# PSEUDO-RANDOM NUMBER GENERATOR (PRNG) COMBINED WITH TEXT BASED WATERMARKING FOR CRYPTOGRAPHY APPLICATION

LEW CHEE HON

FACULTY OF COMPUTER SCIENCE AND
INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

# PSEUDO-RANDOM NUMBER GENERATOR (PRNG) COMBINED WITH TEXT BASED WATERMARKING FOR CRYPTOGRAPHY APPLICATION

LEW CHEE HON

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTER SCIENCE AND
INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

## UNIVERSITY OF MALAYA

## ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **LEW CHEE HON**

Registration / Matric No: **WHA060003**

Name of Degree: **Doctor of Philosophy (Ph.D)**

Title of project paper/Research report/ Dissertation/Thesis ("this work")**:**
**"Pseudo-Random Number Generator(PRNG) Combined with Text based Watermarking for Cryptography Application"**

Field of study: Digital watermarking

**I do solemnly and sincerely declare that:**

1. I am the sole author/writer of this work.
2. This work is original;
3. Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the work and its authorship have been acknowledged in this work;
4. I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work.
5. I hereby assign all and every right in the copyright to this work to University of Malaya ("UM"), who henceforth shall be owner of the copyright in this work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained.
6. I am fully aware that if in the course of making this work I have infringed any other action as may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature: Date:

Subscribed and solemnly declared before

Witness's Signature Date:

Name: Supervisor (Supervisor)

# ABSTRACT

Text digital watermarking,which is an important research direction for information hiding, can be used for copyright protection of text documents. Digital text protection received increasing attention in the last decade due to massive digital data documents via internet, hence digital watermarking was considered as a potential solution for solving cryptography problems since many E-commerce applications protect their data transactions through the internet. These transactions include sensitive document transfers, digital signature authentications, digital watermarking for data security protection, and digital data storage and linkages. This thesis proposed a digital text based watermarking algorithm combined with Pseudo-Random Number Generator(PRNG) for cryptography application. Related work was surveyed in digital watermarking, cryptography and design methodology before implementing the text based watermarking method (embedded and extract/detection of watermarks).The implementation results on text based digital watermarking combined with cryptographic techniques are presented. This thesis intends to provide a reference finding for newcomers in security designs and to promote more activities in these security issues. This experiment demonstrates that better accuracy of extracted watermark due to text based watermarking is obtained when combined with PRNG. The accuracy of the extracted watermark under text with PRNG exceeded 97%. The sentence selected for embedding watermarking bit are 85% higher in speed due to faster embedding watermarking bit string.The relation of information-hiding capacity is 1:5.5 which means that for every 5.5 terms of text, one bit of watermark can be hidden data. A comparison was made between a PRNG generator with the proposed RSA key generator and of text based watermarking in the proposed work. Therefore, security is improved as it is difficult to detect watermark (imperceptibility) when combined with the developed PRNG.

