Improved Text Watermarking Technique

Prabhjot Kaur Cheema[1], Kamaljit Kaur[2]
Student, Masters of Technology[1], Assistant Professor[2]
[1.2]Department of Computer Science and Engineering
Sri Guru Granth Sahib World University
Fatehgarh Sahib, Punjab, India.

Abstract: Text Watermarking is the technique which is used for copyright protection of text documents these days. Many techniques have been used in context of security till now such as steganography, cryptography, hash function and access control methods. Watermarking is found to be the easiest and efficient technique for many languages like English, Chinese, Turkish and many more. In the proposed technique the idea presented by researchers for English language using natural language watermarks is enhanced by making the technique more robust against tampering attacks. The aim of the proposed technique is to create unique watermarks by senders so that attacker is unable to interpret it and the information can be transferred safely. So, watermark key can be created using combination of count of nouns, pronouns, model verbs and conjunctions along with author name. Moreover, double encryption is applied using RSA and AES algorithms to make the watermark unpredictable for attackers and secured.

Keywords: Watermarks; Encryption; RSA; AES; XOR; Key; Tampering Attack

I. INTRODUCTION

With the rapid advancement in science, technologies are increasing with the passage of time. Security is the area of major concern as data is transmitted from one place to another. Data may consist of and image, audio, video or text files. So, researchers are working on watermarking techniques to preclude the important information from unauthorized users. The watermark is a common term which is ordinarily used to keep the data protected in context of terms of the copyright issue. The blueprint of watermark was found in the late 90's when the matter of copyright came into light. Taking an example, suppose there is a person M who has an official document and he put it on the internet. There is a person N with bad connotation steal the document and changed it little bit and afterwards started selling it, as it was his own. M came to know that N is selling his document. So, in this matter M has to give evidence that he is the real possessor of the document. For this reason, we need text watermarking techniques. Watermarking can be done in two ways. It can be visible watermarking which means watermark is visible to everyone or it can be invisible watermarking in which bits hidden inside the text documents. In the proposed technique invisible text watermarking technique is used. [6]

II. RELATED WORK

Researchers have done lot of work in the field watermarking but it has been revealed that less work is done in context to text documents. However, digital library, e-books and also commercial management schemes are attaining popularity these days. So, watermarking the text document will be an crucial factor in these diligences for the purpose of copyright protection.

Ingemar et al. gave overview of applications and properties of watermarking. Watermarking depends on type of application so firstly the context on which watermarking is to be applied should be known. So, this paper concluded that

the evaluation of a watermarking algorithm is possible only if the context is indicated first. [1]

Mercan et al. presented a paper in which the use of natural language watermarking which uses the structure of sentence constituents to insert a watermark in text is discussed. It is different from other text watermarking techniques as it does not modify the appearance of text elements such as lines, words or characters. So, basically this paper concluded the difference of image based and text based natural language watermarking techniques used for watermarking. [2]

Patil and Patil proposed a text watermarking algorithm using structural approach for protecting the text from unauthorized use and provide copyright protection. A text watermarking method based on structural approach using occurrences of double letters in text faces with the problem that it is not applicable for all types of text documents. So, in the proposed algorithm text and image are taken as input and generate watermark key. But this method is applied by taking both image and text because structural text watermarking is not applicable for all types of text documents because it uses alphabetic watermark. [3]

Zunera et al. presented a zero text watermarking scheme in the international conference of 2010. According to this scheme, existing text watermarking algorithms are not robust against random insertion and deletion attacks on text. With the increasing volume of attack, the survival of watermark in text becomes challenging and hence development of novel text watermarking algorithm that can be used for copyright protection of textual contents is done in this paper. The comparison of results with other existing algorithms of the same contrast is done. The results are found to be effective enough to get proceeded for modification.[4]

Makarand et al. proposed a watermarking scheme on the basis of Natural Language. It was a fantastic idea to introduce Natural Language into the contrast of encryption. Till now, the text watermarking is implemented for English, Chinese, Turkish and Arabic language text using different

methods. This paper includes a novel watermarking techniques for English language text documents. Focus is on grammatical rules like conjunction, pronouns and modal verbs to generate encrypted watermark message. It uses AES (Advance Encryption Algorithm) for encryption and decryption. Hence his method is quite effective and can be considered for future development process. [5]

The proposed approach presented in our paper enhances the idea given by Makarand et al. and enhances the security level by applying double encryption on the algorithm which in turn gives better results and more robustness against tampering attacks.

