Jurnal Teknologi

Full Paper

A NOVEL ZERO-WATERMARKING SCHEME FOR TEXT DOCUMENT AUTHENTICATION

Morteza Bashardoost, Mohd Shafry Mohd Rahim*, Narges Hadipour

UTM-IRDA, Digital Media Centre, Faculty of Computing, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

Article history
Received
3 December 2013
Received in revised form
2 July 2014
Accepted
25 November 2014

*Corresponding author
shafry@utm.my

Graphical abstract

Abstract

The demand for copyright protection of text documents is extremely vital especially when the text is transmitted over the Internet. One of the best practical resolutions to maintain the security of document media is digital watermarking. Several approaches have been proposed to guarantee the safety and protection of the documents against dishonest copying and distribution. This paper suggests an innovative zero-watermarking scheme for the purpose of authentication and tamper detection in simple text documents. The algorithm generates a watermark based on the effective characters of the text contents which can be extracted later using extraction algorithm to identify the status of tampering in the text document. Experimental results demonstrate the effectiveness of the algorithm against random tampering attacks. Watermark pattern matching and watermark distortion rate are used as evaluation parameters on multiple text samples of varying length.

Keywords: Watermarking, text document, authentication, Tamper detection, Zero-watermarking, plain text, information security

1.0 INTRODUCTION

The massive attraction of the Internet and other electronic networks, demonstrates the marketable prospective of distributing documents in the digital world. The electronic spreading of information is quicker, cheaper, and requires less determination than producing paper photocopies and transferring them. Other elements that promote the electronic-data distribution consist of the capability of using a computer to explore for a particular information, and the aptitude of modifying the distributed data for the receivers. Furthermore, recipients could be able to select the favorite method of presentation. Accordingly, digital newspapers, e-books, articles, journals and newsletters are increasingly distributed on Internet.

The characteristics that make Internet appealing, as a spreading medium is the simplicity of manipulating information in a digital form, also seem to make security of intellectual property challenging. Once the producer distributes the electronic content upon payment, the consumer might share this matter on his web site, send it to a Usenet newsgroup or forward it to contacts for a lesser rate or even possibly for free. Currently, electronic content producers must choose between two existing options. On one side, they are really eager to exploit the Internet to make greater incomes regarding to a higher customer number, and also lesser distribution and sending charges using the web. On the other side, they would miss profits because of cyber piracy.

Digital watermark can be defined as a verification code that will remain beside the data visible or invisible and it is inside the data before and after any process such as decryption unlike conventional cryptographic techniques. To protect intellectual property with digital format, digital watermark is useful and its invisibility is an important point. The copyright information that copyright protection put into the digital data, it becomes intact and when there is any question about the digital data, the information refers to original owner.

Comparing to the other file types like video, audio and image, documents are more likely to be copied, reproduced or modified. Consequently, text document requires more proficient techniques for copyright protection and tamper detection. Traditional watermarking algorithms modify the contents of the digital medium to be protected by embedding a watermark. This traditional watermarking approach is not practical for plain text. A specialized watermarking approach such as zero-watermarking would do the needful for plain text. In this paper, we propose a novel zero- watermarking algorithm which utilizes the contents of text itself for its authentication. A zero-watermarking algorithm does not change the characters of original data, but utilize the characters of original data to construct original watermark information.

The paper is organized as follows: Section 2 provides an overview of the previous work done on text watermarking. The proposed embedding and extraction algorithm are described in detail in section 3. Section 4 presents the experimental results for the tampering (insertion, deletion and re-ordering) attacks. Performance of the proposed algorithm is evaluated on multiple text samples. The last section concludes the paper along with directions for future work.

2.0 EXPERIMENTAL

Text watermarking techniques have been proposed and classified by many literatures based on several features and embedding modes of text watermarking. We have examined briefly some traditional classifications of digital watermarking as in literatures. These techniques include image based, content based, format based, features based, synonym substitution based, and syntactic structure based, acronym based, noun-verb based, and many others of text watermarking algorithms that depend on various viewpoints. We can categorize the existing text watermarking approaches into four main categories.

