International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:4, No:10, 2010

# Improved Zero Text Watermarking Algorithm against Meaning Preserving Attacks

Jalil Z., Farooq M., Zafar H., Sabir M., and Ashraf E.

***Abstract*—**Internet is largely composed of textual contents and a huge volume of digital contents gets floated over the Internet daily. The ease of information sharing and re-production has made it difficult to preserve author's copyright. Digital watermarking came up as a solution for copyright protection of plain text problem after 1993. In this paper, we propose a zero text watermarking algorithm based on occurrence frequency of non-vowel ASCII characters and words for copyright protection of plain text. The embedding algorithm makes use of frequency non-vowel ASCII characters and words to generate a specialized author key. The extraction algorithm uses this key to extract watermark, hence identify the original copyright owner. Experimental results illustrate the effectiveness of the proposed algorithm on text encountering meaning preserving attacks performed by five independent attackers.

***Keywords*—**Copyright protection, Digital watermarking, Document authentication, Information security, Watermark.

## I. Introduction

SECURITY of digital contents has gained tremendous importance in current digital era. Internet has become an essential part of our daily life for the transfer of different forms of data such as emails, news papers, articles, websites, images, audios, videos, commercials, and opinion blogs. Most of the information over the Internet is in the form of text and the copyright protection of text is one of the major concerns of its creator\author.

In order to protect copyrights, digital watermarking came up as a solution for the identification of the owner of the concerned copyright material. In case of audio, video, and images; digital watermarking has been used for decades. However, no significant work has been done regarding the copyright protection of plain text documents.

Text is a very important form of the Internet. It is part of e-books, websites, articles, news, chats, emails, and SMS. Text documents face many threats such as copying, tampering, plagiarism, reproduction, and paraphrasing attacks. The best solution to address these problems is digital watermarking, which not only helps in authentication of the digital material but also in its protection. Digital watermarking can be used to identify the owner of the copyright material which may be in the form of audio, video, image, a plain text. There are two forms of digital watermarking, visible and invisible but the later one is considered more robust. Digital watermark is an identification code embedded in the data. It mean that unlike conventional cryptographic techniques it remain present within the data even after the decryption [1].

The problem of Digital Text Watermarking has been studied in the past but a practical and efficient text watermarking algorithm is not yet provided for meaning preserving attacks. The main contributions of this paper to the watermarking community are:

- A zero text watermarking algorithm for copyright protection of plain text documents is proposed.
- There is no restriction about the type and length of text.
- Pure alphabetical watermarks are used which are more convenient to be used for plain text
- No changes are made in the text rather attributes of the text are used in the proposed approach.
- This approach towards medium size files like emails, short articles and news is robust and practical to identify the original copyright owner of the contents.

This paper is organized as follows: Section 2 gives an overview about the earlier work on text watermarking. The propose algorithm for embedding and extraction are discussed in detail in section 3. In section 4, the experimental results for intelligent meaning preserving attacks (insertion and deletion) performed by five different attackers are provided. Efficiency of the proposed algorithm is analyzed by five different attacks on the same text. The last section concludes the paper along with directions for future research.

## II. State of the Art

Text watermarking is an emerging domain for research. A robust and practical solution may open new horizons to the information security world. Many watermarking techniques have been developed since 1993, which includes text watermarking that uses text image, synonyms based, noun-verb, word and sentence structure based, acronyms based schemes and many others. These schemes can be placed in the following categories; image-based schemes, syntactic schemes, semantic schemes and structural schemes.

Jalil Z. is with Department of Computer Science, National University of Computer and Emerging Sciences, Islamabad, Pakistan. (email: zunera.jalil@nu.edu.pk)

Farooq M. and Zafar H. are with Air University, Islamabad, Pakistan.(email: 91102@students.au.edu.pk)

Sabir M. and Ashraf E. are faculty members at Department of Computer Science and Software Engineering, Air University Islamabad, Pakistan. (email: maria.sabir@mail.au.edu.pk, erum.ashraf@mail.au.edu.pk )

In image-based schemes, the binary watermark is embedded in text image. Brassil, et al. [2, 3] was the first to propose a text watermarking scheme using text image and binary watermark. Then, Maxemchuk, et al. [4-6] and Low, et al. [7, 8] analyzed the efficiency of these schemes. Afterwards, Huang and Yan [9] introduced an algorithm based on an average inter-word distance in each line.

