International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 10 (2015) pp. 26483-26498

http://www.ripublication.com

A Zero Text Watermarking Algorithm for Protecting Plain Text Documents using Combined Image and Text through Compression and Encryption

Jaseena K.U. [a], Murugan R.[b,c] , Dr. John T Abraham[d]

a. Assistant Professor, Department of Computer Applications, M.E.S. College, Marampally, Kochi, India jaseena.mes@gmail.com
b. Research and Development Centre, Bharathiar University, Coimbatore, Tamilnadu, India.
c. Associate Professor, Department of Computer Applications, M.E.S. College, Marampally, Kochi, India mes.murugan@gmail.com
d. Assistant Professor, Bharata Mata College, Thrikkakara, Kochi, India

Abstract

Internet has become the most popular medium for transferring various digital forms of multimedia data. By the advent of Internet, digital data can be modified and illegally used with ease. Every day, huge amount of information is transferred over the Internet. Thus, the intellectual property right of digital data, which is transmitted over the Internet, requires more attention. Digital watermarking provides authentication and copyright protection for digital contents over the Internet. Text is an important medium for transferring information over the Internet, which requires safeguards. In this paper, there is an attempt which suggests a text watermarking algorithm that uses a combined image plus text watermark and compression as well as encryption techniques to protect the text document. Initially the watermark is embedded in the text and a watermark key is generated by using the properties of text. In order to increase the efficiency of text document, the document is compressed using Huffman Encoding Algorithm. The text document is then encrypted by using RSA algorithm to enhance the security of text. The reverse of the above steps is done to extract the watermark. Experimental results demonstrate the effectiveness of the proposed algorithm.

Keywords: security, watermarking, authentication, watermark embedding and extraction, compression, Huffman coding, encryption, RSA cryptographic algorithm.

1. INTRODUCTION

The rapid growth of Internet has increased the use of electronic data and the development of various Internet technologies and has eased the modification and illegal distribution of digital contents. Thus, the copyright protection of digital data transmitted over the Internet has become an important and conspicuous research topic in recent years. Current copyright laws are not adequate for dealing with digital data over the Internet [6]. This has led to the development of watermarking techniques. In the past few years, many watermarking algorithms had been proposed for authentication and copyright protection of digital contents.

Digital watermarking is a technique for inserting information into an image or text or audio, which can be later extracted for various purposes including identification and authentication [7]. A digital watermark is a pattern of bits inserted into a digital image, audio, video or text file that identifies the copyright information. It can be a unique logo or signature of an individual, who owns the copyright of a digital content [4]. The process of embedding a watermark in a text that uniquely identifies the original copyright owner of the text is called Digital Text Watermarking. Text watermarking protects digital content from copyright violations, illegal copying, tampering, redistributions, and other similar infringements.

Text documents are easier to copy, reproduce and tamper as compared with images, audios and videos and thereby are prone to different types of attacks [8]. Hence, text needs to be fully protected. Text watermarking techniques not only prevent the text from illegal copying, forgery, and redistribution, but also help to protect copyright. Text watermarking algorithms use either an image watermark or a textual watermark. But watermarks consisting of both image and text make the text more secure and has better robustness [1]. So in this paper we have used a combination of both image and text watermark instead of using plain textual or image watermark in order to achieve better robustness.

The paper is organized as follows. The introduction section is followed by the fundamental concepts of digital watermarking in Section 2. A description of the proposed watermarking (embedding and extraction) algorithm is presented in Section 3. The experimental results and its interpretations are given in section 4 and finally section 5 concludes the paper.

2. DIGITAL WATERMARKING

Information hiding techniques came up to safeguard the information from illegal usage, illegal copying, redistribution of copyrighted documents, tampering, forgery, and authentication. Cryptography, steganography, and watermarking emerged as sub disciplines of information hiding to serve particular purposes. Digital watermarking provides authentication and copyright protection for digital contents over the Internet.

Based on human perception, watermarks can be classified into visible (perceptible) watermarking and invisible (imperceptible) watermarking [4]. Visibility is a term associated with the perception of the human eye. In visible watermarking, watermarks are embedded in such a way that they are visible when the content is viewed. Invisible watermarks cannot be seen but they can be recovered with an appropriate decoding algorithm. Invisible watermarks are more robust than visible watermarking. Invisible watermarks can again be robust or fragile. Robust watermarking is a technique, in which modification to the watermarked content will not affect the watermark in any way. As opposed to this, in fragile watermarking, watermark gets destroyed when watermarked content is modified or tampered with.