2.1 Image-Based Approach

In this category of methods towards digital document watermarking, the text document considered as an image and the process of embedding follows the principals of digital image watermarking. In the majority of the approaches of this group, scanning the text document or conversion to the image file is the first step of algorithm. Then the watermark will be inserted by utilizing one of the image watermarking techniques. Therefore, the visual format of the document is the main elements of watermarking in the image based approaches.

Brassil, *et al*. [1] issued the line-shift approach that reallocates a line in the vertical or horizontal direction to embed the watermark. Word-shifting which is similar method that also proposed by Brassil, *et al*. [2], changes the position of the words to the right or left base on the value of watermark. Brassil, *et al*. [3] also proposed a feature-coding method which intends to change some features of the document base on the embedded watermark.

Huang and Yan [4] proposed an alternate strategy that works based on the average inter-word space of every line of document. The feature and the pixel level approaches, that mark the content by changing the stroke characteristics like width or serif, were proposed in [5]. Also the methodologies which uses converted grayscale image of content was proposed in [6].

Authors of [7] used the fractal theory to embed the watermark into binary document image. Another attempt to protect the content of text document was performed by [8] that utilizes the combination of image plus text watermark. In [9] reversible watermarking approach was proposed that relies on identification of specific parts of image by using Pixel Histogram shifting and Dynamic Prediction Error Histogram Shifting (DPEHS).

2.2 Syntactic Approach

Textual content consists of sentences. Sentences are comprised of words as well, and words could be nouns, verbs, articles, prepositions, adjectives, adverbs and so on. Sentences include various syntactic compositions that are determined by language and its particular conventions. Utilizing syntactic transformations on textual content structure in order to insert watermark has been introduced recently as one of the several strategies towards text watermarking.

Mikhail. J. Atallah, *et al*. first presented the natural language watermarking system by using the syntactic composition of text [10] where the particular syntactic tree is created and conversions are demonstrated on this tree in order to insert the watermark.

Hassan *et al*. suggested the natural language watermarking procedure simply by doing the morphosyntactic modifications to the text document [11]. The document is initially converted to a syntactic tree diagram in which the hierarchies and the functional dependencies are created very revealing and watermark is inserted.

Hassan *et al*. later reviewed morphosyntactic resources for text document watermarking and also designed a new syntax-based natural language watermarking system [12] that initially converts the unmarked text to a syntactic tree and subsequently functions on the sentences within syntax tree structure and then completes binary modifications under control of Wordnet to prevent semantic falls.

In [13], a synonym substitution has been proposed to embed watermark by replacing certain words with their synonyms without changing the sense and context of text. C. Kim [14] developed the method of text watermarking for Korean by syntactic analysis.

2.3 Semantic Approach

Semantic-based approaches of text document watermarking focus on the semantic arrangement of the elements of the document such as verbs, nouns, pronouns, adjectives, the spelling of the words, synonym, acronyms, etc.

The first attempt of semantic base watermarking was performed by Atallah *et al*. [15]. In [16], the authors stated the similarities and also differences of the image-based and semantic based watermarking techniques. They also investigated how the image-based techniques can be applied on the natural language watermarking domain.

Later, Topkara *et al*. [17] presented a new sentence-level approach that implements very small changes to the document when the style of document is essential. Also, they used word-level marking to improve the resilience property of sentence-level scheme.

A noun-verb dependent scheme for text watermarking, which is suggested in [18], makes use of nouns and verbs within a phrase parsed with a grammar parser utilizing semantic networks. Another sentence based text watermarking, which depends on the conception of orthogonality between the features of the sentences, was proposed by Mercan *et al*. [19] . In another approach Mercan *et al*. proposed a new technique [20] of text watermarking that relied on idiosyncrasies to embed the watermark in the document.

Multiple approaches [21], [22] were developed stand on the linguistic semantic phenomena of presuppositions. An enhanced semantic watermarking technique, which introduces Text Meaning Representation (TMR) string reordering system before watermark insertion, was proposed in [23]. The enhanced algorithm has the attribute of confronting sentence against reordering attack and has a better robustness.

The watermarking base on predefined semantic and syntactic rules proposed in [24] that uses the structural means of HTML to embed the formulated watermark.

