ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014

Comparison of Text Watermarking Approaches with the Proposed Approach Based on Encryption Techniques used for Creating Watermarks

Prabhjot Kaur Cheema[1], Kamaljit Kaur[2]

Student, Mtech(CSE), Sri Guru Granth Sahib World University, Sirhind, India [1]

Assistant Professor, Computer Science Department, Sri Guru Granth Sahib World University, Sirhind, India [2]

Abstract: Due to advancement in science new technologies are emerging at a fast pace, so there is need of security while transferring data files or documents. Many techniques like steganography, hashing, cryptography and access control have been used so far. Watermarking is one of efficient technique used these days. Watermarking can be used in the areas of images, text, audio and videos. Text Watermarking is basically used for copyright protection of text documents. There are different methods used to create watermarks for the text documents. Syntactic, semantic, structural and natural language watermarking approaches are used for this purpose. Most of the languages like English, Turkish, Chinese and Arabic language text are using these techniques. Natural language watermarking is implemented to make watermarks using simple English language rules. It is a simple method and easy to implement. In the proposed technique components of English language like noun, pronoun, model verbs and conjunctions of user's choice along with author id are used to create watermark of user's choice. Moreover, encryption techniques RSA and AES and hybrid algorithm of both are applied to encrypt watermark and to enhance its security level to protect it from tampering attacks and to prove the most robust algorithm.

Keywords: Watermarks, RSA, AES, XOR, Key, Tampering, Robustness.

I. INTRODUCTION

Protection of the documents in the world of internet is very important these days. Watermarking refers to the terminology where we make the particular document copyrighted to prevent the misuse of the document. The document may be anything like Image, Text, Audio or Video. [1] 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. So, researchers are working on watermarking techniques to preclude the important information from unauthorized users. 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.

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 a crucial factor in these diligences for the purpose of copyright protection. [2] Text watermarking is done using the following steps as shown in Figure 1.

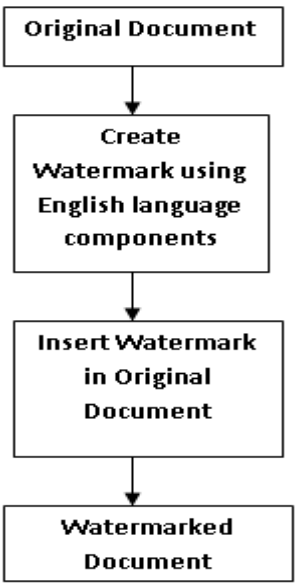


Figure 1. Watermarking process

Text watermarking can be used in the real world in wide number of applications. [4] With the increasing use of Internet all over the world for information sharing the need for text security is crucial. The emerging concepts of digital

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014

libraries, e-business, e-learning, and e-government, e-books, has made text watermarking a necessity. Legal documents, web sites, certificates, business plans, books, articles, company documents, SMS, emails and confidential contents can be protected by text marking algorithms. Text watermarking can be used for a number of purposes. Authentication, copyright protection, copy prevention, covert communication, tamper detection, and fingerprinting are some of the applications of text watermarking. [3]

II. PROPOSED WORK

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. [5] We have used noun, pronoun, model verb and conjunction.

Step 1: Text document is taken as an input
Step 2: Count of total number of noun, pronoun, conjunction and model verbs present in the document is calculated.
Step 3: Author name is entered and thus author name is concatenated with count calculated above
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
Step 5: RSA algorithm is applied to encrypt the key
Step 6: AES is applied on the output given by previous step
Step 7: XOR is applied on the output given by AES which is the actual watermark created by the proposed method
Step 8: Watermark is embedded in the text document
Step 9: Watermark is made invisible at the front end
Step 10: Tampering attack is applied on the original and the watermarked document to check the robustness of proposed algorithm. [2]