In syntactic schemes, the syntactic structure of a text is used to embed the watermarks. Mikhail J. Atallah, et al. was the first one to propose the natural language watermarking scheme by using syntactic structure of text [10, 11]. The syntactic tree is build to embed the watermark by application of transformations, while keeping all the text properties intact. In order to watermark the text, Hassan, et al. used morpho-syntactic alterations [12]. Hassan also provided an overview of syntactic tools available for text watermarking [13].

In semantic schemes, the watermark is embedded using the semantics of text and its language properties. Atallah et al. proposed the semantic watermarking techniques one decade earlier [14]. After that algorithm based on synonym substitution was proposed by which certain words are replaced with their synonyms when watermark was embed [15]. Another approach based on nouns and verbs was proposed by Sun, et al. for text watermarking [16]. Another scheme was proposed by Topkara, et al. is by using abbreviations, acronyms and typos to embed the water mark in the text watermarking [17]. Presuppositions is a unique linguistic approach to watermark the text [18] by which the structure, meaning and rearrangement are detected to embed watermark bits. Another algorithm was developed using Text meaning representation (TMR) [19].

Recently the most practical and robust zero watermarking approach has been adopted using text structure is named as the structural approach. In this approach, text is not altered to embed watermark information. A text watermarking technique to protect copyrights of text documents by using existence of double letter (aa-zz) in the text, have been proposed [20]. Similarly, new watermarking algorithm named as zero watermarking based on structural components has recently been proposed [21].

Text watermarking solutions are not resistant to text retyping and paraphrasing attacks. In this paper we proposed a zero text watermarking algorithm which is resistant towards meaning preserving attacks and then analyze the performance under intelligent meaning preserving attacks (insertion, deletion, and paraphrasing) performed independently separately by different attackers. Previously we proposed a zero watermarking algorithm [22] and studied its performance under random attacks. However, in real life plain text encounters intelligent meaning preserving attacks. The theme of intelligent attack is preservation of the main theme and meaning of text. We have now maintained most occurring non vowel character (MONV) in a circular list.

## III. Proposed Algorithm

The proposed algorithm utilizes existence of non vowel characters to watermark the text document. The original copyright owner of the text document uses an algorithm named as embedding algorithm to generate a key based on given watermark and structure of the text. This algorithm is known as zero watermarking algorithms since it generates the author's key by using properties of the text without altering it.

The text document is first analyzed and prepositions are identified. The occurrence count of all prepositions is obtained and then average frequency preposition (AFP) is identified. This AFP is then used to create the partitions of text. After this, the occurrence count of all alphabetical characters in each partition is obtained and the highest occurring non-vowel ASCII characters are identified and populate MONV list. This MONV list is then used to generate an author key based on watermark provided by the owner.

In order to protect the copyrights of the owner this author key is registered with a certification authority (CA). The original watermark and author key is time stamped and set aside with the CA. If in case anyone violates the copyrights of the author, this generated key is used to identify the original copyright owner by using the extraction algorithm to resolve the copyright clash. It may be possible that one text may have more than one claim. So, in this case the earlier registered key will be regarded as the original one. The proposed algorithm works for both intelligent insertion and deletion attacks on the text.

Watermark used in generating the key is to be pure alphabetical and must be carefully chosen to identify the owner or an association or group that owns the copyrights of the text. Watermarking process involves two steps, watermark embedding and watermark extraction. The first step, watermark embedding is performed by the original author and the second step extraction is done later by CA on the behalf of writers claim.

### *A. Embedding Algorithm*

The algorithm used to embed the watermark in the text is called embedding algorithm. In this algorithm watermark is inserted logically in the text and finally generates the key based on AFP and non-vowel characters.