# ABSTRAK

Teks Watermarking digital, yang merupakan arah penyelidikan penting maklumat bersembunyi, boleh digunakan untuk perlindungan hak cipta dokumen teks. Perlindungan teks digital yang diterima semakin meningkat perhatian dalam dekad yang lalu disebabkan oleh besar-besaran dokumen data digital melalui internet, jadi digital Watermarking satu penyelesaian yang berpotensi untuk menyelesaikan masalah kriptografi. Sejak banyak E-dagang aplikasi melindungi transaksi data mereka melalui internet. Urus niaga ini termasuk sensitif pemindahan dokumen, pengesahan tandatangan digital, Watermarking digital untuk perlindungan keselamatan data dan penyimpanan data digital dan hubungan. Tesis ini dicadangkan digital algoritma Watermarking berasaskan teks digabungkan dengan nombor pseudo-rawak Generator (PRNG) bagi permohonan kriptografi. Kami meninjau kerja-kerja yang berkaitan di Watermarking digital, kriptografi dan metodologi reka bentuk, Watermarking berasaskan teks kemudian dilaksanakan kaedah (terbenam dan ekstrak / pengesanan tera air) .Kami menunjukkan hasil pelaksanaan kami pada berdasarkan Watermarking digital teks digabungkan dengan teknik kriptografi. Tesis ini adalah bertujuan untuk memberikan dapatan rujukan untuk designer keselamatan pendatang baru dan untuk menggalakkan lebih banyak aktiviti dalam isu-isu keselamatan. Eksperimen ini menunjukkan bahawa ketepatan yang lebih baik daripada yang diekstrak watermark kerana Watermarking berasaskan teks digabungkan dengan PRNG. Ketepatan diekstrak watermark di bawah teks dengan PRNG melebihi 97%. Ayat yang dipilih untuk menerapkan sedikit Watermarking adalah kelajuan 85% lebih tinggi disebabkan oleh sedikit Watermarking lebih cepat embedding tali. Hubungan kapasiti maklumat bersembunyi ialah 1: 5.5 yang bermaksud untuk setiap 5.5 segi teks, satu bit watermark boleh terdiri daripada data tersembunyi. Kami membuat perbandingan antara penjana PRNG dengan RSA penjana utama dan perbandingan Watermarking berasaskan teks dan kami yang dicadangkan cadangan kerja Oleh itu, kami meningkatkan keselamatan dengan membuat ia lebih sukar untuk dikesan watermark(imperceptibility) digabungkan dengan PRNG.

# ACKNOWLEDGMENTS

I would like to express my gratitude to my exceptionally outstanding supervisor Dr. Woo Chaw Seng for his guidance, support and his continuous enthusiasm and interest during my course of study. I am also indebted to my preceding supervisor Dr Ng Liang Shing for his advice and suggestions.

I would also wish to express my profound gratitude to both my parents (father and mother); for their overwhelming support and encouragement throughout my PhD studies.

## TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# Chapter 1: Introduction

## 1. Introduction

Cryptographic applications rely on Pseudo-Random Number Generators (PRNG) to generate secret keys such as session key pairs, keys, and passwords. They also require random numbers that are public, such as salts. Random numbers are generated so that they are unpredictable to an attacker. PRNG produces a long bit sequence of random numbers collected from real random data to seed a PRNG. Cryptography can protect/decrypted the content, but the content has no further protection after decrypted. Therefore, digital watermarking able to embed information into its data which makes it hard to delete or modify. In this thesis, text based watermarking techniques are presented based on the Pseudo-Random Number Generator (PRNG) for Cryptography applications. In this thesis, a review of some of the keywords which are used in current literature are presented accordingly for readability and clear distinction. In addition to that, the application development of watermarking technology and random number generation concepts are also briefly discussed. Digital watermarking terminologies and random number generation used in current literature are also introduced in attempt for clear distinction of information.

**Digital Watermarking (data hiding)** is the process of embedding data into multimedia elements such as images, audio or video and texts [Cox, Miller, and Bloom, 2002]. Digital watermark is content a code about ownership or authenticity that is hidden inside its data. [Manpreet Kaur,Sonika Jindal,Sunny Behal, 2012]. Section two discusses in detail digital watermarking concepts, definitions and history, characteristics, applications and differences in information hiding techniques.

**Pseudo-Random Number Generator (PRNG)** produces pseudo-random numbers by using cryptographic algorithm that cannot be guessed by an attacker [Chaitin,G.J. (1975)]. Random numbers are typically used to generate session keys, and their quality is critical for the quality of the resulting systems. Perfect randomness is a scarce resource with the use of the current pseudo random generators for all cryptographic applications, and hence, it is fitting to use Pseudo Random Bits Generator (PRBG). Therefore a good random number generation will performed well in cryptographic security [B.A.Wichmanna, I.D. Hill.(2006)]. Pseudo-random numbers plays an important role in fields such as statistical simulation systems, secret communication systems, modern cryptography systems, Monte Carlo methods and so on.

**Cryptography** is the study of secret writing. [Kelsey, Schneier, Wagner, Hall. (1998)] It is the study of methods of sending messages in distinct forms so that, only the intended recipients can remove the disguise and read the intended messages. A cipher is a form of secret writing which means plaintext is transformed into Ciphertext [Kelsey, Schneier, Wagner, Hall. (1998)]. This process is called encryption of the plaintext into ciphertext.

