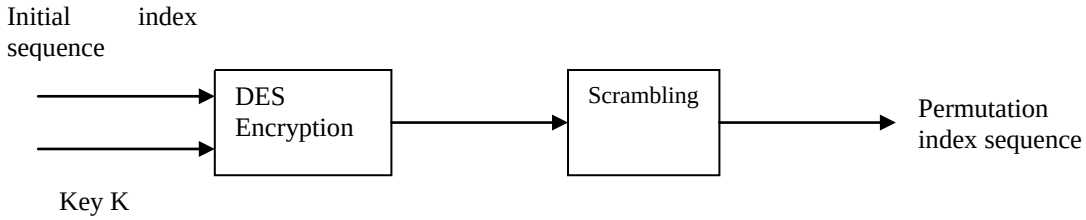

Fig 3 Encryption Scrambling Process

In this module, according to the first paragraph the sentence number sorted in ascending order as the period of initial index sequence, and then encryption, scrambling to get a new index series. Wherein, encryption using DES encryption, scrambling algorithm based on queue transform scrambling algorithm, this algorithm has the fast inverse scrambling, convenient and several transformation can be achieved very good scrambling effect, and can be combined with encryption technology advantages, have very high security.

Hypothesis 1: the input for the period of initial index sequence G= {g1, g2,…, gy} and key K, then the output: scrambling after the index sequence of G′={g1′,g2′,…,gy′}

Hypothesis 2: the input to a sentence initial index sequence is g={s1,s2,… , sy} then the output: scrambled sentence index sequence of g′={s1′,s2′,…,sy′}

2 watermarks are embedded module

This module introduces the multiple correlation of embedded. Each watermark bit is the embedded position not only with the sentence number, but with a sentence number and a watermark bit embedding location. Each watermark bit is embedded, and the method is: from the new period of index sequence in order to select a segment, and then according to the section of the index number and sentence number on the calculated to the embedded sentence index, from which the new sentence index selection sequences in the sentence, the sentence length, and the implementation of the sentence pattern transformation, by changing the length of the sentences the watermark information embedded therein.

Watermark embedding algorithm 2 is as follows:

The original input: text T, scrambling after the segment index sequence G'', scrambled paragraphs sentence index sequence gi' and the secret watermark sequence Ws; ws={w1, w2, ... , wn} where wi is the first i bits of watermark.

Output: Embedded watermark text T(w)

Begin counter=0;

for i=1;i<=n;i++ ;

do

j=(i+counter) mod |pi| ; |pi|pi sentences in a number

bj=|si+j| ; Calculation of section i; the j sentence length bj

bj,1=wi ;

Watermark bit change bj low

The application of natural language text watermarking algorithm is proposed by sentence pattern transformation on the sentence pattern transformation;

counter=counter+j mod |pi|;

end

3. A WATERMARK EXTRACTION AND DETECTION

3.1 Watermark extraction

Extraction algorithm 3 is equivalent to the reverse process of embedding algorithm, algorithm is as follows:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Input: watermark text T (W) and key K

Output: we extracted watermark

Begin

The watermark text T (W) were identical to the scrambling;

counter=0;

for i=1;i<=n;i++

do

j=(i+counter) mod |pi|;

bj=|si+j|

wi =bj,1

counter=counter+j mod |pi|;

end

we={w1,w2,…， wi,…wn} ； The extracted watermarkWe

Key of key_p as an artifact of the seed of the random number generator to generate pseudo-random sequence PP, the watermark text $T(w)$ ASCII code sequences b_1 and p_1 sequences extracted watermarking

$$w_p = \left\{ w_p(i) \middle| \; w_p(i) = b_1(i) \cdot p_1(i), 1 \le i \le n \right\}$$

3.2 Watermark detection

Anyone can use the public key to generate pseudo random sequence key_p p_1 detected data open detection, extraction of watermarking w_p ; at the same time the copyright owner can also use private key $key_s\_1(L_1, I_1, J_1, count_1)$ for replacement by p_1 a pseudo random sequence w_s , which treats the data key detection, extract the secret watermark w_s of course, can also use w_e symmetry detection.

4. SAFETY PERFORMANCE ANALYSIS

4.1 Exhaustive attack

For the asymmetric watermarking schemes, the embedded watermark including public watermarking and secret watermark in two parts, public watermarking needs to open to public inspection, so the security of this scheme depends on whether can infer the secret watermark by public watermarking. While the secret watermark is the first by the original watermarking information in accordance with the private key key_s_2 queue scrambling, and then by a pseudo random sequence of p_2 modulation. Pseudo random sequence p_2 is composed of pseudo random sequence p_1 in accordance with the private key key_s_1 the queue scrambling to get.