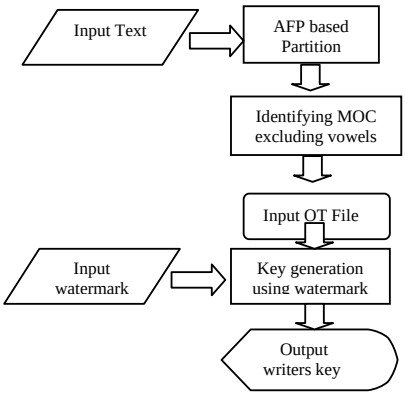


Fig. 1 Watermark embedding process

Fig. 1 shows the watermark embedding process. We used AFP instead of HFP because highest occurring prepositions

also have high chances of deletion. In HFP, created partitions are in large number and the watermark accuracy is less but by using AFP lesser number of partitions is made, so the extracted watermark accuracy is not degraded much.

The watermark embedding algorithm is as follows:

*1. Input OT and WM.*
*2. Preprocess WM to make it in pure alphabetical form. .*
*3. Split the text in words.*
*4. Count AFP from the text.*
*5. Divide the OT into partition based on AFP.*
*6. Calculate the MONV in each partition.*
*7. Generate the key by first placing the AFP in the key and further populate key as follow:*
*Repeat for all watermark characters*
*If WM [ i ] belongs to MONV list,*
*then Key [ j ] = 0;*
*Key [ j+1 ] = of that MONV.*
*if PN repeats then pick the next partition of that character*
*else if WM [ i ] does not belong to MONV list,*
*then Key [ j ]= 1;*
*Key [ j+1 ] = (C(Classical cipher))*
*h. Output Key.*
***OT (Original Text); WM (Watermark); AFP (Average frequency Preposition); MONV (least occurring non-vowel character); PN (Partition number)***

In this algorithm first AFP is identified from the text. This AFP is then used to create partitions of the text. In the next step, the frequency of each non vowel character is calculated in each partition and prepared the MONV list by selecting MONV in every partition. Further, the key generation process generates the author's key by using the letters of watermark and MONV list as stated in the above algorithm. Finally this key is timed stamped and registered with the CA along with the original watermark. This proposed algorithm is specifically for copyright protection of medium size documents such as emails, newspaper articles, blogs, etc.

### *B. Extraction Algorithm*

When copyright clash occurs then the extraction algorithm is used. This algorithm extracts the watermark from the noisy text so it is known as extraction algorithm. CA uses the extraction algorithm to identify any type of copyright clash. The author key is used to extract the watermark from the tempered text using this algorithm. After that both, the original and the extracted watermarks are compared to calculate the watermark accuracy in order to judge the original owner of the text document. The general idea of watermark extraction process is shown in fig. 2.

First, this algorithm reads the author key and creates the partition of text based on AFP obtained from the author registered key. In the next step, the occurrence of each non vowel character is counted in each partition and list is generated by identifying the maximum occurring non vowel characters as done in the previous embedding algorithm. The watermark is obtained by using contents of author's key as explained earlier in step 5 of the extraction algorithm. Finally, CA decides who the original author is and resolves the copyright clash.

The complete extraction algorithm is as follows:

*1. Input AT and key.*
*2. Tokenize AT into words.*
*3. Split AT into partition based on the AFP obtained from the key.*
*4. Calculate the MONV in each partition.*
*5. Generate WM and populate WM as follows (example)*
*if (key = = 015)*
*Pick the MONV from 15 partitions and place it in WM e.g.: k.*
*if (key = = 1g)*
*Place the character using reverse classical cipher in WM e.g.: x.*
*6. Output WM*

***WM (watermark); AFP (Average frequency preposition); MONV (least occurring non-vowel character); AT (Attack text file)***

