Using Combined Pseudo-Random Number Generator with Digital Text-based Watermarking for Cryptography Application

Chee Hon Lew, Chaw Seng Woo

***Abstract*—We focus on text based watermarking techniques based on Pseudo-Random Number Generator(PRNG) for Cryptography application. We survey related workin digital watermarking, cryptography and design methodology, then develop our own text based watermarking method (embedded and extract/detection of watermarks). Our implementation result have shown that better accuracy of extracted watermark and PRNG random bit sequence made its strengthen the security of protecting data. Our RSA Key generator therefore holds potential for future implementations of PRNG in practical parallel applications such as parallel grid computing, parallel genetic programming, parallel cryptography, and parallel computation analysis. This paper is intended to provide a reference finding for newcomer's security designer and to promote more activities in these security issues.**

***Index Terms*—Digital watermarking, Pseudo-Random Number Generator (PRNG), Text based Digital Watermarking Techniques, and Cryptography**

I. INTRODUCTION

Most of the research work that has been done in this field looks at increasing the output speed in computing the cryptographic results in on TRNG platform[1]. As we read other research related work in Research and Implementation of RSA Algorithm in Java[9]and Text Watermarking Using Combined Image-plus-Text Watermark[2,3]. These two papers discuss two separate independent studies on RSA algorithm in Java and text watermarking using combined image based watermarking. So far, the limited research studies of text-based digital watermarking techniques based on Pseudo-Random Number Generator (PRNG) for Cryptography application. Therefore, we need to do research and experimental undertaken within the context of design and implementation of text based watermarking combined with cryptographic techniques.

From the work that has been accomplished using a PRNG platform implementations, Pseudo random bit sequence integrated in text based watermarking. So nobody know it bit sequence in random. PRNG perform better in spend performance and measure in text/character in bit and compare with other methods such as semantic text watermarking[4].

Manuscript received November 20, 2012; revised December 20, 2012.
C.H.Lew is Ph.D Candidate with Department of Artificial Intelligence, University of Malaya, Malaysia.
E-mail: cheehon2006@siswa.um.edu.my
C.S.Woo is with the Department of of Artificial Intelligence, University of Malaya, Malaysia. E-mail:{cswoo@um.edu.my}

There are a few advantages of combining Watermarking with PRNG techniques:

- Good security for combined these two Watermarking with PRNG due to complex algorithms.
- It make hacker hard to predict it public key due to combination RSA algorithms and Watermarking algorithms.
- It can reusable and update source code whenever it is necessary.
- It can protected user and software developer for authentication and data.

This paper is organized as follows: Section II review literature on text-based digital watermarking techniques. Section III explain our research methodology and contributions. Section IV describes the proposed model of Text based Digital Watermarking Techniques embedded in PRNG algorithms for Cryptography applications. Section V shows our experimental work and results. Finally, we provide our conclusion in Section VI.

II. LITERATURE REVIEW

This section gives a literature review of some of the text-based Digital watermarking techniques such as using image based techniques(Line shift coding), Syntactic synonyms (natural language); semantics method(Noun verb based tree of sentence) - Sub Paragraph A-C,refers).

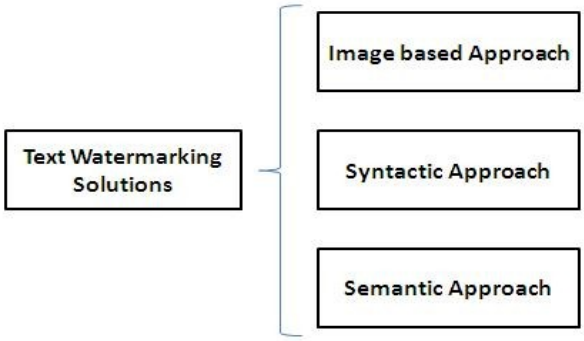

Fig. 1. Text watermarking solutions[5]

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

A. Image-based Techniques

In this approach towards digital text watermarking, the text document image is used to embed the watermark. Brassil, et al. proposed a few methods to watermark text document by using text image [6]. The first method proposed by Brassil[6] was the line-shift algorithm which moves a line upward or downward (left or right) depending on binary signal (watermark) to be inserted as shown in Fig. 2.