Watermarking can also be classified based on the type of document to be watermarked [5]. The classifications include Image Watermarking, Video Watermarking, Audio Watermarking and Text Watermarking. The process of embedding a digital watermark into a digital text document, that carries information unique to the copyright owner or the creator of the document, is called Digital Text Watermarking.

Text watermarking is an emerging area of research. The text watermarking algorithms which have been developed so far can be classified into the following categories [1].

1. Image based methods
2. Syntactic methods
3. Semantic methods
4. Structural methods.

2.1 Image based methods

In image-based methods of text watermarking, the binary watermarks are embedded in text image. Brassil, et al. [10, 11] was the first person to propose text watermarking schemes using text image and binary watermark. He proposed a few methods to watermark text documents using text image. The technique was the line shift algorithm which moves a line upward or downward (left or right) depending on binary signal to be inserted. The detection algorithm was non-blind, in which the original document must be available. The second method was the word shift algorithm, which moves the words horizontally thus expanding spaces to embed the watermark. The algorithm can operate in both blind and non blind modes. The third method is feature coding algorithm, in which certain text features are altered to insert watermark bits in the text. After that, Maxemchuk, et al. [15-17] and Low, et al. [18, 19] analyzed the performance of these methods.

Then, an algorithm based on an average inter-word distance in each line was proposed by Huang and Yan [20]. The distances are adjusted according to the sine wave of a specific phase and frequency. Thereupon, the feature and the pixel level algorithms were proposed, in which it marked the documents by altering the stroke features such as width or serif [29]. The algorithm, which uses the gray scale image of a text was also proposed [30]. Another algorithm, which watermarks the text document image using edge direction histogram was also developed [31].

A text watermarking algorithm, based on word classification and inter-word space statistics was proposed by Young Won Kim et al. [32]. In this algorithm all the words in a text are classified depending on some text features and adjacent words and are grouped to form a segment. The segments are classified based on class labels of the words within the segment. Adnan M. Alattar et al. developed an algorithm to watermark electronic text documents containing justified paragraphs and irregular line spacing [33]. The algorithm that exploits the printed text document to identify the source printer was also proposed [34].

All the above text watermarking algorithms which are based on text image are not robust against text reproduction and copy paste to notepad attacks [2]. Now a days with effective and increasing use of OCR (Optical Character Recognition), these methods are a failure. The use of OCR can destroy the changes made by shifting words upward and downward, to the document margins, to the fonts, serif and features of text. Hence all changes made in text structure, feature and font to embed watermark will get destroyed.

2.2 Syntactic methods

In syntactic methods, the syntactic structure of a text is utilized to embed the watermark. Mikhail J. Atallah, et al. was the first person, who proposed the natural language watermarking scheme by using syntactic structure of text [12,13]. In this method, the syntactic tree is built and transformations are applied to embed the watermark. All inherent properties are maintained during watermark embedding. Hassan, et al. utilized morphosyntactic alterations to watermark the text [21]. The text is first converted into a syntactic tree diagram and watermark is embedded. He stated that agglutinative languages like Turkish are easier to watermark than English language. At this juncture [22], Hassan presented an overview of syntactic tools available for text watermarking and also developed a syntax based natural language watermarking method. The syntactic solutions are more suitable to agglutinative languages like Korean, Turkish and Arabic.

Text watermarking algorithms using syntactic structure combined with natural language processing (NLP) algorithms, is an efficient approach for text watermarking. But NLP is very low [2].

2.3 Semantic methods

In semantic schemes, the watermark embedding is done by utilizing the semantics of text. Atallah et al. proposed an algorithm in 2000 using semantic watermarking techniques [23]. Then algorithm based on synonym substitution was proposed by Topkara, et al [24]. In [25], another algorithm, based on nouns and verbs was proposed for text watermarking. Topkara, et al. developed another algorithm using abbreviations, acronyms and typos to embed the watermark in text watermarking [26]. An algorithm based on presupposition was proposed in [27]. Presuppositions are identified and then transformations like passivization, topicalization, extraposition and preposing are applied to embed watermark in the text. The text pruning and grafting algorithms are also proposed in the past. In [28], another algorithm was proposed using Text Meaning Representation (TMR).