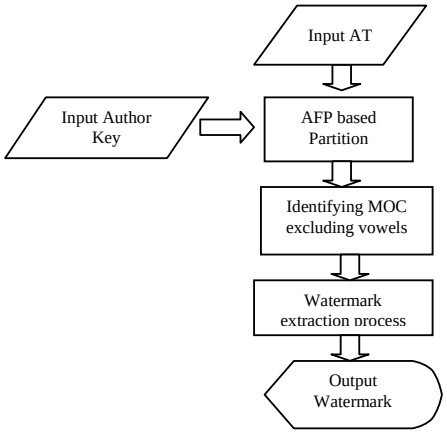


Fig. 2 Watermark extraction process

## IV. Experimental Results

In order to examine the performance of the proposed algorithm, we selected an article from the web [23] as a sample to perform different types of attacks by five different individuals. The sample text was given to five different individuals to make meaning preserving intelligent attacks. The individuals were selected randomly with different English language proficiency and educational background. Five attacked samples were obtained. The statistics of the original text ample and of each of the attack text samples are shown in table 1, where attacked samples are named as A1, A2, A3, A4, and A5.

TABLE 1
Details of Original File and Samples

| Attributes | Original | A1 | A2 | A3 | A4 | A5 |
|---|---|---|---|---|---|---|
| Words | 351 | 273 | 315 | 290 | 174 | 294 |
| Characters | 1990 | 1560 | 1781 | 152 | 1042 | 1618 |
| Sentences | 18 | 15 | 14 | 13 | 11 | 12 |
| Preposition | 44 | 43 | 45 | 22 | 21 | 38 |

Attack on the original file can only alter the characteristics of the original file but the whole theme of text remains same. When the attacker intends to violate the copyrights, he/she will perform intelligent attacks in order to retype or reproduce the text and all the attacked samples were varied based on attack volume. Furthermore, in this algorithm we used two

watermarks of different lengths, in order to check the alteration occurred by the characteristic of the watermark. The features of both watermarks are shown in table 2.

TABLE II
DETAILS OF WATERMARKS

| Attributes | Watermark1 | Watermark2 |
|---|---|---|
| Words Count | 112 | 40 |
| Characters Count | 717 | 269 |
| Sentences Count | 6 | 2 |
| Preposition Count | 12 | 3 |

Both the watermarks which we have used in our evaluation are as follows:

Watermark 1

"*The abrupt way in which President Pervez Musharraf announced the decision to build the dam, overruling the objections of the smaller provinces, has polarized public opinion. On 26 May 2008, Federal Minister for Water and Power of Pakistan Raja Pervez Ashraf has said that Kalabagh Dam will not be constructed. He said due to opposition from Khyber Pakhtunkhwa, Sindh and other stakeholders, the project are no longer feasible. The Prime Minister of Pakistan, Syed Yousuf Raza Gillani announced that the fate of the project would be decided by a plebiscite. The decision came after Pakistan faced extreme power crisis and acute water shortages. The government is currently finding alternative locations for the dam.*"

Watermark 2

"*Pakistan has a rich and vast natural resource base, covering various ecological and climatic zones; hence the country has great potential for producing all types of food commodities. Agriculture has an important direct and indirect role in generating economic growth* ".

After attacking the original text, watermark had been extracted from the attacked files. It is clearly shown that the most of the attacked files are suffered from deletion attack because in table 1, all the characteristics are decreased as compared to the original one. After the extraction of watermark from the attacked samples, it is revealed that when watermark1 is used the results are up to the mark. It is also observed that on one hand we are decreasing the length of watermark, and on the other hand watermark accuracy is also decreasing. So for better results a watermark of appropriate length is needed. The watermark accuracy on all attack samples and their average results are shown in table 3.

TABLE III
DETAILS OF ACCURACY OF WATERMARKS

| Samples | Watermark 1 | Watermark 2 |
|---|---|---|
| A1.txt | 70.73 % | 65.74 % |
| A2.txt | 65.83 % | 60.33 % |
| A3.txt | 93.51 % | 85.16 % |
| A4.txt | 68.28 % | 66.66 % |
| A5.txt | 81.28 % | 58.23 % |
| **Average** | **75.92 %** | **67.22 %** |

The watermark accuracy of the attacked samples are manipulated in different figures both for watermark1 and watermark2 are shown in fig.3 and fig.4 respectively The analyses of fig.3 shows that watermark1 gives the accuracy up to 79.92%. This shows that if attacks are made intelligently and watermark used is of appropriate length, then this algorithm gives the best results to solve the copyright protection issues especially in medium size files such as articles etc.