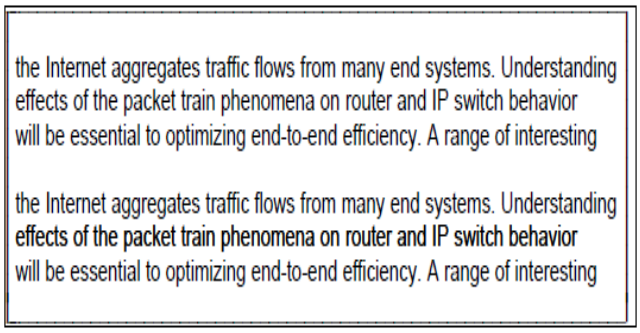

Fig. 2. Line shift coding [5]

B. Syntactic Techniques

Text is composed of sentences. Sentences are composed of words, and words can be nouns, verbs, articles, prepositions, adjectives, adverbs etc. Sentences have different syntactic structure which depends on language and its conventions. Applying syntactic transformations on text structure to embed watermark has also been one of the approaches towards text watermarking in the past.

Mikhail. J. Atallah, et al. first proposed the natural language watermarking scheme using the syntactic structure of text [7]-[8] where the syntactic tree is built and transformations are applied on this tree to embed the watermark. All the inherent properties of the text are preserved while embedding watermark. The watermarking process is shown in Fig. 3.

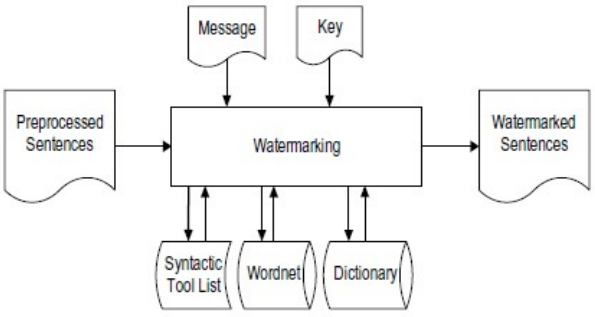

Fig. 3. Syntactic sentence level watermarking[5]

C. Semantic Method

The semantic watermarking schemes focus on using the semantic structure of text to embed the watermark. Text contents, like verbs, nouns, prepositions, words spelling, acronyms, sentence structure, grammar rules, etc. are exploited to insert watermark in the text.

Atallah et al. were the first to propose the semantic watermarking schemes in 2000[9]-[10]. Later, the synonym substitution method was proposed in which watermark is embedded by replacing certain words with their synonyms without changing the context of text[11]. Xingming, et al. proposed noun-verb based technique for text watermarking[12] which exploits nouns and verbs in a sentence parsed with a grammar parser using semantic networks. Fig. 4 shows the parse tree for noun-verb based transformation.

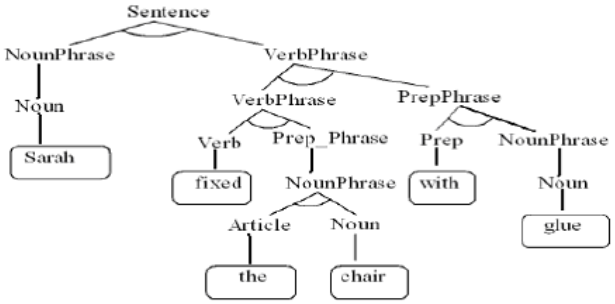

Fig. 4. Noun verb based tree of sentence "Sarah fixed the chair with glue"[5]

III. RESEARCH METHODOLOGY

The purpose of this research is to develop a text based digital watermarking that can be synthesized by using PRNG as standard Cryptography design tools. A text based digital watermarking that can be synthesized using PRNG standard Cryptography design tools will enable security designers to address digital content protection privacy concerns more efficiently. Developing a digital PRNG composed of standard Cryptography design tools is important because:

- It alleviates the need for embedding a text based digital watermarking design.
- The PRNG can be incorporated with other digital cryptographic components.
- No external components are required for a text based digital watermarking implementations.

The contribution of this research paper is two sides:
My contribution work does in Cryptography area as below:

1. A better understanding of Pseudo Random number primitives will make it easier to design and use PRNGs securely.
2. Modification and improvement of PRNG based on Open source RSA Algorithms.

My contribution work include in Watermarking area as below:

1. Better understanding concepts and methods of text based watermarking combined with PRNG for Cryptography application.
2. Modification and improvement of text based Watermarking based on Open source Watermarking Algorithms.
3. It make hacker hard to predict it public key due to combination open source RSA algorithms and watermarking algorithms.