The text watermarking methods based on semantics are language dependent. The synonym based methods are not resilient to the random synonym substitution attacks. Sensitive texts like legal documents, poetry, and quotes are difficult to watermark [2].

The above mentioned algorithms are not applicable to all types of text documents under random tampering attacks and are not designed specifically to solve tamper detection problems. Hence to avoid the above mentioned drawbacks, new methods, which use text structures to embed watermarks, have been proposed.

2.4 Structural methods

Structural schemes of text watermarking are the recently used watermarking approach which uses the text structures to embed watermarks. In this scheme, the text is not altered when the watermark is embedded into it. These types of text watermarking schemes are called zero watermarking [3]. Many text watermarking techniques utilizing existence of double letters in the text have been proposed for protecting text documents [1,4, 9]. Z. Jalil et al. proposed a zero text watermarking algorithm based on structural approach for protecting text documents [2].

Text watermarking solutions are not robust against random tampering attacks such as insertion, deletion and re-ordering attacks. In this paper, we propose a zero text watermarking algorithm based on non-vowel ASCII characters that uses combined image and text watermark to protect the text document.

3. PROPOSED ALGORITHM

In this paper, we propose an invisible zero watermarking algorithm, which uses combined image and text watermark to protect the text document. In this algorithm, the occurrences of non-vowel ASCII characters existing in the text are utilized to embed the watermark [2]. The original owner of text logically embeds the watermark in a text and generates a watermark key. The watermark key along with the watermarks is kept with the Certification Authority (CA), where the original author is registered. If any violation to the copyrights of the owner occurs, the generated key is used to identify the original owner and thus resolve conflicts. This is done by using the extraction algorithm. Sometimes, there is a chance of more than one claim for a single text. In that case, the one that was registered earlier will be considered as the original one.

A compression scheme is also applied to the text documents. This not only reduces the size of the document but also increases the speed of data transfer. Here we use Huffman Encoding Algorithm. The text document is then encrypted using RSA algorithm to enhance the security of text document. The reverse of the above steps is done to extract the watermark. The extraction of watermarks involves three steps. 1) The cipher text is decrypted to get compressed plain text document back. 2) The compressed document is decompressed to retain plain text document. 3) The watermark is extracted using watermark key from the plain text with the help of watermark extraction algorithm. After extraction, the original and extracted watermarks are compared to calculate the watermark accuracy in order to prove authenticity of the original owner of the text document.

In the proposed algorithm, the text is first partitioned based on partition size (Pr). Pr is considered as a delimiter to form text partitions. Depending on the value of GS (Group Size), partitions are combined to form text groups. Then the occurrence of all non-vowel ASCII characters is calculated in each group and maximum occurring non-vowel ASCII character is identified in each group to create MONV (Maximum Occurring Non-Vowel) list. This MONV list and combined image and text watermark is used to generate the watermark key. This watermark key is registered with a certification authority (CA), a trusted third party for copyright protection. The watermarks and watermark key is kept with the CA along with time and date. Later this key is used in the extraction algorithm to identify the original owner. In general, the watermarking process involves two stages, Watermark Embedding and Watermark Extraction.

3.1 Steps of Embedding Process

The steps of embedding process are shown in Figure 1. The algorithm which is used to embed the watermark in the text and to generate watermark key is called embedding algorithm. It takes the combined image and text watermark as input, produces watermark key and compressed and encrypted watermarked text as output.