2.4 D_AZero-based Approach

Text watermarking techniques based on Zero-based watermarking are content features dependent. There are several approaches that designed for text documents. The authors of [25] proposed an algorithm for tamper detection in plain text documents based on length of words and using digital watermarking and certifying authority techniques.

Another algorithm has been proposed by [26] for improvement of text authenticity in which utilizes the contents of text to generate a watermark and this watermark is later extracted to prove the authenticity of text document.

The approach that has been proposed by [27] presented for copyright protection of text contents based on occurrence frequency of non-vowel ASCII characters and words.

In [28], the proposed algorithm attempts to protect all open textual digital contents from counterfeit in which is insert the watermark image logically in text and extracted it later to prove ownership.

In [29], Chinese text zero-watermark approach has been proposed based on space model by using the two-dimensional model coordinate of word level and the sentence weights of sentence level.

The above mentioned approaches and algorithms are not applicable to all types of text documents under random tampering attacks and are not designed specifically to solve tamper detection problem; hence we propose a zero-watermarking algorithm which incorporates the contents of text for its protection and tamper detection.

3.0 PROPOSED SCHEME

This paper presents an improved intelligent approach of English text zero-watermarking based on selected characters and Markov matrix for content authentication and tampering detection of text documents. An improved approach depends on character mechanism and Markov model to improve the performance, complexity and accuracy of tampering detection. An improved approach should perform watermark generation, embedding, extraction and detection processes under higher accuracy and security measures. An improved approach hybrid text zero-watermarking techniques and soft computing tools for natural language processing and protect the digital text documents. A Markov model uses for text analysis and extracts the interrelationship between its contents as probabilistic patterns based on character level and first order of Markov model in order to generate the watermark information. This watermark can later be extracted using extraction algorithm and matched with watermark generated from attacked document using detection algorithm for identifying any tampering and prove the authenticity of text document.

A Watermark Generation and Embedding

The watermark generation and embedding algorithm requires the original text document (To) as input which provided by the author, then as a pre-processing step it is required to perform conversion of capital letters to small letters. A watermark pattern is generated as the output of this algorithm. This watermark is then stored in watermark database along with the main properties of the original text document such as document identity, author name, current date and time. This stage includes involves following algorithms, which are pre-processing and Effective Characters List (ECL) extraction, text analysis, and watermark generation and embedding as shown in Figure 1.

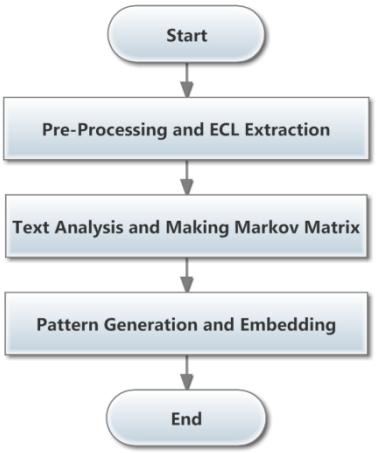

Figure 1 Watermark generation and embedding

1 Pre-processing and ECL Extraction

This algorithm requires the original text document as inputs, and provides Markov matrix as outputs as it demonstrated in Figure 2. The given document first processed to convert all capital letters to lowercases. Then, based on the Effectiveness Ratio (ER), list of characters which compose ER percent of document is generated. Finally, the Markov matrix is produced according to the ECL. Building the states and transition matrix is the most base part of text analysis and watermark generation using Markov model. A Markov matrix that represents the possible states and transitions available in given text is constructed without reputations. In this approach, each character from ECL represents as state and transition in the Markov matrix. During building process of Markov matrix, the proposed algorithm initialize all transition values by zero to use these cells later to keep track of the number of times that the i^{th} character is followed by the j^{th} character within given text document.