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

IV. PROPOSAL MODEL

Our implementation model is as shown in Fig. 5

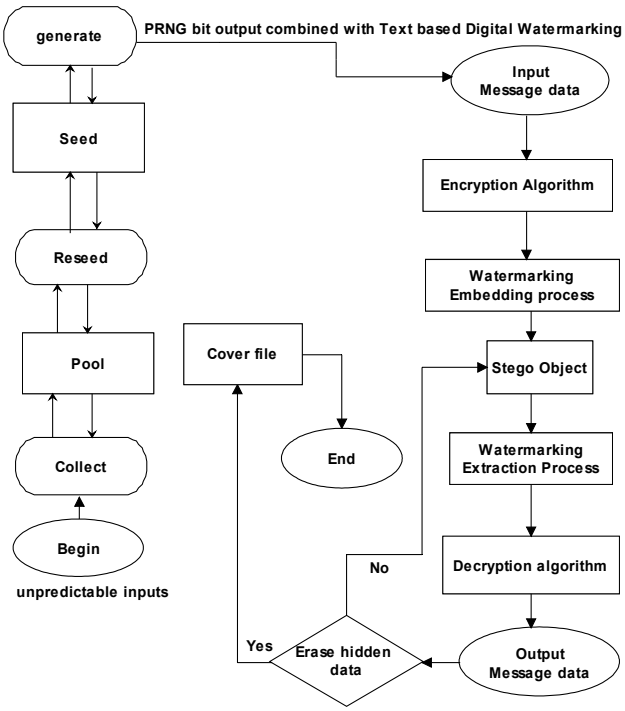

Fig.5. Proposed model of Text based Digital Watermarking Techniques embedded in PRNG algorithms for Cryptography applications.

Although our primary emphasis is on evolution of new text-based watermarking technique but we are also aiming towards providing a blended solution comprising of RSA-based encryption and text based digital watermarking dully added by compression before encryption.

The above left part is Generalized PRNG with periodic reseeding. PRNG with periodic reseeding that depicts a possible architecture for implementing catastrophic reseeding. The part of the internal state that is used to generate outputs should be separate from the entropy pool. The generation state should be changed only when enough entropy has been collected to resist iterative guessing attacks, according to a conservative estimate.

The above right part is proposed implementation model of Text based Digital Watermarking Techniques. Although our primary emphasis is on evolution of new Text-based Digital Watermarking Technique but we are also aiming towards providing a blended solution comprising of Encryption and Text based Digital Watermarking dully added by PRNG algorithm before encryption. To focus more on the subject and proposing a closest possible feasible secure digital content solution, The combination of hidden data-plus cover (known as stego object) that hold the hidden information upon watermarking extraction process. No message data can be erased upon done RSA decryption algorithms. We shall be using PRNG algorithms for embedding the best available compression with cryptographic hash function (SHA-2) and encryption algorithms/techniques(RSA encryption/decryption method).

V. RESULT AND PERFORMANCE ANALYSIS

The following sequence of steps identifies RSA Key Generator adopted in this work.

1. Definition of the problem.
2. An algorithm 1 **: Multiple-precision library (BigInt.js)** is a suite of routines for performing multiple-precision arithmetic in JavaScript.
3. An algorithm 2 : **Modular reduction library (Barrett.js)** is a class for performing Barrett modular reduction computations in JavaScript.
4. An algorithm 3 : **RSA library (RSA.js)** is suite of routines for performing RSA public-key computations in JavaScript. It link algorithm 1 and algorithm 2 into algorithm 3 experimental testing for plaintext, ciphertext, verification and message keys.
5. The RSA key generator application is written in Delphi 7, which is Object Pascal. This includes a re-implementation of the multiple-precision library and the Miller-Rabin test for primality. Keys are generated using Algorithm 8.1[17]-[18].

The following is procedures for compiling and testing RSA Key Generator.

1. Before testing “RSA Key generator testing file”, make sure all files (BigInt.js,Barrett.js, RSA.js) including in the same directory.
2. Open “RSA-testing” file by using Internet Explorer browser (such as Microsoft IE9, Google Chrome, Firefox).
3. You should testing this file as below Fig. 6:

