Sensors & Transducers, Vol. 157, Issue 10, October 2013, pp. 331-337

www.sensorsportal.com

Sensors & Transducers

http://www.sensorsportal.com

Research of Multiple Text Watermarks Technique in Electric Power System Texts

[1] Xiao-Xi XING, [1] Qing CHEN, [2] Lian-Xi FU

[1] School of Optical-Electrical and Computer Engineering,
University of Shanghai for Science and Technology,
Shanghai, China, 200093

[2] State Grid Jibei Electric Power Co. Ltd., Beijing, China, 102401

[1] Tel: 18221207516

[1] E-mail: wealthxiaoxi@163.com

Received: 21 July 2013 2013 /Accepted: 25 October 2013 /Published: 31 October 2013

Abstract: Aiming at the reliable transmission security problems of electric power system texts, multiple text watermarking algorithm is first applied into protecting it. Take protecting the transmission of electricity price in the in electric power system text and authenticating the content from the sender for example, and robustness watermarking algorithm is employed to transmit the electricity price with covert communication which makes it possible to resist different kinds of attacks. The semi-fragile watermark is utilized to confirm the identity of the sender which can show whether the text is stolen or misused, and find where the contents are tampered. The experiments show that multiple text watermarking algorithm can protect the data effectively aiming at different purpose which achieves its application value. Those also show the validity of multiple text watermarking technique used in the reliable transmission of electric power system texts and that this technique has a vast application prospect. *Copyright © 2013 IFSA.*

Keywords: Text watermarking technique, Multiple watermarking, Power system, Content authentication, Semi-fragile watermarking, Robustness watermarking.

1. Introduction

With the popularity of modern technology in the power system and the development of multimedia technology, information security has become an urgent problem to be solved in the power system. Digital watermarking technology as a powerful means to solve the security of multimedia information has been gradually applied in electric power system. Image [1, 2] and video [3] watermarking technology both can be used for covert communication of important information about electricity bargain price and so on to improve the security of information transmission. Audio watermarking technology [4] can protect the copyright of audio files to prevent its theft or misuse. Mesh watermarking technology [5, 6] can protect the copyright of mesh model with proprietary intellectual property rights. Literature [7] proposes an idea that text watermarking technology can be applied into reliable transmission of power system texts, but did not practice it.

Currently, text watermarking technologies can be mainly categorized two groups: the format and content. Maxemchuk and Brassil [8] proposed an algorithm which changes line spacing, the character spacing and character [9] to embed watermarking information slightly, changes of which cannot be

identified by the naked eye. These three methods have high imperceptibility, but cannot guarantee the robustness and capacity is the best. Though the robustness of text watermarking algorithm based on the content is very strong, it will cause the ambiguity of natural language [10], and its imperceptibility is not high, and it is easy to cause the opponent's attack. These algorithms can't cater to the characteristics of strong robustness, large capacity and high imperceptibility. Therefore, the method of embedding multiple watermarks is employed to meet the needs of different performances.

Aiming at improving the security of convert communication of important information in electric power system and confirming the source of the files by using content authentication method, a multiple watermarks algorithm applied into the reliable transmission of electric power system text is proposed. The algorithm based on changing character spacing which has a strong robustness is used to hide electricity bargain price, and semi-fragile watermarking embedding algorithm based on changing color RGB value of characters which has a high imperceptibility is also employed to confirm the identity of the sender. The proposed multiple watermarking algorithm is tested with embedding, extracting watermarking information and attack experiments respectively. And the experimental results show that the system is highly feasible, can improve the security of covert communication, and find precisely where the content is tampered when the text suffers from malicious attacks. It is suitable for reliable transmission of electric power system texts and content authentication.

2. Text Watermarking Technology

Digital watermarking technology is an effective security technology used for copyright protection, covert communication, content authentication, quality detection of multimedia information. Multimedia redundant space is utilized to embed watermarking information to achieve the purpose of security protection

2.1. Text Watermarking Algorithm

Text watermarking technology, based on human visual system (HVS), by fine-tuning the text format or content [11, 12], embeds the watermarking information. Watermarking algorithm based on fine-tuning mainly the text format fine-tunings the line spacing, characters spacing and character properties to embed watermarking information. This kind of watermarking algorithm has good robustness, and can resist certain attack. It also has high imperceptibility, large capacity. Watermarking algorithm based on changing the content of a text to embed the watermarking information mainly rely on the synonyms and equivalent sentence replacement to achieve the watermarking embedding, The robustness of this algorithm is higher than that based on changing format, however, due to the changes of the content, its imperceptibility is poorer and it is less capacity.