## 1.1 Research Background and Motivation

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 True Random Number Generator (TRNG) platform [D.Eastlake, S.Crocker and J.Schiller. (1994)]. Progressive readings will also lead readers to works by [Jiezhao Peng, Qi Wu. (2009)] and [Zunera Jalil and Anwar M. Mirza. (2010). This papers discusses two separate independent studies on RSA algorithm in Java and text watermarking using combined image based watermarking. However, due to the lack of research studies on text-based digital watermarking techniques based on PRNG for Cryptography application, the need to conduct this experimental research on design and implementation of text based watermarking combined with cryptographic techniques was necessary.

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

- Good security due to complex algorithms when combined.
- It gives a hacker a difficult and hard time to predict its public key due to the combination of RSA algorithms and Watermarking algorithms.
- It can protect users and software developers for authentication and data.

This thesis is organized as follows; Chapter 1, 2, 3 describes the introduction, overview of digital watermarking and cryptography and literature review of related works. Chapter 4 and Chapter 5 further explains on the research methodology and design methods, proposal implementation model of Text Based Digital Watermarking Techniques embedded in PRNG algorithms for Cryptography applications. Chapter 6 shows the implementation and testing of the research, experimental results and performance analysis. And finally in Chapter 7, the conclusion and future research direction work is provided for review.

The **RNG (Random Number Generator)** provides a sequence of random numbers. **True (Physical) Random Number Generator (TRNG)** requires a source of entropy (randomness). [Charles Wright (2004)].

**Pseudo-Random Number Generator (PRNG)** produces random numbers by using cryptographic algorithm. Strong PRNG cannot be predicted by a hacker. [SSH Communications Security, 2004]. Therefore, PRNG uses an algorithm generate genuine random numbers that are used for stimulation of random processes and statistical methods. Perfect randomness use the current PRNGs for all cryptographic applications, and hence, it is fitting to use pseudo random bits generator (PRBG). [Parijat Naik, 2002].

## 1.2 Research Aim and Objectives

This research aims to provide an in-depth understanding of text based digital watermarking techniques based on PRNG. An embedded algorithm will be designed for data security using text based watermarking techniques combined with cryptographic techniques. Thus, the text based watermarking method (embedded and extract/detection of watermarks) can be implemented. Besides that, the following objectives points will be discussed in depth:

- A novel PRNG approach was adopted towards text based digital watermarking.
- RSA based on PRNG techniques generated public made keys, which when embedded with text based digital watermarking are used for text based watermarking results.
- Results are compared with the accuracy of extracted watermark under text with or without PRNG.
- Show output of performance analysis such as performance of the designed system, comparisons of performance for coverage and information hiding capacity, comparisons of performance for average edit, comparisons between text based watermarking method and previous approaches in terms of capacity, robustness and security.
- Comparisons of a PRNG generator with the proposed RSA key generator.

## 1.3 Scope of Research

In order to ensure that this research is conducted in accordance to the objectives, the scope of this research is confined to the followings:

Firstly, this research will be emphasizing on Pseudo-Random Number Generator (PRNG as Deterministic Generators) in aspect of random number generator. Therefore, other aspects of random number generator such as “Quasi” random number generator are not within the concerns of this research. Furthermore, other types of advance cryptography algorithms such as randomness in complexity theory and algorithms, elliptic curve, quantum cryptography are also not considered in this research.

Secondly, this research will be focusing on the implementation of the PRNG key generator based on the pseudo-random number generator concept for applying text based digital watermarking techniques. Hence, the developed PRNG key generator application is created by Borland Delphi and integrated with RSA algorithms that encrypt data with maximized 2048 bits. This scope will help the development process to focus on to the stated objectives.

Lastly, since the area of this applied research interest is rather novel, hence , this research will focus on text based watermarking techniques using the Pseudo-Random Number Generator (PRNG) for Cryptography. Therefore, justification and verification of its encryption will be revealed upon completion of testing output results.