III. PROPOSED APPROACH

In the proposed approach steps followed in generating a watermark are as follows:

A. Select a text document:

First of all document is selected on which watermarking is to be done. In this approach we have used text documents of extension .txt.

B. Steps in Generating Watermark:

Watermark is generated using the English language components. We have used noun, pronoun, model verb and conjunction.

a. Step 1: Text document is taken as an input
b. Step 2: Count of total number of noun, pronoun, conjunction and model verbs present in the document is calculated.
c. Step 3: Author name is entered and thus author name is concatenated with count calculated above
d. Step 4: Key is made up of combination of count of all possible combinations of noun, pronoun, model verb and conjunction concatenated with author name
e. Step 5: RSA algorithm is applied to encrypt the key
f. Step 6: XOR is applied on the output given by RSA
g. Step 7: AES is applied on the output given by previous step which is considered as the watermark.
h. Step 8: Watermark is embedded in the text document.
i. Step 9: Tampering attack is applied on the original and the watermarked document to check the robustness of proposed algorithm.
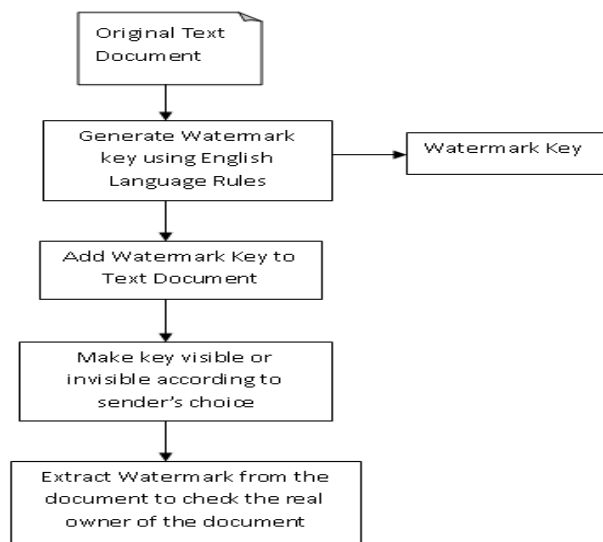

Figure 1. Flowchart of Proposed Technique

IV. RESULTS AND DISCUSSIONS

The watermark key is generated for the text documents using the key as given below:

Watermark Key= Concatenation of (Author name + Count of any combination (noun, pronoun, model verb or conjunction)

Figure 2 shows the original text document, and Figure 3 shows the part of document where watermarked bits are present

Figure 2. Original text document

Figure 3. Watermarked part of document

C. Robustness:

To check the robustness of the proposed algorithm we have compared the original document with the watermarked document on the basis of tampering attack. Figure 4 shows the effect of tampering attack on original document. Figure 5 shows the effect of tampering attack on watermarked document.
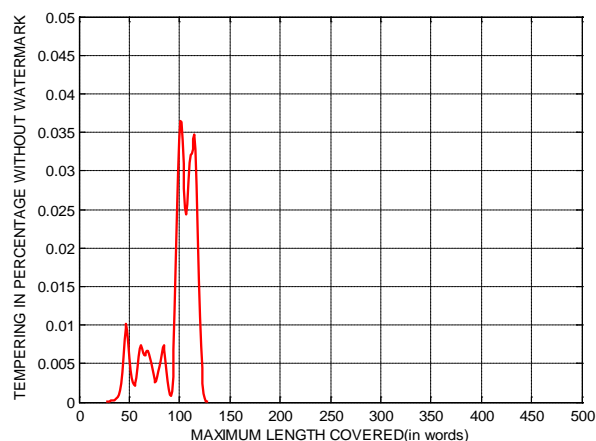