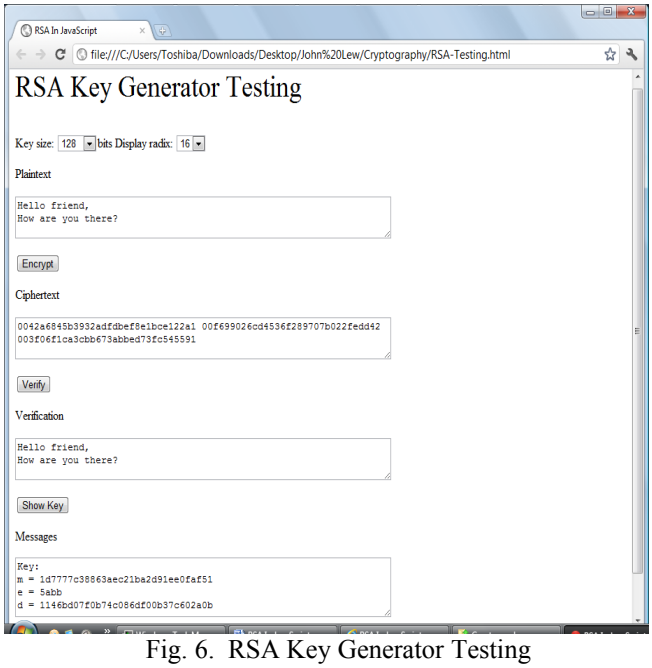

Fig. 6. RSA Key Generator Testing

4. Use Borland Delphi 7 or 8 edition for compiling RSA Key generator souce code. Run and compile all source code files(Main.pas, BigMath.pas, RSA.pas) before it can execute RSAkeygenerator(main program).
5. After run “RSAKeyGenerator” execution file, you should testing this file as below:

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

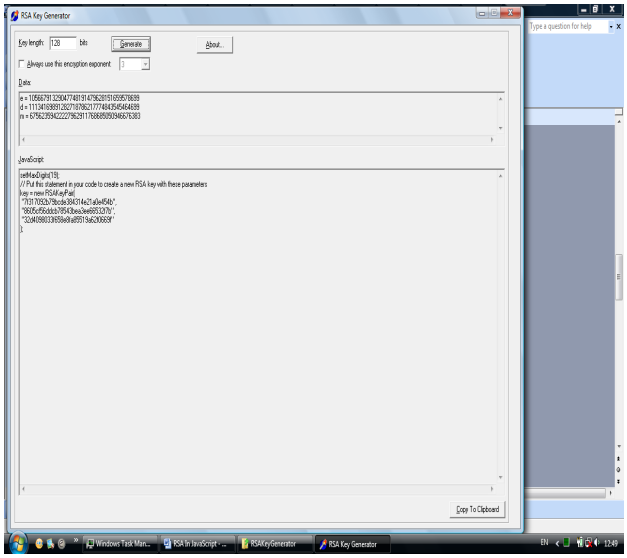

Fig. 7. RSA Key Generator Testing for generating new RSA key

6. In order to secure RSA Key Generator Testing(Fig. 6), editing this source code by using Notepad or HTML editor. Add the following JavaScript lines as below:

```
setMaxDigits;
// Put this statement in your code to create a new RSA
   key with these parameters
key = new RSAKeyPair(
 "3",
 "18e78c00696284b3923a6d7401aee5b",
 "255b52009e13c7108d7b57f0a764527"
);
```

7. Repeat testing "RSA Key generator testing file". Done it.

A. Further Modification and Improvement

In this section, we shown as below:

A.1) The modification and improvement of PRNG based on open source RSA algorithms as below:

i)On RSA key Generator Testing source code file, we edited a new RSA key pair statement. Due to RSA Key Generator Testing, it generating new RSA public key and private key. It always keep changing a new RSA public key and private key after done repeatedly RSA Key Generator Testing .

ii)Due to speed performance constraint, modified maximum to 2048-bits RSA key(Common Cipher strength 128-bit keys). So it can faster speed of encryption and accuracy performance.

iii)Due to RSA based PRNG encryption and decryption, PRNG random bit sequence made its strengthens the security of protecting data.

A.2) The Modification and improvement of text based Watermarking based on open source Watermarking algorithms as below:

i)Semantic Techniques is used for text based watermarking techniques. The semantic watermarking schemes focus on using the semantic structure of text to embed the watermark, so faster speed and accuracy embedd the text contents.

ii)We edited open source watermarking algorithms, we add up RSA based PRNG algorithms embedded with text based watermarking algorithms. Nobody know it combined PRNG random bit sequence embedded with text watermarking algorithms.

iii)After compiled RSA key Generator Testing, it generating Public key and Private key. Then put public key in text based watermarking pool in Fig. 8. Invisibility and erase hidden data, due to PRNG random bit sequence. it make hacker hard to predict or crack it public key.

This paper also developed an experimental tool with JAVA on windows according to the above embedding and extracting algorithm for watermarking, the interface is shown in Fig. 8 below.