2.2. Text Watermarking Framework

Digital watermarking system is mainly composed of four parts: watermarking generating Gen, watermarking embedding Emb, watermarking attack Att, watermarking detection Det. By detecting the text after the attack, the watermarking information can be extracted correctly to determine the robustness of watermarking system. We represent the watermarking attack model as [13] WAM,

$$WAM = <\{C, M, W, K, C_W, W'\}, \{A, C_W', WatAlg, WatDet, WatPar\}, \{Gen, Emb, Det, Att\}>$$

In order to facilitate the understanding of this model, some variables are described as follows:

C: watermarking carrier;

C_W: watermarked carrier;

M: watermarking message;

W: watermarking;

W': extracted watermarking;

A: attack;

C_W': attacked watermarked carrier

$$C_W' \leftarrow C_W + A \tag{1}$$

K: secret key;

K^{gen}, K^{emb}, K^{det} represent the set of secret keys used in the process of watermarking generating, embedding and detection, respectively.

WatAlg denotes watermarking algorithm; *WatDet* denotes watermarking detector; *WatPar* denotes watermarking parameter.

In a set, the small letter represents an element of the set of the corresponding capital letter.

$$Gen: W \leftarrow Gen([C], M, [K^{gen}]) \tag{2}$$

$$Emb: C_w \leftarrow Emb(C, W, K^{emb}) \tag{3}$$

$$Att: A \leftarrow Att([WatAlg], [WatDet], WatPar) \tag{4}$$

$$Det: ([W'], [yes, no]) \leftarrow Det(C_W', K^{det}, [W]) \tag{5}$$

3. Solutions Based on Text Digital Watermarking

In the reliable transmission of texts in electric power system, the imperceptibility of the algorithm is required to be very high, the text cannot be changed with perceptible changes, which may cause the attacker's interest, and the content of the text cannot

generate ambiguity after being changed. Therefore, algorithm based on content is not suitable and this paper uses the algorithm based on fine-tuning text format to embed the watermarking information. In order to protect the texts in the security of reliable transmission in the electric power system and increase the trust between users, semi-fragile watermarking and robust watermarking are embedded at the same time. The semi-fragile watermarking having high imperceptibility and high security which can resist some attacks can be used to authenticate contents after texts are received, distinguish the identity of the sender, embed the personal information, send date and so on, and find where the contents are tampered. Robust watermarking has a high ability to resist attacks which can be used for covert communication of important information, and requires a large watermarking capacity. It employs the method of loop embedding watermarking information, which ensures the text is, as long as there is at least one watermarking information available, still can be extracted correctly. Therefore, the semi-fragile watermarking embedding algorithm based on changing character colors is used for content authentication, while the robust watermarking algorithm based on character spacing is used for hiding important information.

4. The Text Watermarking Algorithm

As described in literature [14], the RGB value combinations varying from (0,0,0) to (60,60,60) is very close to the Word default black, the naked eye cannot distinguish changes. Maxemchuk [15] points out that horizontal word displacements of 1/150 in and less readily go unnoticed, vertical line displacements of 1/300 in and less go unnoticed by readers.

4.1. Based on Change the Font Color RGB Values's Semi-Fragile Watermarking Algorithm

The purpose of this algorithm is to authenticate the content of the transmitted texts. Therefore, semi-fragile watermarking can be employed. It is required to have the following features: 1) Imperceptibility. After embedding watermarking information, the text does not produce visible changes, because capacity and imperceptibility are mutually contradictory and the increase in capacity will result in the decline in imperceptibility. Therefore, this algorithm sacrifices part of capacity to ensure imperceptibility; 2) Detecting and signing tampered contents. It is an important property of the semi-fragile watermarking. The tampered content must be precisely positioned and displayed; 3) Security. The key space should be large enough to resist exhaustive attack method.

According to the test of human cone's sensitivity to color shows that the sensitivity of the naked eye to different RGB components of color is different, and the human eye is most sensitive to red, followed by green and blue. Therefore, the algorithm changes lower two bits of the font color of G components and lower two bits of B components to achieve the watermarking embedding. By using this algorithm, though each two characters can be embedded with a watermarking byte without calculating error correction coding at the expense of a certain capacity, the imperceptibility can be better protected.