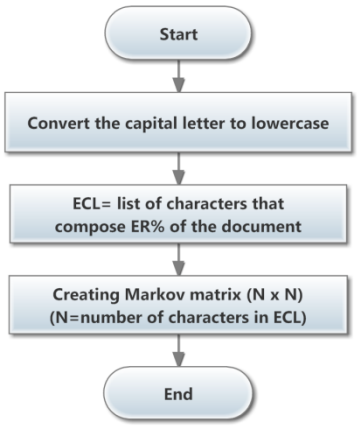

Figure 2 Pre-processing and ECL extraction

2 Test Analysis

This algorithm takes the pre-processed text document and initialized Markov matrix as input, and provides the watermark patterns as output. After the Markov matrix was constructed, text analysis process should be done using Markov model by finding the interrelationship between characters of the given text document. In the other word, the proposed algorithm computes the number occurrences of the next state transitions for every present state. Matrix of transition probabilities represents the number of occurrences of transition from a state to another.

3 Watermark Generation and Embedding

After performing the text analysis and extracting the probability features, the watermark is obtained by identifying all the values in the above Markov matrix. These values are sequentially concatenated to generate a watermark pattern, denoted by WMP_O. Furthermore, the ECL string is joined to the start of WMP_O for the tamper detection purpose in the extraction stage.

The embedding process will be done logically during text analysis process by keeping the tracks of all transitions and its values shown in the Markov matrix. In which the cells of transitions contains the number of times that the i_{th} character is followed by the j_{th} character within given text document. These tracks can be used later by detection algorithm for matching it with those tracks that will be producing from the attacked text document. This watermark is then stored in the watermark database along with some properties of the original text document such as document identity, author name, current date and time. After watermark generation as sequential patterns, an MD5 message digest is generated for obtaining a secure and compact form of the watermark, which is used for authentication purpose.

B Watermark Extraction and Detection

The watermark detection algorithm is on the base of zero-watermark, so before detection for attacked text document TA, the proposed algorithm still need to generate the attacked watermark patterns. When received the watermark patterns, the matching rate of patterns and watermark distortion are calculated in order to determine tampering detection and content authentication. This stage includes two main processes which are watermark extraction and detection.

Extracting the watermark from the received attacked text document and matching it with the original watermark will be done by the detection algorithm. The proposed watermark extraction algorithm takes the attacked text document, and performs the same watermark generation algorithm to obtain the watermark pattern for the attacked text document as shown in Figure 3.

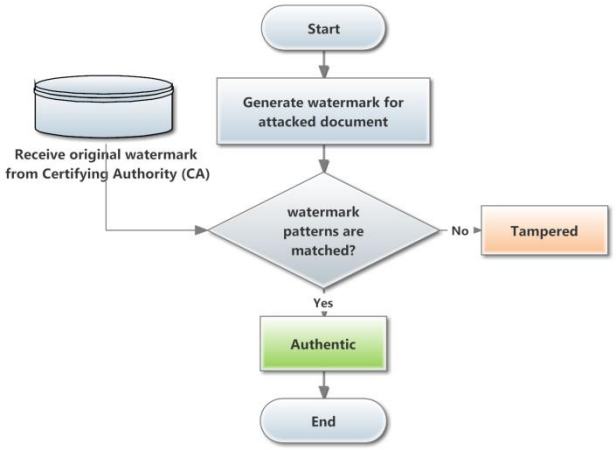

Figure 3 Watermark extraction and detection

1 Watermark Extraction Algorithm

In this algorithm the proposed approach takes the attacked text document (TA), original watermark patterns or original text document as inputs and the procedure is similar to that of watermark generation. Output of this algorithm is attacked watermark patterns (WMPA).

2 Watermark Detection Algorithm

After extracting the attacked watermark pattern, the watermark detection is performed in three steps:

- Primary matching is performed on the MD5 compressed watermark pattern of the original document WMPO, and the attacked document WMPA. If these two compressed patterns are found the same, then the text document will be called authentic text without tampering. If the primary matching is unsuccessful, the text document will be called not authentic and tampering occurred, then we proceed to the next step.
- Secondary matching is performed by comparing the components associated with each state of the overall pattern. This algorithm compares the extracted watermark pattern for each state with equivalent transition of original watermark pattern. To measure the Pattern Matching Rate, first the PMR is calculated for each transition (PMR_T) and subsequently the PMR_S is determined for each state. This process can be described by the following mathematical equations (1) and (2).