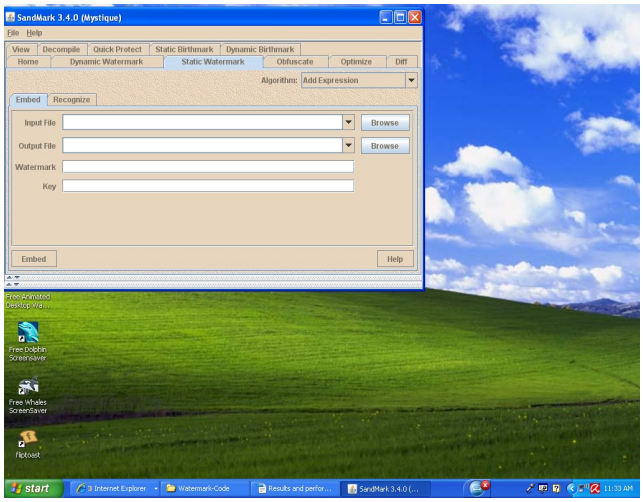

Fig. 8. Digital watermarking tool

We evaluate the performance of algorithm under both Text with/without PRNG that means random bit sequence for random insertion, deletion and reordering of word characters to and from the text. The average accuracy of extracted watermark under text with/without PRNG is shown in Table II and Fig. 9.

Table I: Text categories[4]

	Text Category	Number of Sentence
1	Small Size Text(SST)	< 20
2	Medium Size Text(MST)	[21,100]
3	Large Size Text(LST)	[101,1000]
4	Very Large Size Text(VLST)	> 1000

We used 10 text samples from data set designed and used in [15]. Each text samples are divided in four categories based on text length in Table 2[4], each text category contains 5 sample. These sample has been randomly selected from Reuter Corpus[15], e-book, newspaper, article, research paper, novels, web content, reports, etc.

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Table II: Accuracy of Extracted watermark under Text without/with PRNG.

Text category	Accuracy of extracted watermark	
	Image-plus-Text without PRNG[3]	Text with PRNG
1. Small Size Text(SST)	85.31%	97.20%
2. Medium Size Text(MST)	81.88%	97.46%
3. Large Size Text(LST)	90.22%	97.32%
4. Very Large Size Text(VLST)	92.60%	97.23%

A major improvement is that the semantic techniques is used for text based watermarking techniques. The semantic watermarking schemes focus on using the semantic structure of text to embed the watermark, so faster speed and accuracy embedd the text contents, like verbs, nouns, prepositions, words spelling, acronyms, sentence structure, grammar rules. Due to RSA based PRNG, textual watermarking is more secure sensitive to tampering attacks than text with PRNG. On section V sub-section A, we mentioned further modification and improvement of PRNG based on open source RSA algorithms and text based Watermarking based on open source Watermarking algorithms that affected the improvement in the Table II.

The most important difference is that Jalil and Mirza[10] measured the performance using Image-plus-Text based digital watermarking without PRNG. Rather than we measured the performance using text-based digital watermarking with PRNG.

The results given in Table II and show that better accuracy of extracted watermark due to text based watermarking combined with PRNG. The accuracy of extracted watermark under text with PRNG exceed 97%. Textual watermarking is more secure sensitive to tampering attacks than text with PRNG. Hence the accuracy of text without PRNG is lesser than text with PRNG. Experiments were also performed under text with/without PRNG on all text samples and percentages accuracy of extracted watermark for text with/without PRNG is shown in Table II and Fig. 9.

We have adopted a novel approach in text watermarking where text with/without PRNG are combined to form text based digital watermarking. There is no such previous work on combined text with PRNG in this research area, so no comparison have been found yet. Thus, there is no benchmark text available to facilitate comparison.

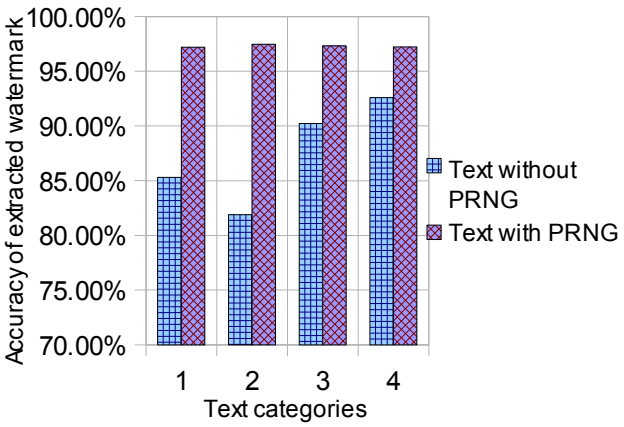

Fig.9. Accuracy of extracted watermark under Text without/with PRNG

Table III: Comparison between our text based watermarking method and previous approaches in term of capacity, robustness and security.