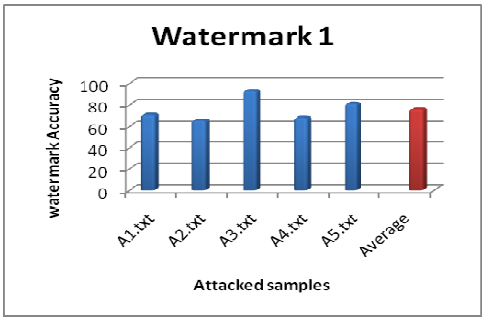


Fig. 3 Details of Accuracy of Watermark1

Further analysis was made by changing the length of watermark because it may be possible that changing the watermark results would be effected. So, it is clearly seen in the fig.4 that the watermark accuracy of short length watermark is less than the long one.

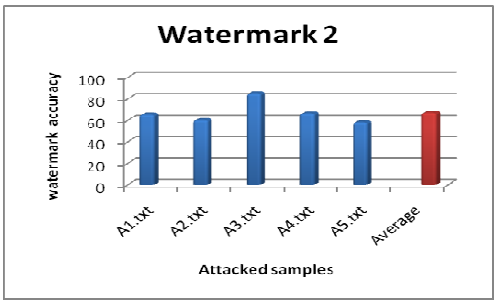


Fig. 4 Details of Accuracy of Watermark2

## V. CONCLUSION

The existing text watermarking solutions are not robust against intelligent meaning preserving attacks. In this paper, we have proposed a text watermarking algorithm using occurrence frequency of prepositions and non-vowel ASCII characters to safeguard text from the attackers. The embedding algorithm embeds the textual watermark in text and generates a key. This key is later used by a CA to extract that watermark. We evaluated the performance of the algorithm for intelligent meaning preserving attacks on five different attacked samples obtained by five different attackers using two watermarks. The results show that our algorithm is robust, secure, and efficient against intelligent tampering attacks for both watermarks.