$$PMR_T(i,j) = \left| \frac{WMP_O(i,j) - (WMP_O(i,j) - WMP_A(i,j))}{WMP_O(i,j)} \right| \quad (1)$$

$$PMR_S(i) = \left| \frac{\sum_{j=1}^{n} (PMR_T(i,j))}{TotalStatePatternCount(i)} \right| \quad (2)$$

After we get the pattern matching rate of every state, we have found the weight of every state from whole states in Markov matrix. Finally, the PMR is calculated by equation (3), which represents the pattern matching rate between the original and attacked text document. The watermark distortion rate, which is presented in equation (4), refers to tampering amount occurred by attacks on contents of attacked text document, this value represent in WDR.

$$PMR = \frac{\sum_{i=1}^{n} \left| \frac{PMR_S(i) * transitionsFrequency(i)}{totalNoOfTransitions} \right|}{N} \quad (3)$$

$$WDR = 1 - PMR \quad (4)$$

4.0 EXPERIMENTAL RESULT

We used 4 samples of variable size text from the data set designed in [30] for our experiments. These samples have been collected from Reuters' corpus, e-books, and web pages. Insertion and deletion of words and sentences was performed at multiple randomly selected locations in text.

The values of PMR ranges between 0 (the lowest) and 1(the highest) with desirable value close to 1. The values of WDR also ranges between 0 (the highest) and 1(the lowest) with value close to 0 as desirable value. The proposed approach tested for deletion, insertion and reordering attacks and the experimental results are discussed in following sections.

A Deletion Attack

The ER tested for the values 70%, 80%, and 90% to select the effective characters list (ECL). Pattern matching rate and watermark distortion rate are shown in Table 1 for the 3 levels of deletion attack.

It can be observed in Table 1 that tampering with text is always detected, whether it is low, moderate or high. Watermark distortion rate greater than 0 indicate that watermark has been distorted as a result of tampering. Furthermore, for all the tested cases the accuracy of the tamper detection improved when the Effectiveness Ratio increased.

B Insertion Attack

Table 2 represents the values of PMR and WDR for the insertion attack on the same input documents as previous experiment.

The results of reordering prove the accuracy of tamper detection in our proposed method. Similar to deletion attack, the precision of tamper detection becomes more satisfactory by increasing the ER value. Moreover, the algorithm is more or less effective for Large and Very Large Size Text categories.

C Reordering Attack

We have performed 3 levels of randomly reordering attack on each sample document of our dataset. We have applied our proposed approach using different ER values. The matching and distortion rates of retrieved watermark of this experiment are shown in Table 3.

It can be clearly observed that the PMR and WDR values for reordering attack are even more accurate than deletion and insertion attacks. In fact, the character based Markov model helps to keep the transition patterns of original and attacked documents more similar to each other.

The value of distortion rate indicates that the text has been tampered and is not authentic. This proves that the accuracy of watermark gets affected even with minor tampering and the evident watermark fragility proves that text has been attacked.

Table 1 PMR and WDR values for deletion attack

Category	Word Count	Deletion Volume	Effectiveness Ratio for ECL					
			70%		80%		90%	
			PMR	WDR	PMR	WDR	PMR	WDR
[SST]	179	10%	0.8136	0.1864	0.8301	0.1699	0.8792	0.1208
		20%	0.6288	0.3712	0.6531	0.3469	0.7371	0.2629
		50%	0.3177	0.6823	0.2986	0.7014	0.3510	0.6490
[MST]	469	10%	0.8091	0.1909	0.8113	0.1887	0.8535	0.1465
		20%	0.5806	0.4194	0.6273	0.3727	0.6640	0.3360
		50%	0.2992	0.7008	0.3186	0.6814	0.3255	0.6745
[LST]	2018	10%	0.7772	0.2228	0.8381	0.1619	0.8449	0.1551
		20%	0.5618	0.4382	0.5977	0.4023	0.6125	0.3875
		50%	0.2632	0.7368	0.2689	0.7311	0.3755	0.6245
[VLST]	4647	10%	0.5933	0.4067	0.6334	0.3666	0.8002	0.1998
		20%	0.4281	0.5719	0.5441	0.4559	0.5772	0.4228
		50%	0.2115	0.7885	0.2703	0.7297	0.2998	0.7002