## 1.4 Research Importance and Contributions

This research has important practical and theoretical implications for Cryptographic application. The implications are as follows:

- A PRNG can permit possible attacks by hacker if it uses a simple design structure. Therefore, it is necessary to create a one of its kind, primitive and unique cryptographic.

- To implement a text based watermark that takes place in making a tradeoff between robustness and imperceptibility.
- To propose a new text based watermarking method by using Pseudo-Random Number Generator (PRNG) for Cryptography.
- Practical demonstration of the text based watermarking is not only data hiding in the watermark, but can also be used for cryptography application
- Presents a research finding for text based watermarking techniques using Pseudo-Random Number Generator (PRNG) for Cryptography upon completing testing output results.

Original published contributions are listed in the *Research Publication* of the Preliminaries section. There are 7 publications in total with 3 journals paper and 4 conferences paper, respectively. All have been presented in refereed conferences and accepted in the corresponding proceedings or journals. The contents of these publications are further discussed in the subsequent chapter of this thesis.

To the existing body of knowledge, and researcher's awareness, this research paper makes the following two important research contributions:

My contribution work contributes effectively in the Cryptography area:

- A better understanding of Pseudo Random number primitives, which will make it easier to design and use PRNGs securely.
- The modification and improvement of PRNG based on Open source RSA algorithms.

My contribution work contributes effectively in the Watermarking area:

- Provides a better understanding of concepts and methods of text based watermarking combined with PRNG meant for Cryptography application.
- The modification and improvement of text based Watermarking based on Open source Watermarking Algorithms.
- It causes a hacker to face difficulty in predicting its public key due to the combination of open source RSA algorithms and watermarking algorithms.

### 1.5 The Research Methodology and Design Tools

The purpose of this research is to develop a text based digital watermarking that can be synthesized by using Pseudo-Random Number Generator(PRNG) as standard Cryptography design tool. A text based digital watermarking that can be synthesized using PRNG standard Cryptography design tool 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 of the following reasons:

- 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.

### Design Tools

The design tools are written in Delphi 8 . Apart from that, the user interface of RSA Key Generator Testing is created by using a HTML Editor with integrated JAVA Scripts. The RSA Key Generator is used for generating new a RSA key for the public and private key. Apart from that, an open source text based digital watermarking algorithms is used as an experimental tool with JAVA on embedding and extracting algorithm for watermarking. The user interface of digital watermarking tool is also based on an open source watermarking combined with PRNG.

### 1.6 Outline of Thesis

This thesis is divided into seven chapters:

Chapter 1: Introduction: Chapter 1 provides a brief overview of the research problems, discusses the overview of watermarking and random number generator, stresses on research aims and objectives, confers on research importance and contributions and explains the research methodology and design tools used.

Chapter 2: Overview of Digital Watermarking and Cryptography: This chapter illustrates a general background of the digital watermarking cconcept, definition and history, characteristics, applications and differences information hiding techniques and cryptography concept, the modern mathematical cryptography, the concept of Secret and Public Key Cryptography and some specific cryptography applications also thoroughly presented. Apart from that, detailed description of digital watermarking model, an overview of text based watermarking techniques, client-side watermarking embedding and secure watermark detection is also presented in this chapter.

Chapter 3: Literature Reviews on Relevant Works: Related work in PRNG and Text based watermarking, advantages of PRNG and watermarking combination, and the combination of PRNG and Watermarks for improved security, key researchers in the combinations of PRNG and Watermarking is reviewed and surveyed. A description of PRNG algorithms for Cryptography application of the application of cryptography, related to random number generator is presented. This chapter illustrates the reported use of cryptography applications (comprising pseudo-random number generator, lottery number generator, and zero knowledge proof) in cryptographic applications.

Chapter 4: Research Methodology and Design Methods: As mentioned in Section 1.6, there are three research phases identified for this research, and therefore, Chapter 4 presents the research activities involved in each of these phases. Several data collection procedures are discussed in depth here. Several constructs which form the structure of the framework are discussed detail in Section 4. The general Framework for Watermarking is also elaborately discussed. Apart from that, this chapter also introduces several concepts related to the proposed text based digital watermarking technique and details of the proposed embedding procedure using PRNG key generator.