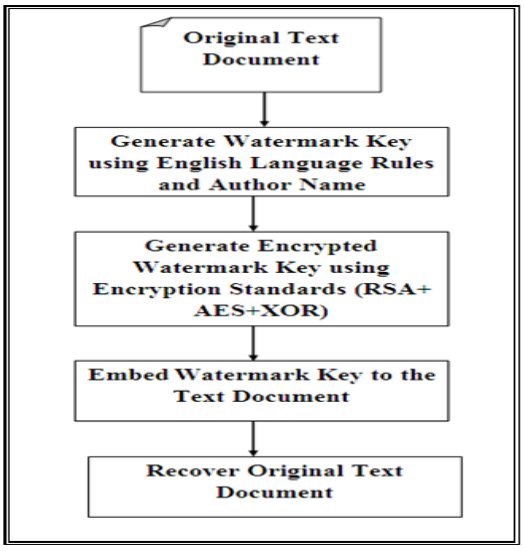


Figure 2. Flowchart of Proposed Technique

III. 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 3 shows the original text document of 4 KB, and Figure 4 shows key generated using AES

Well, John has a quite different, not necessarily more elaborated theology. There is some evidence that he must have known Luke, and that the content of Q was known to him, but not in a 'canonized' form. This is a new argument to me. Could you elaborate a little? The argument goes as follows: Q-oid quotes appear in John, but not in the almost codified way they were in Matthew or Luke. However, they are considered to be similar enough to point to knowledge of Q as such, and not an entirely different source. Assuming that he knew Luke would obviously put him after Luke, and would give evidence for the latter assumption I don't think this follows. If you take the most traditional attributions, then Luke might have known John, but John is an elder figure in either case. We're talking spans of time here which are well within the range of lifetimes. We are talking date of texts here, not the age of the authors. The usual explanation for the time order of Mark, Matthew and Luke does not consider their respective ages. It says Matthew has read the text of Mark, and Luke that of Matthew (and probably that of Mark). As it is assumed that John knew the content of Luke's text. The evidence for that is not overwhelming, admittedly.....

Figure 3. Original text document

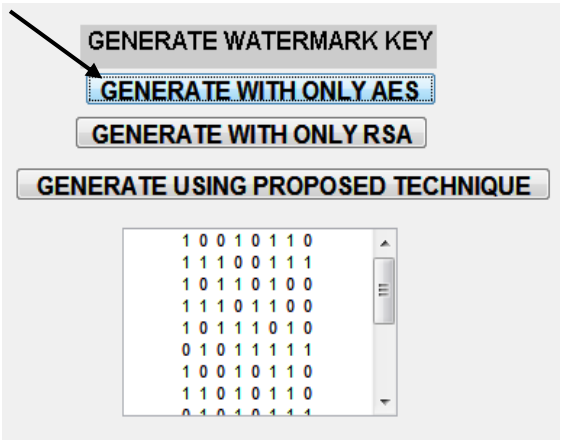


Figure 4. Key generated using AES

Figure 5 shows the visible key present in the text document at the backend and Figure 9 shows that key is made invisible at the front end.