From this analysis, secret watermark is a public watermarking through two queue scrambling transformation is obtained, and the queue scrambling itself includes four parameters, namely $K(L, I, J, count)$, which transforms the way L is 8, $count$ is not the only number of iterations, and the (I, J) can be set as a reference point in a cohort of an arbitrary point, the parameter selection of large space, can be said to be infinite, so by the public watermarking through exhaustive attack launched secret watermark, in the calculation is not feasible

In addition, the public key and the original text related to prevent forgery, public key, different works by different public key watermarking open detection; second, due to the one-way hash function, the attacker to the public get a fake original text in the calculation is not feasible; third, by the public through an exhaustive infer private key in calculation is not feasible. It ensures that the watermarking scheme has very high security.

4.2The revised text attack

The programme is based on a natural language sentence pattern transformation, the watermark is embedded in the structure of the text and not in text format (word spacing, row spacing, and characters), and so this attack on the watermarking scheme will not be affected.

4.3Change the text structure attack

In general, change the text structure is mainly the addition or deletion of one or several sentences, paragraphs or sentences into. The use of DES encryption scrambling algorithm is to segment index sequence scrambling before, according to the first paragraph sentence numbers in ascending order undertook sort and record. So even if an attacker disrupting section sequence, or by adding or deleting a sentence that the changes in the number of segments of the sentence, the watermark

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

extraction can recover the initial segment of the sequence and the number of index. At the same time, the scheme introduces embedded multiple correlation. But each watermark bit is embedded position not only with the sentence number, but with a sentence number and a watermark bit embedding location. Moreover, the watermark is scattered throughout the text, the attacker may result from deletions of a sentence and the destruction of a watermark bit, but only to remove all the hidden watermark bit sentences that may lead to the entire text of utter devastation, the text is not available. So the probability to attack and destroy the whole watermark is very small.

5. CONCLUSIONS

This paper is based on the replacement, asymmetric watermarking idea, proposed one kind based on the queue scrambling asymmetric watermarking scheme. The program uses the DES encryption algorithm, the infinite key space queue scrambling transformation on the section and the index of the sentence sequence scrambling processing, embedded with multilevel association embedding mechanism to improve watermark scheme; the original text of the Hash value as the watermark detection key, public key and the original text, can prevent a forged a public key and a watermark; and the public key can be of the same work in different versions of the generation of different embedding watermark, for piracy tracing, and these different versions from the same public key public key detection; detection performance is good, the watermark to common attacks has good robustness.

REFERENCE

[1] Zou Xiaoxiang, Li Jintao, Peng Cong. Asymmetric digital watermarking technology research [J]. computer engineering and application of.2002, 16:7-10, 54

[2] Wang Bingxi, Chen Qi, Deng Fengsen. Digital watermarking technology [M]. Xi'an: Xi'an Electronic and Science University press, 2003:6-17102-103.

[3] Yan Weiqi, Zou Jiancheng, Qi Dongxu. A digital image scrambling based on DES new method. Journal of North China University of Technology, 2002,3, 14 (1): 1-7.

[4] Huang Zhenhua, Kou Weidong, Zhang Jun. Based on the feature vectors and the equations solution of the asymmetric watermarking algorithm. Computer engineering and applications 2006 (32): 104-105114.

[5] Guo Xiping, Zhang Wen, horse Shengfeng, Wang Wei a text watermarking algorithm based on sentence length 2007,43 (32) Computer Engineering and Applications computer engineering and Applications

[6] Zhang Yu, Liu Ting, Chen Yi identity natural language watermarking [J], 2005,19 (1) Journal of Chinese information processing

[7] Geying people, Zheng Gang use of a character changes in the text digital watermarking method [J] micro computer application, 2005,21 (3) :36-39.

[8] Shu Minglei Fang Wangsheng, based on changing the characters of fragile text watermarking Journal of Jiangxi University of Science and Technology, 2006,12:28-31.

[9] M.P.Vani, Computer Aided Interactive Process of Teaching Statistics Methodology - II, IEIT Journal of Adaptive & Dynamic Computing, 2011(3), Jul 2011, pp:18-21. DOI=10.5813/www.ieit-web.org/IJADC/2011.3.4

[10] Zhou Y.Y, Measuring Service Quality at University's Libraries, IEIT Journal of Adaptive & Dynamic Computing, 2011(3), Jul 2011, pp:22-25. DOI=10.5813/www.ieit-web.org/IJADC/2011.3.5