Chapter 5: Implementation of RSA Based PRNG and Text-based Watermarking: Chapter 5 will discuss and present the proposed RSA based PRNG algorithm description including modification and improvement of RSA key generator. In addition to that, this chapter will dissect the proposed text based digital watermarking algorithm description including modification and improvement text based watermarking algorithm.

Chapter 6: Result and Performance Analysis: in this chapter, the researchers will be able to show the user interface of RSA Key Generator Testing, RSA Key Generator Testing for generating new RSA key, the user interface of Digital watermarking tool. The experimental results and findings of this research will be further scrutinised. The output of performance analysis such as performances of the proposed system, comparison of Performance for coverage and information hiding capacity, comparison of performance for average edit, along with the comparison between the proposed text based watermarking method and previous approaches in term of capacity, robustness and security will be presented accordingly.

Chapter 7: Conclusion and Future Work: This chapter concludes the research conducted through the written presentation of the final discussion on results is obtained, and reasons and recommends possible further research directions.

# CHAPTER 2: BACKGROUND OF WATERMARKING AND CRYPTOGRAPHY

## 2. Digital Watermarking Overview

Digital watermarking technology is plays an important role in cryptography, computer science, communications, and signal processing. Digital Watermarking provides solution to developers, scrambling for content protection once done with data encryption.

### 2.1.1 Watermarking History

Watermarks were used as anti-fake measures on money in the 18th century. The first example of digital watermarking was identified by Emil Hembrooke of a patent filed for identifying music works in 1954. The term "digital watermarking" was first used by Komatsu and Tominaga in 1988 . Copyright protection of content, increased watermarking interest.[Ravi Sharma, 2012].

### 2.1.2 Watermarking Application

- *Broadcast monitoring:* Identifying when and where works are broadcasted by recognizing watermarks embedded in them.
- *Owner identification:* Embedding the identity of a work's copyright holder as a watermark.
- *Proof of ownership:* Using watermarks to provide evidence in ownership disputes.
- *Transaction tracking:* Using watermarks to identify people who obtain content legally but illegally redistribute it.
- *Content authentication:* Embedding signature information in content that can be later checked to verify it has not been tampered with.

### 2.1.3 Watermarking Properties

The suitability of a given watermarking system for a given application may be judged in terms of the following properties of that system:

- ***Embedding effectiveness:*** The probability that the embedder will successfully embed a watermark in a randomly selected Work.
- ***Fidelity:*** The perceptual quality of watermarked content.
- ***Data payload:*** The amount of information that can be carried in a watermark.
- ***Blind or informed detection:*** Whether (a) the watermark detector can detect a watermark in a Work without extra information (*blind detection*), or (b) the detector requires some information related to the original version of a watermarked work.

### 2.1.4 Characteristics of Watermarks

Several requirements must be satisfied, in order to achieve a maximum protection of intellectual property with watermarked media.

- **Undeletable**: The watermark must be difficult or even impossible to be removed by a malicious cracker, at least without obviously degrading the host signal.
- **Statistically undetectable:** A pirate should not be able to detect the watermark through a comparison of several watermarked signals belonging to the same author.
- **Robustness:** Watermark should be recoverable, and commonly used for transmission and storage. The watermark should be retrievable even if common signal processing operations are applied, such as signal enhancement, geometric image operations and noise filtering. (Hao-Tian Wu and Yiu-Ming Cheung, 2005).
- **Unambiguous**: Retrieval of the watermark should unambiguously identify the owner,

and the accuracy of identification should degrade gradually in the face of attacks.

- **Imperceptibility:** A watermark should be integrated with digital data such that it can not be distinguished. The watermark should be perceptibly invisible or transparent (imperceptible).

### 2.2 Definition of Random Number Generator, Cryptography.

**Random Number Generators(RNG)** is used to generate an array of numbers that have a random distribution. In practice, a random number generator does not generate numbers in random. Therefore, a random number generator require an initial value, or seed. Initializing the generators with the same seed will give the same sequence. Some types of random number generators such as truly random number generator, pseudo-random generator and quasi random number generator. However, this research will be emphasizing on True Random Number Generator(TRNG as Non-Deterministic Generators) and Pseudo-Random Number Generator (PRNG as Deterministic Generators) . Therefore, other “Quasi” random number generator are not within the concerns of this research. These two types of random number generators will be discussed further in depth more detail in Section 5.