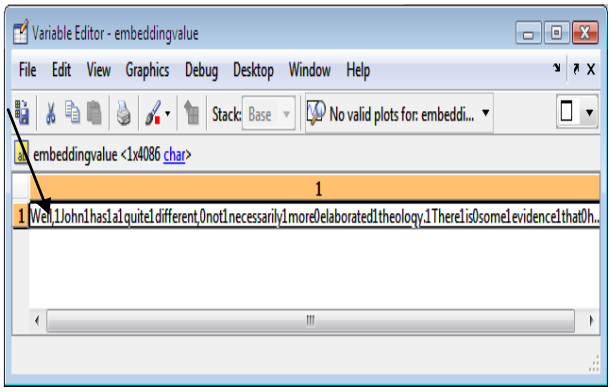


Figure 5. Key embedded at the backend generated using AES

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014

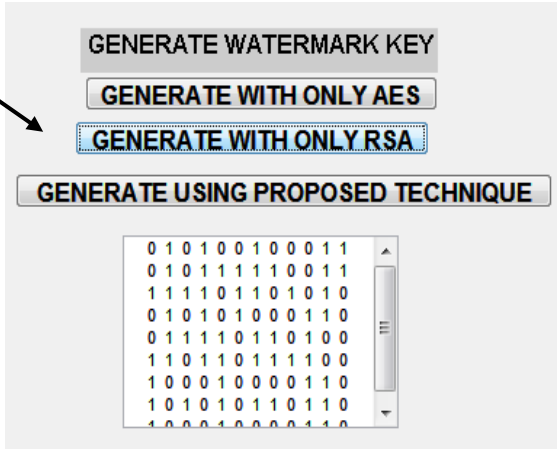


Figure 6. shows key generated using RSA

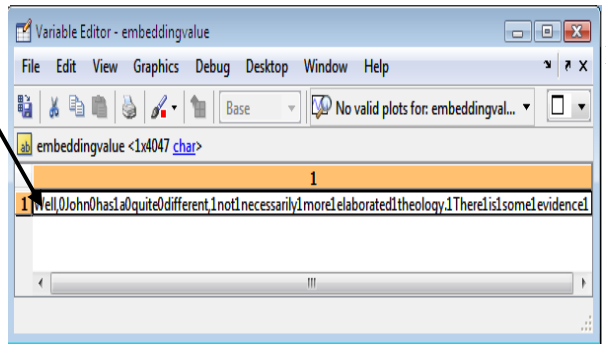


Figure 7. Key embedded at the backend generated using RSA

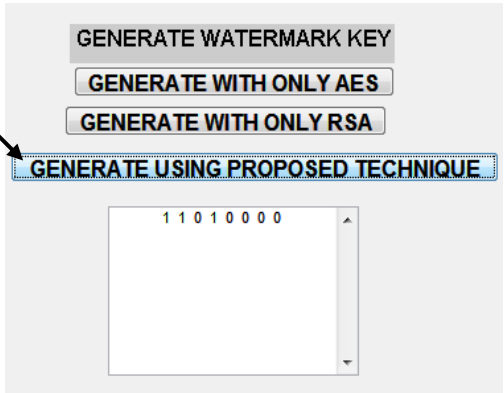


Figure 8. shows key generated using Proposed Approach

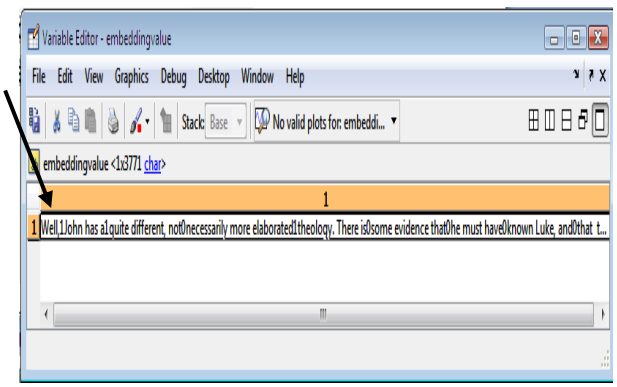


Figure 9. Key embedded at the backend generated using Proposed Approach

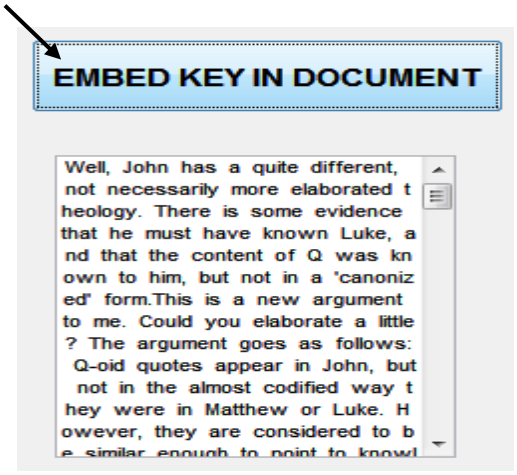
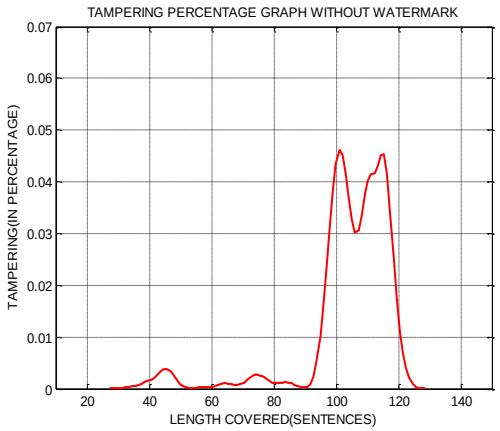