4.1.1. Pretreatment

In order to enhance the robustness and security of the watermarking information, it needs encrypting and error correction coding before being embedded. Take transmitting the electricity bargain price of electric power system for example, where the watermarking information is the sender's information "State Grid Tel0000".

The watermarking information and the key to be embedded are converted into binary sequences

$$W = w_1 w_2 w_3 \cdots w_n \qquad w_i \in \{0,1\}, 1 \le i \le n \tag{6}$$

$$K = k_1 k_2 k_3 \cdots k_m \qquad k_i \in \{0,1\}, 1 \le i \le m \tag{7}$$

A one-to-one circular modulo encryption method is used to obtain a new watermarking sequence:

$$M = m_1 m_2 m_3 \ldots m_n \tag{8}$$

$$m_i = w_i \oplus k_{(i \bmod m)} \quad 1 \le i \le n$$

Take advantage of the Hamming code $a_6 a_5 a_4 a_3 a_2 a_1 a_0$ to encode the encrypted watermarking sequence. S_1, S_2, S_3 represents the correction factor in the supervisory relationship.

Table 2 shows that when $i \bmod 4 = 0$, a new binary sequences is generated:

$$H = h_1 h_2 h_3 \ldots h_t , \tag{9}$$

where $t = 7 * l / 4 \qquad h_i \in \{0,1\}, 1 \le i \le t$

Table 1. Error correcting codes and wrong code position corresponding relation table.

S_1, S_2, S_3	Fault code position	S_1, S_2, S_3	Fault code position
001	a_0	101	a_4
010	a_1	110	a_5
100	a_2	111	a_6
011	a_3	000	none

Table 2. Information bits and supervision bit correspondence table.

information bit	supervision bit	information bit	supervision bit
$a_6a_5a_4a_3$	$a_2a_1a_0$	$a_6a_5a_4a_3$	$a_2a_1a_0$
0000	000	0100	110
0001	011	0101	101
0010	101	0110	011
0011	110	0111	000
$a_6a_5a_4a_3$	$a_2a_1a_0$	$a_6a_5a_4a_3$	$a_2a_1a_0$
1000	111	1100	001
1001	100	1101	010
1010	010	1110	100
1011	001	1111	111

4.1.2. Watermarking Embedded

1) In order to make the position where the watermarking information is embedded random to prevent the attackers from getting the watermarking information, Linear Congruence Generator (LCG) is employed. Pre-embedding space D of a text is calculated, and the RGB values of all the characters are changed to (0,0,0), the most common used color in Word. The Linear Congruence Generator is used to generate the pseudo random sequence $\{y_j\}$, $j = 1, 2, 3 \ldots t$

The basic Linear Congruence Generator iterative formula is:

$$\begin{aligned} &y_{n+1} = (a * y_n + c) \bmod m \quad (m > 0) \\ &y_0 \in [0, m-1] \end{aligned}, \tag{10}$$

where y_0 is the seed, a is known as the multiplier, c is the increment and m is the modulus. The currently generated random number, y_{n+1} ,depends only on the previous one, y_n ,when the values of a , c and m are given.

2) Step 1: Traverse the word text and, for the character j (j <N), it is embedded with watermarking interval identifier sgnsart. RGB values are modified to (1,3,3) and (1,3,4) respectively.

Step 2: Select character j . If j <N, go to step 5; Otherwise, judge interval identifier. If the identifier embedding has been completed, go to step 3. Otherwise, go to step 1;

Step 3: If $h_i = 1$, set the RGB value of the current character to (1,1,1); set the RGB value of the next character to (1,1,2);

If $h_i = 0$, set the RGB value of the current character to (1,2,2); set the RGB value of the next character to (1,2,3);

Step 4: Repeat step 1-3 to embed watermarking information.

Step 5: Complete the embedding watermarking information and save the text.

4.1.3. Watermarking Detection

Step 1: Input key and convert it into binary sequence K ;

Step 2: Traverses the text and find where of the RGB value of font color is modified. According to the rules of embedding, extract "0", "1" and obtain the binary sequence S ;

Step 3: Sequence S is decoded and corrected to get the binary sequence M ; if $i \bmod 7 = 0$, calculate correction factor. If each bit of the three-bit correction factor is 0, the watermarking has not been tampered. If other values are obtained, find where the content is tampered, make a wrong code corrected and remove supervision bit.