### Cryptography

Cryptography is the study of secret writing. As Figure 2.1 , a cipher is secret writing that means plaintext is transformed into Ciphertext. This process is called encryption of the plaintext into ciphertext. The reverse process is called decryption.[Samuel S.Wagstaff Jr, 2003]

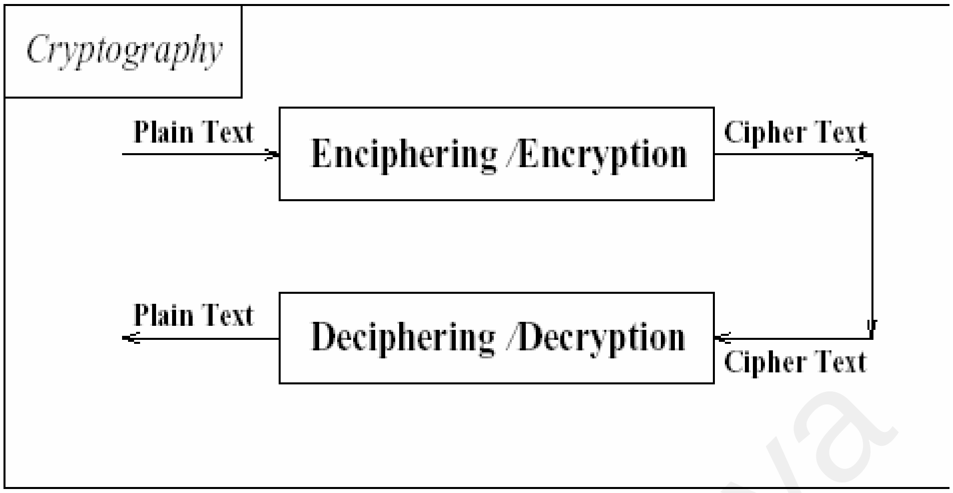


**Figure 2.1 Basic Principle of Cryptography[S. P. Mohanty (1999)]**

Ciphers are divided into two categories: Substitution and Transportation Ciphers.

**Substitution Cipher** replaces letters or large blocks of text with substitutes of usually the same length. In a simple substitution cipher, the same alphabet is used for plaintext and a ciphertext provides the substitution rule. For example, suppose the letters of the alphabet are arranged in a circle (with A following Z) and the message is encrypted by replacing each plaintext letter by k = 5 . The cipher is often called **Cesar Cipher** because Julius Caesar used it. Thus the message word “SECRET” would be enciphered as “XJHWJY”.

A **transportation cipher** rearranges the characters in the plaintext to form the ciphertext. The letters remain unchanged. For example, the plaintext of “HELLO WORLD” would be re-written as: [Samuel S.Wagstaff Jr, 2003] HLOOL ELWRD, in ciphertext.

(Resulting in the ciphertext “HLOOLELWRD”.)

The Product cipher is created by the composition of several ciphers whose types alternate between substitution and transportation. Substitution and transportation ciphers each have certain weakness which may be resolved through this alternating process. For an example, for the composition of the two ciphers above, first through the use of the Caesar cipher -the plaintext “SECRET” is first changed into “XJHWJY” and this is written into the Matrix. [Samuel S.Wagstaff Jr, 2003] while the ciphertext is read as “HYJJXW”.

X J H
W J Y
And the ciphertext is “HYJJXW”.

The Data Encryption Standard (DES) and Advanced Encryption Standard(AES) are product ciphers. The key for a transposition cipher is the fixed permutation of the letters in a block. The key for simple substitution cipher is the fixed permutation of the alphabet.[Samuel S.Wagstaff Jr, 2003]

### 2.3 The Objective of Cryptography

Cryptography is used to provide security as explained below:

- **Confidentiality**. Only authorized people can see the protected data.
- **Data integrity**. Assets can only modified by authorized parties or only in authorized ways. In this context, modification include writing, changing, changing status, deleting and creating
- **Authentication**. The recipient of a message should be able to verify its origin. Both the recipient and the sender should able to authenticate each other.
- **Non-repudiation**. The sender should not be able to later deny sent message. This is important in electronic commerce applications, where it is important that a consumer is not able to deny authorization of a purchase.