## REFERENCES

[1] A. Khan, A. M. Mirza and A. Majid, Optimizing Perceptual Shaping of a Digital Watermark Using Genetic Programming, Iranian Journal of Electrical and Computer Engineering, vol. 3, pp. 144-150, 2004.
[2] J. T. Brassil, S. Low, N. F. Maxemchuk, and L. O'Gorman, Electronic Marking and Identification Techniques to Discourage Document Copying, IEEE Journal on Selected Areas in Communications, vol. 13, no. 8, pp. 1495-1504, October 1995.
[3] J. T. Brassil, S. Low, and N. F. Maxemchuk, Copyright Protection for the Electronic Distribution of Text Documents, Proceedings of the IEEE, vol. 87, no. 7, pp.1181-1196, July 1999.
[4] N. F. Maxemchuk, S. H. Low, Performance Comparison of Two Text Marking Methods, IEEE Journal of Selected Areas in Communications (JSAC),vol. 16 no. 4 1998. pp. 561-572, May 1998.
[5] N. F. Maxemchuk, "Electronic Document Distribution," AT&T Technical Journal, September 1994, pp. 73-80. 6.
[6] N. F. Maxemchuk and S. Low, Marking Text Documents, Proceedings of the IEEE International Conference on Image Processing, Washington, DC, , pp. 13-16, Oct. 26-29, 1997.
[7] S. H. Low, N. F. Maxemchuk, and A. M. Lapone, Document Identification for Copyright Protection Using Centroid Detection, IEEE Transactions on Communications, vol. 46, no.3, pp 372-381, Mar. 1998.
[8] S. H. Low and N. F. Maxemchuk, Capacity of Text Marking Channel, IEEE Signal Processing Letters, vol. 7, no. 12 , pp. 345 -347, Dec. 2000.
[9] D. Huang and H. Yan, Interword distance changes represented by sine waves for watermarking text images, IEEE Trans. Circuits and Systems for Video Technology, Vol.11, No.12, pp.1237-1245, Dec 2001.
[10] M. J. Atallah, C. McDonough, S. Nirenburg, and V. Raskin, Natural Language Processing for Information Assurance and Security: An Overview and Implementations, Proceedings 9th ACM/SIGSAC New Security Paradigms Workshop, Cork, Ireland, pp. 51–65, September, 2000.
[11] M. J. Atallah, V. Raskin, M. C. Crogan, C. F. Hempelmann, F. Kerschbaum, D. Mohamed, and S.Naik, Natural language watermarking: Design,analysis, and a proof-of-concept implementation, Proceedings of the Fourth Information Hiding Workshop, vol. LNCS 2137, Pittsburgh, PA, 25-27 April 2001.
[12] H. M. Meral et al., Natural language watermarking via morphosyntactic alterations, Computer Speech and Language, 23, 107-125, 2009.
[13] H. M. Meral, E. Sevinç, E. Ünkar, B. Sankur, A. S. Özsoy, T. Güngör, Syntactic tools for text watermarking, 19th SPIE Electronic Imaging Conf. 6505: Security, Steganography, and Watermarking of Multimedia Contents, San Jose, Jan. 2007.
[14] M. Atallah, V. Raskin, C. F. Hempelmann, M. Karahan, R. Sion, U. Topkara, and K. E. Triezenberg, Natural Language Watermarking and Tamperproofing, Fifth Information Hiding Workshop, vol. LNCS, 2578, Noordwijkerhout, The Netherlands, Springer-Verlag, October, 2002.
[15] U. Topkara, M. Topkara, M. J. Atallah, The Hiding Virtues of Ambiguity: Quantifiably Resilient Watermarking of Natural Language Text through Synonym Substitutions, In Proceedings of ACM Multimedia and Security Conference, Geneva, 2006.
[16] X. Sun, A. J. Asiimwe. Noun-Verb Based Technique of Text Watermarking Using Recursive Decent Semantic Net Parsers. Lecture Notes in Computer Science (LNCS) 3612: 958-961, Springer Press, August 2005.
[17] M. Topkara, U. Topraka, and M.J. Atallah, Information hiding through errors: a confusing approach. Proceedings of SPIE Security, Steganography, and watermarking of Multimedia Contents IX., pp. 65050 V-1–65050V-12.
[18] B. Macq and O. Vybornova, A method of text watermarking using presuppositions, in Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents, January 2007.
[19] Peng Lu et al., An optimized natural language watermarking algorithm based on TMR, Proceedings of 9th International Conference for Young Computer Scientists, 2009.
[20] Z. Jalil and A. M. Mirza, "An Invisible Text Watermarking Algorithm using Image Watermark", International Conference on Systems, Computing Sciences, and Software Engineering (SCSS 2009), Innovations in Computing Sciences and Software Engineering, published by Springer, ISBN: 978-90-481-9111-6.
[21] Z. Jalil, A. M. Mirza, and T. Iqbal, "A Zero-Watermarking Algorithm for Text Documents using Structural Components", International Conference on Information and Emerging Technologies (ICIET 2010), June 14-16, 2010, Karachi, Pakistan.
[22] Z. Jalil, M. Farooq, M. Arif and A. M. Mirza, "A Zero Text Watermarking Algorithm Using Non-Vowel Alphabets", International Journal of Electrical, Computer, and Systems Engineering (ICCESSE 2010), November 24-26, 2010, Venice, Italy.
[23] DAWN news website article link: http://epaper.dawn.com/ArticleText.aspx?article=01_08_2010_006_006

**Jalil Z.** was born in Islamabad, Pakistan on 23rd July, 1980. She received her B.Sc. degree from Punjab University, Lahore, Pakistan in 1999. She later received M.Sc. degree in Computer Science in 2002 from International Islamic University, Islamabad, Pakistan. Then she earned her M.S. degree in Computer Science in 2007 from FAST National University of Computer and Emerging Sciences, Islamabad, Pakistan. Currently she is pursuing her Ph. D research at Department of Computer Sciences, National University of Computer and Emerging Sciences, Islamabad, Pakistan. Her research interest includes text watermarking, information security, copyright protection and privacy.