Step 4: Cycle modulus is operated on Key K and the binary sequence M to get binary sequence of watermarking and convert it to obtain the watermarking signal.

4.2. Robust Watermarking Algorithm Based on Changing Character Spacing to Embed Watermarking Information

This algorithm is designed for covert communication of file contents which requires high robustness and sufficient capacity for hidden information, high imperceptibility and safety. Therefore, watermarking algorithm based on changing the character spacing is chosen as a robust algorithm used in power system. Based on HVS, the changes in character spacing should not exceed the threshold.

4.2.1 Pretreatment

The pretreatment method is the same as is used in the algorithm based on changing the RGB values of character. But here the hidden message is "electricity bargain price: 0000YUAN".

4.2.2. Watermarking Embedded

1) The linear congruence method is used to make the position where the watermarking information is embedded random, text character spacing is unified and is set to "no". In order to prevent the character spacing from being modified which will cause line overflow in the text, binding two adjacent characters, and indenting and widening their spacing respectively are used.

2) Step 1: Traverse the word text and, for the character j (j<N), it is embedded with watermarking interval identification sgnsart. Here the character spacing is increased by 0.1 pounds, the next character spacing is decreased by 0.1 pounds.

Step 2: Select character j . If j <N, go to step 5; Otherwise, judge interval. If the identity embedding

identification has been completed, go to step 3. Otherwise, go to Step 1;

Step 3: If $h_i = 1$, set the current character spacing to increase by 0.05 pounds; set the next character spacing to reduce by 0.05 pounds;

If $h_i = 0$, set the current character spacing to increase by 0.15 pounds; set the next character spacing to reduce by 0.15 pounds;

Step 4: Repeat step 1-3 to embed watermarking information.

Step 5: Complete the embedding watermarking information and save the text.

4.2.3. Watermarking Detection

The watermarking detection method is the same as is used in the algorithm based on changing the RGB values of character.

5. Experimental Results and Analysis

For both algorithms, watermarking embedding and extraction experiments are made to verify whether the change of text in visual happens after the embedded watermarking information is embedded into it, and check whether watermarking information can be extracted correctly. Through attack experiments, robustness of both algorithms is tested.

5.1. Watermarking Embedding and Extracting Experiment

The algorithm based on changing character RGB values of color is tested by embedding experiment, and the watermarking information is "State Grid Tel0000", while the algorithm based on changing character spacing is also tested by embedding experiment, and the watermarking information is "electricity bargain price: 0000YUAN".Both of the keys are "2213".

By extracting test, all watermarking information can be extracted correctly on the texts without suffering attacks. And through the visual, it is impossible to find the text has been modified after being embedded watermarking information.

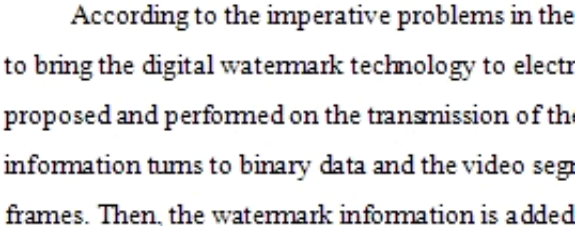
According to the imperative problems in the
to bring the digital watermark technology to electr
proposed and performed on the transmission of th
information turns to binary data and the video segr
frames. Then, the watermark information is added

Fig. 1. The original text.

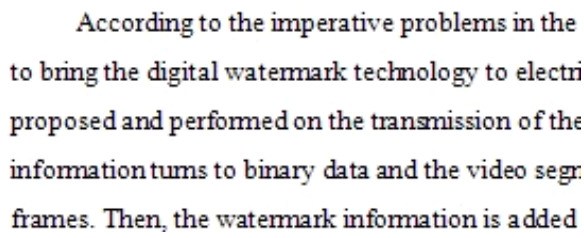
According to the imperative problems in the
to bring the digital watermark technology to electri
proposed and performed on the transmission of the
information turns to binary data and the video segr
frames. Then, the watermark information is added

Fig. 2. Text after embedding watermarking information.

5.2. Attack Experiment

The purpose of digital watermarking attack is designed to make it impossible to detect and extract watermarking information correctly. The robustness of watermarking algorithm can be determined by checking whether the watermarking information can be detected and extracted correctly after being attacked.