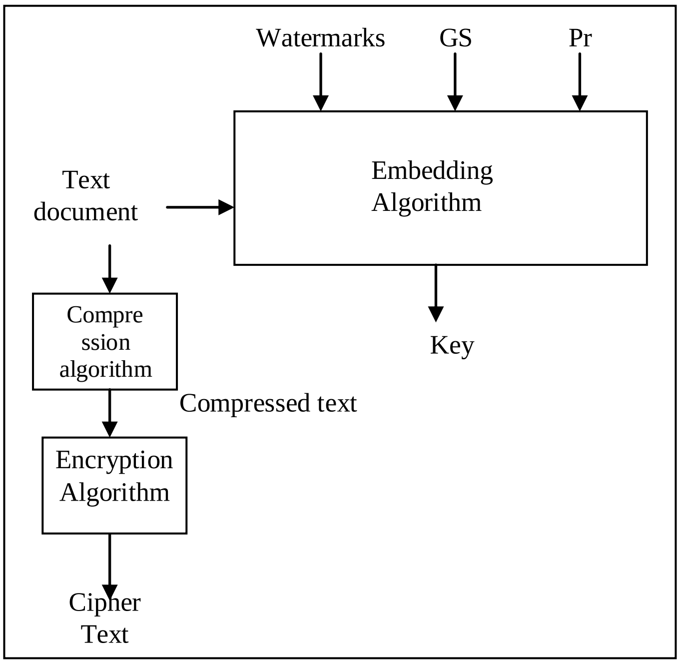

Figure. 1. Steps of embedding process

1. Inputs are combined image plus text watermark and text document.
2. Preprocess text which includes, discarding white spaces, special characters, digits etc to make the watermark pure alphabetical.
3. Preprocess image which converts image to grayscale and scaling to standard size (100 x 100 pixels).
4. Then Convert image to plain text by normalization process.
5. The two textual watermarks (watermarks obtained after text preprocessing and image preprocessing) and partial key containing a partition size (Pr) and group size (GS) is given as input to the embedding algorithm.
6. Generate a watermark key using properties of text (using algorithm described below).
7. Compress the text document and watermark key using Huffman coding algorithm to reduce the size of the information to be transmitted.
8. Encrypt the text document using RSA encryption algorithm to increase security of text.

Preprocessing of text is the process of removing white spaces, special characters, digits etc to make the watermark pure alphabetical. During image pre processing, image is first converted in to grey scale and then scaling to 100x100 pixels. After image pre processing, image is converted in to plain text by normalization process. The two textual watermarks (watermarks obtained after text preprocessing and image preprocessing), partition size (Pr) and group size (GS) is given as input to the embedding algorithm.

3.1.1 Algorthm: Watermark Embedding

The algorithm used for embedding watermark as in [2] is presented below.

1. Input W, GS, Pr and T.
2. Split W into W_{Img} and W_{Txt}
3. Preprocess W_{Img} and W_{Txt}
4. Convert W_{Img} to text WT
5. Merge WT and W_{Txt} to generate W.
6. Make partitions of T based on Pr
7. Make groups of text based on GS,where No.of groups = No. of partitions/GS
8. Count occurrence of non-vowel ASCII characters in each group and find Maximum Occurring Non-Vowel (MONV) in each group
9. Generate Watermark Key using steps from 10 to 12.
10. While (j<watermark_length) repeat step 11 to 12
11. if (W_j€ MONV list)

    Key(i)=0,key(i+1)= Groupnumber(MONV)List

    else

    Key(i)=1,Key(i+1)=(W_j+k)MOD26, the Cipher text where k is in Z26 and Z26 represents 26 alphabets(a- z)
12. Increment i
13. Output Key

W: Combined watermark, W_{Img}: image watermark, W_{Txt}: text watermark, GS: Group size, Pr: Partition, T: text file, WT: text watermark (image to text), Key: Watermark Key

First, the watermark is split into image and text watermarks. The preprocessing of text and image watermarks is done to make the watermark pure alphabetical (W_{Txt} and WT). The text partitions are made based on Pr and based on GS; they are combined to form text groups. Then, the occurrence of all non-vowel ASCII characters is calculated in each group and maximum occurring non-vowel ASCII character is identified in each group to create MONV (Maximum Occurring Non-Vowel) list. Using this MONV list and textual watermarks, the watermark key is generated using the algorithm described above. Then the watermark key is registered with a certification authority (CA), a trusted third party for copyright protection.

3.2 Steps of Extraction Process

The steps of extraction process are shown in figure 2. The algorithm which extracts the watermarks using watermark key is called extraction algorithm. It takes the watermark key and encrypted and compressed watermarked text as input and extracts the watermark (image and text) from the text.