### 2.4 Overview of Cryptography Concept

"Cryptanalysis" is the art of breaking, attacking or analyzing cryptographic methods. Together, these two fields are called "cryptology" though often the term "cryptography" is commonly used to identify the same act of studying encryption and decryption.[Song Y.Yan, 2002]

## 2.5 The Modern Mathematical of Cryptography

### 2.5.1 Probability Theory

Probability Theory plays an important role in cryptography. Here, some basic ideas from Probability Theory are introduced. Some basic notions that concerns probability definition, probability space and random variables were respectively reviewed. There are many other textbooks on the probability theory, namely Feller[W.Feller,1957], Bauer[H.Bauer,1996], Gan Ylv[R.A Gangoli, D.Ylvsaker,1967], Gordon[H.Gordon,1997], and Renyi[A. Renyi,1970].

### 2.5.2 Definition of Probability

Assume an experiment has set $X = \{x_1,\ldots.x_n\}$ of n possible outcomes. Each time the experiment is performed, exactly one of the outcomes occur.

Each outcome is assigned to a real number P, between o and 1, which is called the probability of that outcome [Samuel S.Wagstaff Jr, 2003]. The summation of the probability of all of the outcomes must be 1. Write p ($x_i$) for probability of xi. So $0 <= p(x_i) <= 1$ for each i and $\sum^n i = p(x_i) = 1$

A subset E of X is called an event [Samuel S.Wagstaff Jr, 2003]. The event E "happens" if the outcome of the experiment is in E. The probability of event E is defined to be sum of the probabilities of the outcome in E, that is E = ∑ x = E p(x). It easy to see that 0 <= p(E) <= 1 and if the probability E doesn't happen , then it is 1 –p(E).

Suppose a known-plaintext attack is made on a cipher with 1,000,000 possible keys. M and C are given and one needs to find the key K for which M = $D_k$ (c) or C = $E_k$(M). If one of the 1,000,000 keys is picked, then in order to determine whether it is a correct key is by testing whether M = $D_k$(c). If it is assumed that the keys are equally likely to be chosen, then each key has the probability of $10^{-6}$ of being the correct one [Samuel S.Wagstaff Jr, 2003].

The event $E_1$ U $E_2$ is the union of two sets $E_1$ and $E_2$. The event $E_1$ U $E_2$ happens if either of two events $E_1$ and $E_2$ happens, and if outcome is in either set. The event $E_1 \cap E_2$ is the intersection of two set $E_1$ and $E_2$ . The event $E_1 \cap E_2$ happens if both of two events $E_1$ and $E_2$ happens, and if outcome is in both sets.

Event $E_1$ and $E_2$ are called mutual exclusive events if they are disjoint sets, that is $E_1 \cap E_2$ is empty. If $E_1$ and $E_2$ are disjoint, then the probability that either $E_1$ or $E_{2\ \text{to}}$ happen is p($E_1$ U $E_2$) = p($E_1$) + p($E_2$).

As an example of these principles [Samuel S.Wagstaff Jr, 2003], assume the keys for the cipher of the preceding paragraphs were 6-digit integers. To find the probability when the first digit of the key is either 3 or a 6, take an example of $E1_{\text{to}}$ be the event, "the first digit to be 3" and $E_{2\ \text{to}}$ be the event, with "the first digit as 6". Since there are 100,000 six digit numbers whose first digit is a 2, $p(E_1) = 100,000 / 1,000,000 = 0.1$. Likewise, $p(E_2) = 0.1$. Since the first digit can not be both a 2 and a 5, then the events are mutually exclusive and the sum of these probabilities is 0.2.