Table 1 PMR and WDR values for insertion attack

Category	Word Count	Insertion Volume	Effectiveness Ratio for ECL					
			70%		80%		90%	
			PMR	WDR	PMR	WDR	PMR	WDR
[SST]	179	10%	0.7919	0.2081	0.8546	0.1454	0.8766	0.1234
		20%	0.6421	0.3579	0.6770	0.3230	0.7628	0.2372
		50%	0.3409	0.6591	0.3551	0.6449	0.3831	0.6169
[MST]	469	10%	0.7677	0.2323	0.7726	0.2274	0.8081	0.1919
		20%	0.6117	0.3883	0.6552	0.3448	0.7115	0.2885
		50%	0.3007	0.6993	0.3775	0.6225	0.4123	0.5877
[LST]	2018	10%	0.6117	0.3883	0.6333	0.3667	0.6693	0.3307
		20%	0.5255	0.4745	0.5382	0.4618	0.5571	0.4429
		50%	0.2210	0.7790	0.2863	0.7137	0.3030	0.6970
[VLST]	4647	10%	0.3183	0.6817	0.5227	0.4773	0.5448	0.4552
		20%	0.2100	0.7900	0.2392	0.7608	0.3106	0.6894
		50%	0.3715	0.6285	0.3938	0.6062	0.4118	0.5882

Table 2 PMR and WDR values for reordering attack

Category	Word Count	Reordering Volume	Effectiveness Ratio for ECL					
			70%		80%		90%	
			PMR	WDR	PMR	WDR	PMR	WDR
[SST]	179	10%	0.8447	0.1553	0.8559	0.1441	0.8802	0.1198
		20%	0.6931	0.3069	0.7223	0.2777	0.7781	0.2219
		50%	0.3807	0.6193	0.4112	0.5888	0.4375	0.5625
[MST]	469	10%	0.7932	0.2068	0.8117	0.1883	0.8443	0.1557
		20%	0.6421	0.3579	0.6729	0.3271	0.7073	0.2927
		50%	0.3498	0.6502	0.3892	0.6108	0.4293	0.5707
[LST]	2018	10%	0.6336	0.3664	0.6481	0.3519	0.6829	0.3171
		20%	0.5381	0.4619	0.5565	0.4435	0.5831	0.4169
		50%	0.2718	0.7282	0.2963	0.7037	0.3632	0.6368
[VLST]	4647	10%	0.4552	0.5448	0.5387	0.4613	0.5701	0.4299
		20%	0.2554	0.7446	0.2735	0.7265	0.3431	0.6569
		50%	0.3889	0.6111	0.4007	0.5993	0.4130	0.5870

5.0 DISCUSSION

The above implementation was an effort to understand how to remove noises to increase the quality of input fingerprint image and which technique will be the best to remove noise in the fingerprint images. Based on the experimental result, we can see that the proposed algorithm give second highest value of PSNR and median filter contributed the highest value of PSNR.

This happened because of median filter removed noise effectively in the filtering process. Thus, it eliminates the impulse noise only if the noisy pixels occupy less than half the area of the neighborhood. While, even though there is median filter in the proposed algorithm, noise still produced after the equalization process. Generally, histogram equalization is a great technique to increase the global contrast between furrows and ridges where it connect the fake broken points created during data acquisition. These broken points created caused by unsatisfactory quantity of ink or poor quality of sensing devices. In order to reduce the noise in the image after equalization process, grayscale enhancement technique was applied but the quality of final image from this process cannot surpass the quality of image from median filter.

6.0 CONCLUSION

The existing text watermarking solutions for text authentication are not applicable under random tampering attacks and on all types of text. With the small volume of attack, it becomes impossible to identify the existence of watermark and to prove the authenticity of information, especially when the document size is large. We have developed a zero-text watermarking algorithm, which utilizes the contents of text to generate a watermark and this watermark is later extracted to prove the authenticity of text document. We evaluated the performance of the algorithm for three types of random tampering attacks, namely deletion, insertion and reordering. Furthermore, the idea of effective characters list (ECL) is exploited to reduce the size of watermark and complexity of the text analysis. However, the higher effectiveness ratio (ER) led to a better estimation for the tamper detection. Results show that our algorithm always detects tampering even when the tampering volume is low.