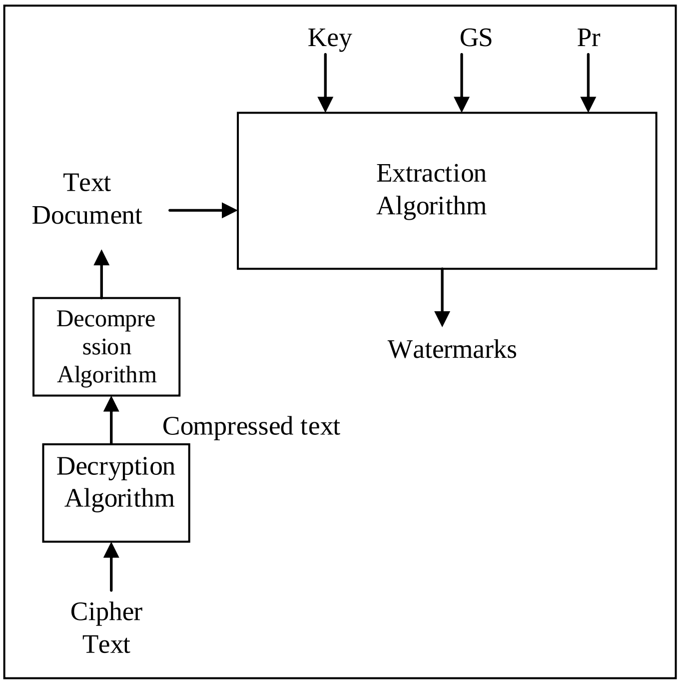

Figure. 2. Steps of extraction process

1. The encrypted and compressed watermarked text is the inputs.
2. Decrypt the text document using RSA algorithm.
3. Decompress watermark key and decompress watermarked text using Huffman coding algorithm.
4. Partition the text using Pr from watermark key.
5. Combine the partitions to make text groups as done previously in the embedding algorithm.
6. Occurrence of non vowel ASCII characters in each group is analyzed and maximum occurring non vowel ASCII character (MONV) in each group is identified.
7. Using the watermark key, watermarks are extracted from the text with the help of extraction algorithm.

3.2.1 Algorithm: Watermark Extraction

The algorithm used for extracting watermark as in [2] is given below.

1. Input Key and T.

2. Read Pr and GS.

3. Make partitions of T based on Pr

4. Make groups of text based on GS i.e. Numberofgroups=Numberof partitions/GS

5. Count occurrence of non-vowel ASCII characters in each group and find Maximum Occurring Non-Vowel(MONV)

6. L=length(Key), I=1, j=1

7. While (I<L) repeat 8 to 9

8. If (Key (I) equal 0)

   W(j)=groupnumber(MONV)

   else

   W(j)= Key(I+1) i.e. reverse cipher letter

9. I=I+1, j=j+1

10. Split W into W_{Img} and W_{Txt}

11. Output W_{Img} and W_{Txt}

Texts are partitioned and grouped based on Pr and GS value respectively. Occurrence of non-vowel ASCII characters is analyzed and maximum occurring Non-Vowel ASCII character (MONV) in each group is identified. Using the algorithm described above, watermarks are extracted.

3.3 Compression & Decompression Process

After generating the watermark key, the original text document and watermark key are compressed so that the size of text and key can be reduced. Compression not only reduces the size of the document but also increases the speed of data transfer. A lossless compression algorithm is preferred since we are dealing with text documents. The proposed system utilizes Huffman coding algorithm to compress text document and watermark key. The decompression process converts the compressed text back to normal form.

The Huffman Coding algorithm was proposed by David A Huffman in 1952 and it makes use of bottom up technique. It is a lossless compression algorithm and it uses variable length coding. It is a prefix code because no Huffman code is a prefix of another Huffman code. In this method more frequently appearing symbols are coded with fewer bits per symbol [14]. So only fewer bits are needed to represent the whole information. The algorithm [14] is given below.

1. Sort all symbols on the list according to their frequency count.
2. Repeat the following steps until the list has only one symbol left
   a. From the list , pick two symbols with lowest freq count
   b. Form Huffman sub tree with these two symbols as child nodes and create a parent node for them.
   c. Assign the sum of children's freq count to parent and insert it in to the list

d. Delete the children from the list

3. Assign a code word to each leaf based on the path from the root.

3.4 Encryption and Decryption

In cryptography, encryption is the process of converting the plain text information into an unintelligible format called cipher text. By encryption the confidentiality of the information can be maintained. Here, we have used RSA algorithm which is simple as well as strong [35].

3.4.1 RSA Algorithm

The RSA algorithm is named after Ron Rivest, Adi Shamir and Len Adleman, who invented it in 1977. RSA Algorithm is based on integer factorization assumption. The RSA algorithm can be used for both public key encryption and digital signatures. The important characteristic of RSA is that it is computationally infeasible to determine the decryption key given only the knowledge of the cryptographic algorithm and the encryption key. RSA involves a public key and a private key. The public key can be known to everyone and is used for encrypting messages. Messages encrypted with the public key can only be decrypted using the private key. It consists of the following procedures: key generation, encryption, and decryption [35].