Approach	Speed	Robustness	Security
Our proposed in this work (Semantic Techniques)	High speed	Robust due to PRNG random bits	Invisibility and erase hidden data
Image-based techniques	Low speed due to image overhead	Not robust	Slightly visible
Syntactic Techniques	Moderate speed	Not robust	Slightly visible

In Table III, our proposed work also considered general comparison(capacity, robustness and security) with differences text-based watermarking techniques previously explained in section II, subsection A-C. Based on our observation, semantics techniques are more suitable techniques than other two approaches techniques due to high speed performance, robust, and invisibility for hidden data.

VI. CONCLUSION

In conclusion, this paper proposed a text based watermarking algorithm based on Pseudo-Random Number Generator (PRNG) for cryptography application. The paper has also reported the experimental results shows that proposed method has good invisibility and robustness to resist deletion, modification attack etc. Moreover this algorithm can also be applied in the information hiding area through cloud computing, as the embedding method in this paper has large information embedding capacity, which can be used for secret communication of confidential information. Finally, our design and implementation of Text based Digital Watermarking combined with Pseudo-Random Number Generator (PRNG) can be recommended for the following future direction on research finding:

- Our future research can extend further studies for advance mathematics in chaotic functions theory. [13].

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

- Further improvement watermarking method based on double random phase encoding technique[16]
- Our future research can extend further studies for double hash function for text based digital watermarking[14].

REFERENCES

[1] D.Eastlake, S.Crocker and J.Schiller, RFC 1750: Randomness Recommendation for Security, Internet Activies Board,1994

[2] Jiezhao Peng, Qi Wu, Research and Implementation of RSA Algorithmin Java, International Conference on Management of e-Commerce and e-Government, 2009

[3] Zunera Jalil and Anwar M. Mirza, Text Watermarking Using Combined Image-plus-Text Watermark, Second International Workshop on Education Technology and Computer Science in 2010.

[4] Zunera Jalil, M. Arfan Jaffar and Anwar M. Mirza, A Novel Text Watermarking Algorithm Using Image Watermark, International Journal of Innovative Computing, Information and Control, Volume 7, Number 3, March 2011.

[5] Zunera Jalil and Anwar M. Mirza, "A Review of Digital Watermarking Techniques for Text Documents", IEEE International Conference on Information and Multimedia Technology, pp. 230-234, 2009

[6] 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, July 1999, pp.1181-1196.

[7] 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, September, 2000, Cork, Ireland, pp. 51–65.

[8] 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 HidingWorkshop, vol. LNCS 2137, 25-27 April 2001, Pittsburgh, PA.

[9] 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, October, 2002, Noordwijkerhout, The Netherlands,Springer-Verlag. *3612*: 958-961, Springer Press, August 2005.

[10] Topkara, C. M. Taskiran, and E. Delp, "Natural language watermarking," Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents VII, 2005.

[11] 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.

[12] Xingming Sun, Alex Jessey Asiimwe. Noun-Verb Based Technique of Text Watermarking Using Recursive Decent Semantic Net Parsers. *Lecture Notes in Computer Science (LNCS)*

[13] S. Behnia,A. Akhavanb, A. Akhshani, A. Samsudin, A novel dynamic model of pseudo random number generator, Journal of Computational and Applied Mathematics 235 (2011) 3455– 3463.

[14] Jiri Fridrich, Miroslav Goljan, Robust Hash Functions for Digital Watermarking, Proceeding of International conference on Information Technology: Coding and Computing, 2000.

[15] Reuter Corpus, [Online] Available: http://about.reuters.com/researchandstandards/corpus/index.asp

[16] S. Behnia, A. Akhavan, A. Akhshani, A. Samsudin, An improved watermarking method based on double random phase encoding technique , Journal of Computational and Applied Mathematics 235 (2011) 3455–3463

[17] Alfred Menezes, Paul van Oorschot, Scott Vanstone,Handbook of Applied Cryptography (Discrete Mathematics and Its Applications) , CRC Press; 1996

[18] Bruce Schneier, Applied Cryptography: Protocol, Algorithms, and source code in C, John Wiley & Sons, Inc. 1996.

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)