Acknowledgement

This research was funded by the Universiti Teknologi Malaysia through Research University Grant and Flagship-COE, and manage by Research Management Centre under Vot No. Q.J130000.2528.09H69 and Q.J130000.2428.02G28.

References

[1] Brassil, J. T., S. Low, N. F. Maxemchuk and L. O. 'Gorman. 1995. Electronic Marking and Identification Techniques to Discourage Document Copying. *IEEE J. Sel. Areas Commun.* 13(8): 1495-1504.

[2] Brassil, J., S. Low, N. Maxemchuk and L. O. 'Gorman. 1995. Hiding Information in Document Images. In *Proceedings of the 29th Annual Conference on Information Sciences and Systems.* Baltimore, Maryland. 22-24 March 1995. 482-489.

[3] Brassil, J. T., S. Low and N. F. Maxemchuk. 1999. Copyright Protection for the Electronic Distribution of Text Documents. In *Proceedings of the IEEE.* 87(7): 1181-1196.

[4] Huang, D. and H. Yan. 2001. Interword Distance Changes Represented by Sine Waves for Watermarking Text Images. *Circuits Syst. Video Technol. IEEE Trans.* 11(12): 1237-1245.

[5] Harouni, M., M. S. M. Rahim, D. Mohamad, A. Rehman and T. Saba. 2012. Online Cursive Persian/Arabic Character Recognition by Detecting Critical Points. *Int. J. Acad. Res.* 4(2): 208-213.
[6] Sharifara, A., M. S. M. Rahim and M. Bashardoost. 2013. A Novel Approach to Enhance Robustness in Digital Image Watermarking using Multiple Bit-Planes of Intermediate Significant Bits. In *International Conference on Informatics and Creative Multimedia (ICICM)*. Kuala Lumpur, Malaysia. 3-6 Sept. 2013. 22-27.
[7] Daraee, F. and S. Mozaffari. 2014. Watermarking in Binary Document Images using Fractal Codes. *Pattern Recognition Letters.* 35: 120-129.
[8] Rameshbabu, K., P. Prasannakumar and K.E. Balachandrudu. 2013. Text Watermarking using Combined Image and Text. *Int. J. Eng. Res. Technol.* 2(12): 3812-3818.
[9] Jadhav, A. and M. O. Sharma. 2014. Reversible Watermarking Technique using Histogram Shifting Modulations. *Int. J. Innov. Technol. Explor. Eng.* 4(4): 29-33.
[10] Atallah, M. J., C. McDonough, S. Nirenburg and V. Raskin. 2000. Natural Language Processing for Information Assurance and Security: An Overview and Implementations. In *Proceedings of the 2000 Workshop on New Security Paradigms.* Ballycotton Co. Cork, Ireland. 18-21 Sept. 2000. 51-65.
[11] Meral, H., B. Sankur, A. S. Özsoy, T. Gungor and E. Sevinc. 2009. Natural Language Watermarking via Morphosyntactic Alterations. *Comput. Speech Lang.* 23(1): 107-125.
[12] Meral, H. M., E. Sevinç, E. Ünkar, B. Sankur, A. S. Özsoy and T. Güngör. 2007. Syntactic Tools for Text Watermarking. In Edward, J.D. and Ping, W.W. (eds.). *SPIE 6505, Security, Steganography and Watermarking of Multimedia Contents IX*.
[13] Topkara, U., M. Topkara and M. Atallah. 2006. The Hiding Virtues of Ambiguity: Quantifiably Resilient Watermarking of Natural Language Text through Synonym Substitutions. In *Proceedings of the 8th Workshop on Multimedia and Security*. Geneva, Switzerland. 26-27 Sept. 2006. 164-174.
[14] Kim, M. 2008. Text Watermarking by Syntactic Analysis. In *Proceedings of the 12th WSEAS International Conference on Computers*. Heraklion, Crete Island, Greece. 23-25 July 2008. 904-909.