3.4.1.1 Key Generation

1. Choose two big primes: p and q;
2. Calculate $n=p*q$;
3. Randomly choose an integer e, satisfying $1<e<\varphi(n)$, $\gcd(e, \varphi(n))=1$. Totient function $\varphi(n)$ denotes the number of positive integers less than n and relatively prime to n. Here $\varphi(n) = (p-1)*(q-1)$. The public key is (e, n);
4. Calculate d, satisfying $ed \bmod \varphi(n) = 1$, the private key is (d, n);

3.4.1.2 Encryption Procedure

1. Partition the message m to groups mi, i=1,2,…, $|mi|= |n|-1$; (|a| means the length of a in binary form)
2. Encrypt each group $c_i = m_i^e \bmod n$.
3. Connect each c_i and get the cipher text c.

3.4.1.3 Decryption Procedure

1. Partition c to c_i, i=1,2,…, $|ci|= |n|-1$;
2. Decrypt each c_i : $m_i = c_i^d \bmod n$
3. Connect each m_i and recover the plain text m.

4. EXPERIMENTAL RESULTS

We have used different values for Pr for our experiments. Group size was taken as 5 in all experiments. The combined image and text watermark used in experiments is shown in Figure3.

Figure. 3. Combined image and text watermark used in experiments

Table I shows the accuracy of extracted watermark for image, text and combined image and text watermarks under tampering attacks when algorithm based on non-vowel ASCII characters[2] are used. For comparing the accuracy of extracted watermarks, we have taken five values for Pr as 100,120,140,160 and 180. The graph corresponding to Table I is shown in Figure 4.

TABLE I. ACCURACY OF EXTRACTED WATERMARK (IMAGE, TEXT AND OVERALL) UNDER RANDOM TAMPERING ATTACK

Pr	**Text %**	**Image %**	**text+image%**
100	96.05	99.66	98.67
120	80.13	99.67	89.71
140	96.16	99.72	98.39
160	97.06	99.66	98.53
180	99.99	100	100
Average	93.88	99.74	97.06

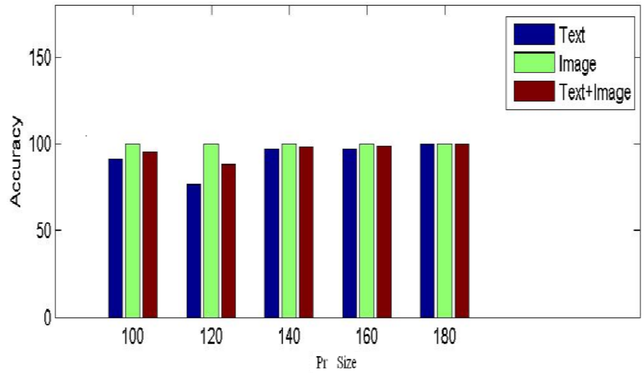

Figure. 4. Graph Corresponding to Table I

From Table I, we can conclude that average accuracy of extracted textual watermark is 93.88% and that textual watermark is more sensitive to tampering attacks (insertion, deletion and reordering attacks) than image watermark. Hence the accuracy of text is lesser than image. But the combined accuracy (average) is 97.06%.

TABLE II. LENGTH OF TEXT DOCUMENT AND KEY BEFORE AND AFTER COMPRESSION

Item	Length Before Compression	Length After Compression
Text taken	2135	1172
Key	20088	8381

Table II shows the length of the text and key before and after compression. The total length of the original key without compression is 20088. But after applying Huffman coding for compression, the total length of the key is reduced to 8381.

5. CONCLUSION AND FUTURE SCOPE

Digital Watermarking is an emerging research area for copyright protection and authentication of electronic documents and media. In this paper, a new watermarking technique is proposed that uses combined image and text watermark to prove authenticity of a text document. To enhance efficiency and security, a lossless compression using Huffman Coding as well as an encryption technique using RSA, are also used. This method proves to be a better way not only to prove the authenticity of the transmitted documents, but also to resolve conflicts related to copyright violation. The proposed algorithm is more resistant to random tampering attacks. Experimental results show that the proposed algorithm is efficient and has better performance.