The texts embedded with watermarking information are tested by attack experiments. Repeat the test to prevent from generating a right result by chance.

1) Attack based on the format. The text format attributes is changed in different degree.

a) Based on changing the RGB values of font color. Color attack on text, wholly or randomly, shows a weak robustness. But color attack on parts of the texts and any other attack on the texts show a strong robustness. And watermarking information can be detected and extracted correctly.

b) Based on changing character spacing. Character spacing attack on whole texts show a weak robustness. But character spacing attack on parts of the texts and any other attack on the texts show a strong robustness. And watermarking information can be detected and extracted correctly.

2) Attack based on the content. Delete and paste operations on text content in varying degrees.

a) Based on changing the RGB values of font color. Sensitive to the tampered text content.

b) Based on changing character spacing. Have a strong robustness and be able to resist attacks.

3) The Fig. 3 shows that if either the text content or the font color changes, watermarking algorithm based on changing RGB values of font color can accurately determine the specific positions where the watermarking bits are damaged to achieve content certification.

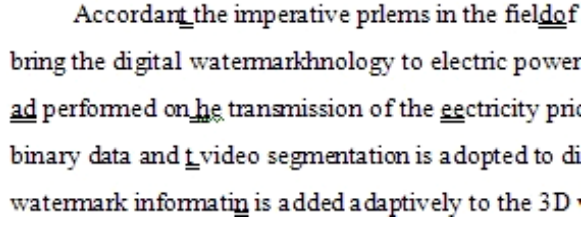
Accordant the imperative prlems in the fieldof
bring the digital watermarkhnology to electric power
ad performed on he transmission of the eectricity pric
binary data and t video segmentation is adopted to di
watermark informatin is added adaptively to the 3D

Fig. 3. Positioning the tampered content.

The experimental results show that the watermarking algorithm based on changing the RGB value of font color show a poor robustness when the content is tampered, and can accurately position the changed contents and characters whose color are changed so that the receiver can identify whether the text has been tampered which can effectively achieve the content authentication. The algorithm based on changing character spacing has a strong robustness and can resist various formats and contend attacks except attacks with changing character spacing. It won't lower text quality and is an effective algorithm with strong robustness which can be used to hide electric power information. Through digital watermarking system composed of two above mentioned algorithms, when the text is edited or suffers from other damages, the electricity bargain price can be still extracted, and the identity of the sender can also be confirmed which helps to confirm the authenticity of texts. The combination of multiple watermarking algorithm may be, to some extents, complementary when texts suffer from attacks, and can achieve various purposes, which makes up for the weakness of a single watermarking algorithm.

5.3. Algorithm Performance Analysis

The most important performance features of digital watermarking technology are capacity, robustness, imperceptibility, security and so on which determine whether the watermarking algorithm has a practical value.

1) Robustness analysis.

a) Algorithm based on changing RGB values of font color to embed watermarking information can resist basic format attacks, but shows a poor robustness when the content is tampered. It can be used for effective content authentication of texts. Once the text content is maliciously modified, it will be hard to extract watermarking correctly.

b) Algorithm based on changing character spacing to embed algorithm watermarking information has a very strong robustness, and it can resist format and content attacks. As for the credible transmission of electronic text in power system, it provides an effective protection to resist the maliciously tampering.

2) Capacity analysis.

Both watermarking algorithms are using (7,4) Hamming code. Each of the 7 code words contains 4-bit information and 3-bit supervision. It can detect two fault codes and correct an error code.

Watermarking algorithm based on font color to code modified lower two bits of G and B component respectively to embed watermarking information, therefore, a character can be embedded with four-bit information. Though the use of error correction coding affects the capacity, the algorithm still has a large capacity.

Watermarking algorithm based on the character spacing to code. Through the theoretical analysis, PW=N (PW is the total capacity provided for loading watermarking information into the text, N is the number of characters of the text), so the number of bits of watermarking information is consistent with that of characters. Although, in order to improve the robustness and security, error correction coding is introduced at the expense of certain data capacity, the requirement of embedding capacity of watermarking information is enough.

3) Imperceptibility analysis.

Both algorithms are designed based on human visual system. The changed values are within the thresholds where the changes are indiscernible by naked eye. In principle, the algorithms have very high imperceptibility. Through embedding watermarking information experiments, it can be seen that the text doesn't be changed or be degraded visually.