[15] Atallah, M. J., V. Raskin, M. Crogan, C. Hempelmann, F. Kerschbaum, D. Mohamed and S. Naik. 2001. Natural Language Watermarking : Design, Analysis and a Proof-of-Concept Implementation. In Ira, S.M. (ed.). *4th International Workshop, IH 2001*. Springer Berlin Heidelberg.
[16] Topkara, M., C. M. Taskiran and E. J. Delpj III. 2005. Natural Language Watermarking. In Edward, J. and Ping, W.W. (eds.). *Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 5681*.
[17] Topkara, M., U. Topkara and M. J. Atallah. 2006. Words Are Not Enough: Sentence Level Natural Language Watermarking. In *Proceedings of the 4th ACM International Workshop on Contents Protection and Security*. Santa Barbara, CA, USA. 28 Oct. 2006. 37-46.
[18] Sun, X. and A. J. Asiimwe. 2005. Noun-Verb based Technique of Text Watermarking using Recursive Decent Semantic Net Parsers. *Lect. Notes Comput. Sci.* 3612: 968-971.
[19] Jiang, L., Z. Xu and Y. Xu. 2012. Commutative Encryption and Watermarking based on Orthogonal Decomposition. *Multimed. Tools Appl.* 70(3): 1617-1635.
[20] Topkara, M., U. Topkara and M. J. Atallah. 2007. Information Hiding through Errors: A Confusing Approach. In Edward, J. D. and Ping, W. W. *Proc. SPIE 6505, Security, Steganography and Watermarking of Multimedia Contents IX*.
[21] Macq, B. and O. Vybornova. 2007. A Method of Text Watermarking using Presuppositions. *Proc. SPIE 6505, Security, Steganography and Watermarking of Multimedia Contents IX*.
[22] Vybornova, O. and B. Macq. 2007. Natural Language Watermarking and Robust Hashing based on Presuppositional Analysis. In *IEEE International Conference on Information Reuse and Integration (IRI 2007)*. Las Vegas, IL. 13-15 Aug. 2007. 177-182.
[23] Lu, P., Z. Lu, Z. Zhou and J. Gu. 2009. An Optimized Natural Language Watermarking Algorithm based on TMR. In *Proceedings of 9th International Conference for Young Computer Scientists*. Hunan. 18-21 Nov. 2008. 1459-1463.
[24] Mir, N. 2014. Copyright for Web Content using Invisible Text Watermarking. *Comput. Human Behav.* 30: 648-653.
[25] Jalil, Z., A. Mirza and H. Jabeen. 2010. Word Length based Zero-Watermarking Algorithm for Tamper Detection in Text Documents. In *2nd International Conference on Computer Engineering and Technology (ICCET)*. Chengdu. 16-18 Apr. 2010. 378-382.
[26] Jalil, Z., A. Mirza and M. Sabir. 2010. Content based Zero-Watermarking Algorithm for Authentication of Text Documents. *Int. J. Comput. Sci. Inf. Secur.(IJCSIS)*. 7(2): 212-217.
[27] Jalil, Z., A. Hamza, S. S. M. Arif and A. Mirza. 2010. A Zero Text Watermarking Algorithm based on Non-Vowel ASCII Characters. In *International Conference on Educational and Information Technology (ICEIT)*. Chongqing. 17-19 Sept. 2010. 503-507.
[28] Tayan, O., M. N. Kabir and Y. M. Alginahi. 2014. A Hybrid Digital-Signature and Zero-Watermarking Approach for Authentication and Protection of Sensitive Electronic Documents. Hindawi Publishing Corporation 2014.
[29] Zhang, S. and S. Yan. 2014. Unicode-based Zero-Watermarking Algorithm for Chinese Documents. In Maozhu, J. and Zhenyu, D. (eds.). *Management Innovation and Information Technology*.
[30] Jalil, Z. and A. M. Mirza. 2010. Text Watermarking using Combined Image-plus-Text Watermark. In *2nd Int. Work. Educ. Technol. Comput. Sci.* Wuhan. 6-7 March 2010. 11-14.