The security of text can be enhanced by using another powerful encryption technique. The more powerful the encryption algorithm is, then it becomes more difficult for an attacker to decode the original information. According to the proposed algorithm, the generated watermark key is long. The disadvantage of a lengthy key is that it will be difficult for the CA to maintain key and also transfer of key between owner of text and CA will not be easy. As a future work, some measures other than compression may be considered to reduce the key length.

REFERENCES

[1] Z.Jalil, A.M. Mirza ,"Text Watermarking Using Combined Image-plus- Text Watermark" ,IEEE, 2010.

[2] Z.Jalil, H.Aziz, S.B.Shahid, M.Arif, A.M.Mirza," A Zero Text watermarking Algorithm based on Non-Vowel ASCII characters",IEEE, 2010.

[3] Z.Jalil,., Farooq M., Zafar H., Sabir M., and Ashraf E.," Improved Zero Text watermarking Algorithm against Meaning Preservation Attacks", World Academy of Science, Engineering and Technology, 2010.

[4] 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.

[5] M. Chandra, S. Pandey, R. Chaudhary, "Digital Watermarking Techniques for Protecting Digital Images", IEEE, 2010.

[6] Ketki Kundu, Shashi Bhushan,"Digital Video Watermarking for Copyright over Internet", Vol. 1, Spl. Issue 2 (May, 2014) e-ISSN: 1694-2329 | p-ISSN: 1694-2345.

[7] Z. Xiao-hua1,M. Hong-yun,L. Fang, "A New Kind of Efficient Fragile Watermarking Technique",Acta Electronica Sinica, 2004.

[8] Z. Jalil, A. M. Mirza ,M. Sabir "Content Based Zero Watermarking Algorithm for Authentication of Text Documents" ,International Journal of Computer Science and Information Technology,V 7, 2010.

[9] Jassena K.U., Anita John, "Text Watermarking using Combined Image and Text for Authentication and Protection", In International Journal of Computer Applications (0975 – 8887), Volume 20 – No.4, April 2011.

[10] J. T. Brassil, S. Low, N. F. Maxemchuk, 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.

[11] J. T. Brassil, S. Low, 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.

[12] M. J. Atallah, C. McDonough, S. Nirenburg, V. Raskin, “Natural Language Processing for Information Assurance and Security: An Overview and Implementations”, Proceedings of 9th ACM/SIGSAC New Security Paradigms Workshop, Cork, Ireland, pp. 51–65, September, 2000.

[13] M. J. Atallah, V. Raskin, M. C. Crogan, C. F. Hempelmann, F. Kerschbaum, D. Mohamed, S.Naik, “Natural language watermarking: Design,analysis, and a proof-of-concept implementation”, Proceedings of the Fourth Information Hiding Workshop, vol. LNCS 2137, Pittsburgh, April, 2001.

[14] Z. Li, M.S.Drew,” Fundamentals of Multimedia”, Prentice Hall of India, 2010.

[15] 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.

[16] N. F. Maxemchuk, “Electronic Document Distribution,” AT&T Technical Journal, September 1994, pp. 73-80.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] H. M. Meral et al., Natural language watermarking via morphosyntactic alterations, Computer Speech and Language, 23, 107-125, 2009.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] Peng Lu et al., An optimized natural language watermarking algorithm based on TMR, Proceedings of 9th International Conference for Young Computer Scientists, 2009.

[29] T. Amano, D. Misaki, " A feature calibration method for watermarking of document images", proceedings og ICDAR, 1999.

[30] A. Bhattachariya, H. Ancin, "Data Embedding in text for a copier system", Proceedings of the ICIP, Vol.2, 1999.

[31] Y. Kim, I.Oh , "Watermarking text document images using edge direction histograms", Pattern Recognition Letters, 2004

[32] Young Won Kim et al., " a text watermarking based on word classification and interword sapace statistics", ICDAR, 2003.

[33] A. M. Alattar, O.M.Alattar, " watermarking electronic text documents containing justified paragraphs and irregular line spacing", proceedings of SPIE,2004.

[34] A.K.Mikkilineni et al. , " Signature embedding in printed documents for security and forensic applications", proceedings of the SPIE, 2004.

[35] Jiezhao Peng1, Qi Wu, "Research and Implementation of RSA Algorithm in Java", IEEE, 2008.