4) Security analysis.

The two algorithm uses encryption algorithm to encrypt the watermarking information. Even if the attacker knows the watermarking embedding algorithm, watermarking information would not be completely obtained without the key. And it embeds the information in the random place which makes it much harder for rivals to attack it.

Through the above performance analysis, the proposed multiple text watermarking algorithm can achieve convert communication of important information and content authentication in power system by taking advantage of the combination of two different text watermarking algorithms.

6. Conclusions

Digital watermarking technology can effectively achieve covert communication and content authentication. The robust watermarking and the semi-fragile watermarking algorithm are used for text transmission of electric power system and can hide electricity bargain price effectively in the transmission to resist all kinds of attacks in the transmission. The semi-fragile watermarking algorithm is used to authenticate the sender, which can increase the trust between the sender and receiver. Experimental results show that the use of multiple watermarks technology can effectively resist attacks for different purposes, protect the data in many ways and improve the security of the data.

Electric power industry is the national pillar industry and, with the improvement of normalization level in the electric power system, information security has emerges gradually. With the development of digital watermarking, digital watermarking technology will have a broad application space in the power system security field as an effective technique to solve the problem of information security.

References

[1]. Junji Wu, Sheng Qi, Jifeng He, et al., Application of wavelet-based digital watermarking in power system information security, *Electric Power Automation Equipment*, Vol. 24, Issue 12, 2004, pp. 40-42.

[2]. Yuancheng Li, Xiaolei Wang, Technology of digital ridgelet transform watermarking in electric power system, *Electric Power Information Technology*, Vol. 5, Issue 10, 2007, pp. 120-123.

[3]. C. Yin, L. Li, A. Lv, et al., Technology of digital video watermarking in electric power system, *Relay*, No. 20, 2007, pp. 40-42.

[4]. Ronghui Tu, Jiying Zhao, A semi-fragile audio watermarking scheme based on digital wavelet transform and quantization and its application in power system, *Transactions of the Chinese Society of Electrical Engineering*, Vol. 25, No. 12, 2005, pp. 78-85.

[5]. S. Zhu, J. Liu, An adaptive watermarking-quantifying algorithm for 3-D meshs model of ultra high voltage equipment, *Transactions of China Electrotechnical Society*, No. 12, 2011.

[6]. S. Zhu, J. Liu, Electric power equipment 3D mesh model adaptive robust watermarking algorithm, *Transactions of China Electrotechnical Society*, No. 12, 2011, 029.

[7]. W. Xianpei, Y. Wenxia, W. Quande, Application of digital watermarking technique to credible delivery of texts in power systems, *Automation of Electric Power Systems*, Vol. 26, Issue 18, 2002, pp. 61-64.

[8]. S. H. Low, N. F. Maxemchuk, Performance comparison of two text marking methods, *IEEE Journal on Selected Areas in Communications*, Vo. 16, Issue 4, 1998, pp. 561-572.

[9]. K. T. Ahern, Invisible encoding of attribute data in character based texts and files: *European Patent EP 1145140*, 2001-10-17.

[10] M. J. Atallah, V. Raskin, C. F. Hempelmann, et al. Natural language watermarking and tamperproofing, *Information Hiding, Lecture Notes in Computer Science*, Vol. 2578, 2003, pp. 196-212.

[11]. Z. Yu, X. Liu, A new digital watermarking scheme based on text, in *Proceedings of the International Conference on Multimedia Information Networking and Security MINES'09*, Vol. 2, 2009, pp. 138-140.

[12]. B. Yang, W. Shi, W. Qi, et al., Methods and apparatus for embedding and detecting digital watermarks in a text texts: *U.S. Patent* 8,107,129, 2012-1-31.

[13]. X. Zhou, Z. Wang, W. Zhao, et al. Attack model of text watermarking based on communications, in *Proceedings of the IEEE International Conference on Information Management, Innovation Management and Industrial Engineering*, Vol. 4, 2009, pp. 283-286.

[14]. X. Wei, Sine-wave-based text watermarking for WORD text, in *Proceedings of the IEEE International Conference on Computer and Information Application (ICCIA)*, 2010, pp. 99-102.

[15]. J. T. Brassil, S. Low, N. F. Maxemchuk, Copyright protection for the electronic distribution of text documents, in *Proceedings of the IEEE*, Vol. 87, No. 7, pp. 1181-1196.