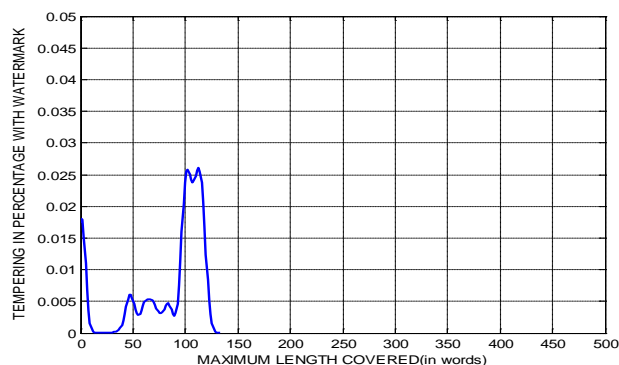Figure 4. Effect of Tampering Attack on Original document

Figure 5. Effect of Tampering Attack on Watermarked document

If we compare Figure 4 and Figure 5 it has been observed that effect of tampering attack is 3.7 % in the former and 2.6 % in the latter. Hence, watermarked document is more robust against tampering attack. Figure 6 shows the accuracy plot of proposed approach and it shows that proposed method is 79% robust against prevention of file from attacks.
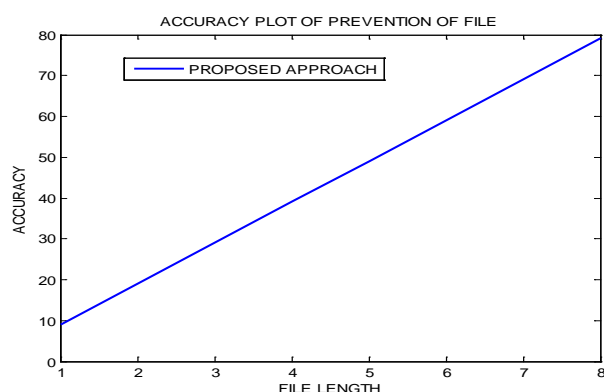

Figure 6. Accuracy plot of proposed approach

V. CONCLUSION

We have proposed a new approach in which a user can create watermark according to the choice using combinations of noun, pronoun, conjunction and model verb. Also encryption methods are used to enhance the security level which gives better results.

VI. REFERENCES

I.J. Cox, M. L. Miller and J. A. Bloom,"Watermarking applications and their properties" in Proceedings of IEEE International Conference on Information Technology: Coding and Computing (ITCC 2000), March 2000, pp. 6-10.

M. Topkara, C. M. Taskiran, E. Delp, "Natural language watermarking", in Proceedings of SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents VII , San Jose, CA, 2005.

Patil B. Devidas and Patil N. Namdeo, " Text Watermarking Using Structural Approach", in Proceedings of Institute of Electrical and Electronics Engineers, 2012.

Z. Jalil, H. Aziz, Saad B. Shahid, M. Arif and Anwar M. Mirza, "A Zero Text Watermarking Algorithm based on Non-Vowel ASCII Characters", in Proceedings of International Conference on Educational and Informational Technology(ICEIT), 2010.

Makarand L. Mali, Nitin N. Patil, J. B. Patil, "Implementation of Text Watermarking Technique Using Natural Language Watermarks", in Proceedings of International Conference on Communication Systems and Network Technologies, doi:10.1109/CSNT, 2013.

A. K. Vanwasi, "Digital Watermarking - Steering the future of security" Edition 2001, available at

http://www.networkmagazineindia.com/200108/security1.htm