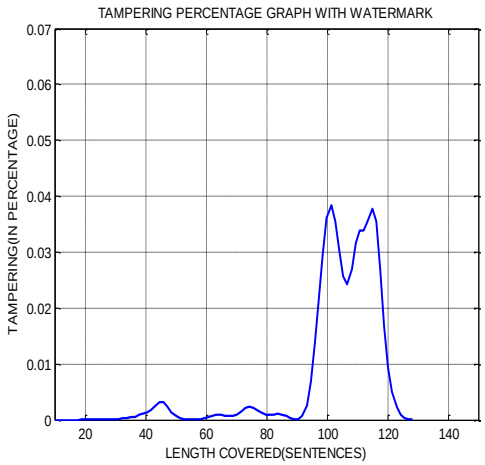
Figure 10. Invisible key at the front end generated using AES,RSA and Proposed Approach

IV. COMPARISON OF PROPOSED TECHNIQUE WITH RSA AND AES

The proposed algorithm is performed by taking files of different sizes in Kilo-bytes (KB). Graph 1 shows the percentage of tampering on original text document and watermarked text document of size 4 KB which is 4.5% and 3.8% respectively.



Graph 1 Percentage of Tampering in Original Document



Graph 2 Percentage of Tampering in Watermarked Document

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014

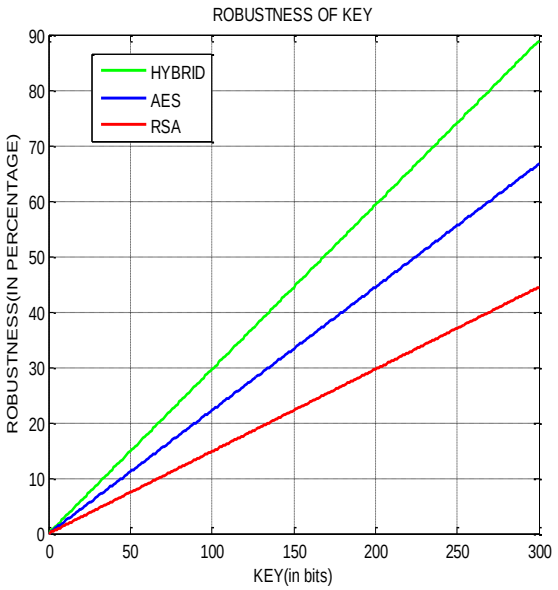
Table 1. shows the results to compare effect of tampering attack on original and watermarked document.

Table 1. Effect of tampering attack

Document Size(IN KB)	Tampering %(Original Document)	Tampering %(Watermarked Document)
1 KB	4.0%	3.0%
4 KB	4.5%	3.8%
6 KB	4.6%	3.6%
9 KB	3.7%	2.6%
13 KB	4.0%	2.8%
15 KB	4.2%	2.8%
17 KB	5.0%	4.1%
22 KB	5.1%	3.9%
25 KB	5.0%	4.4%
36 KB	5.1%	4.1%

Conclusion: The results of Table 1 show that effect of tampering attack is less on the watermarked document. However, it is more on original document.

Graph 3 shows the robustness of key made using Hybrid or Proposed Approach, AES and RSA which shows that watermark key of proposed approach is more robust.



Graph 3 Robustness graph comparing keys of AES, RSA and Hybrid Approach

Conclusion: It can be seen in the Graph 3 that robustness of key of proposed technique is 89%, AES is 67% and RSA is 44%. Thus, the proposed approach gives better result as compared to other approaches. Hence, it proves to be more robust than other approaches.

Table 2. Comparison of Robustness of Key of Proposed Algorithm with RSA and AES

Table 2. Robustness of Key

Size(IN KB)	Hybrid method	AES algorithm	RSA algorithm
1 KB	86%	70%	43%
4 KB	89%	67%	44%
6 KB	89%	67%	45%
9 KB	89%	66%	44%
13 KB	88%	67%	44%
15 KB	89%	68%	44%
17 KB	86%	71%	43%
22 KB	86%	71%	43%
25 KB	84%	73%	42%
36 KB	89%	67%	44%

Conclusion: Table 2 shows that watermark key of proposed algorithm is proved to be robust and gives better results than RSA and AES algorithm for documents of different sizes.

REFERENCES

[1] P. K. Cheema and K. Kaur, "Text Watermarking Techniques" in International Journal of Advanced Research in Computer Science and Software Engineering, Vol.4, no. 4, Issue 4., April, 2014.

[2] P. K. Cheema and K. Kaur, " Improved Text Watermarking Technique" in International Journal of Advanced Research in Computer Science, Vol. 5, No. 5, May-June, 2014.

[3] 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.

[4] 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.

[5] 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.