Assume $E_1$ and $E_2$ are two events. Assume $p(E_2) > 0$. The defined conditional probability of $E_1$ given $E_2$ to be $p(E_1 | E_2) = p(E_1 \cap E_2) / p(E_2)$

Let $E_1$ be the odd, ”the first digit is 3” and $E_2$ be the odd, “the first digit is odd”. Thus $p(E_1) = 0.1$ . Likewise, $p(E_2) = ½$ due to half of first digits are odd. However, if the first digit is 3, then the first digit is odd. Therefore, $E_1 \cap E_2 = E_1$ and $p(E_1 \cap E_2) = p(E_1) = 0.1$. The conditional probability [Samuel S.Wagstaff Jr, 2003] is

$p(E_1 | E_2) = 0.1 / 0.5 = 0.2$

The formula defining condition probability [Samuel S.Wagstaff Jr, 2003] is the form

$p(E_1 \cap E_2) = p(E_1 | E_2) p(E_2)$.

The sample space is the set of all possible outcomes of E, which each having the probability p(E). A random variable is a real value function defined on a sample space [Samuel S.Wagstaff Jr, 2003]. If $x_1,x_2,\ldots$.are all of possible value of r(E), then the probability distribution of r is the function f defined by $f(x_1) = p (r(E)) = x_i)$, the probability that $r(E) = x_i$ . Thus, $f(x_i)$ is the sum of p(E) for all outcomes E for which $r(E) = x_i$ . Several random variable $r_1,r_2,r_3\ldots r_k$ are called mutually independent[17] if for any possible value y1,y2,y3… yk then probability that $r(E) = y_i$ for every $1 <= 1 <= k$ equal the product $p (r_1(E)) = y_1 )$ $p (r_2(E)) = y_2)$ ……… $p (r_k(E)) = y_k)$.

On the other hand, the mean, variance and standard deviation of random variable as a below:

1. The mean or expected value E(r) of random variable r with value $x_1, x_2 \ldots$and probability distribution F(R), then expected values is

$E(F(r)) = \sum F(x_i) F(x_i)$.

2. The variance Var(r) of random variable r with expected with value μ is

$Var(r) = E((r - \mu)^2 = E(r^2) - \mu^2$

3. The standard deviation is the nonnegative square root of the variance of r.

σ(r) = square root of Var(r).

There are several theorems in the Probability Theory such as Markov inequality, Chebyshev's inequality, Hoefding inequality, and a law of large numbers. More information about these theorems can be found in textbooks written by [Wagstaff, 2003], [Delf and Knebl, 2002], [Goldreich, 1999].

## 2.6 Secret and Public Key Cryptography

### The Concept of Public Key Cryptography

In this section, the concept of the public key cryptography is introduced, followed by a discussion on  examples of public-key cryptosystems, such as IPSec, PKI. All other specific cryptography application, are also discussed in the next later sections. In 1976, W.Diffie and M.E. Hellman introduced a public key method for key agreement which is used up till this day [W.Diffie, M.E. Hellamn, 1976]. The discussion on RSA Cryptosystem is in Section 4.4.3

### The Concept of Secret-key Cryptography

In conventional cryptosystem, the same key is called a secret key in both encryption and decryption. By this it means, an individual who has enough information to encrypt a message will automatically have enough information to decrypt message [Wade Trappe, Lawrence C. Washington, 2002]. For example, Advance Encryption Standard(AES) and Data Encryption Standard (DES) is also example of symmetric key algorithm. Cipher are classified according to how the key is used to encipher the plaintext message M.

## 2.7 Some Specific Cryptography Application

### Digital Signatures

Digital signatures are widely used in the public-key cryptography. Without digital signatures, a hacker can forge the documents signature. The signer, who is the person who signs the document, must match the message. Digital signatures will widely be used on electronic transactions and E-cash. Digital schemes consists the signing process and the verification [Trappe and Washington, 2002]. In the following sub-section, digital signatures will be deftly discussed by examples, hash functions, the digital signature standards (DSS), “birthday attacks” on discrete logarithms.

### Hash Function

Hashing is plays an important role in not just algorithm techniques and database design,but also in cryptography. Weak and strong hash functions will be introduced to readers and its use as an aid to generating random number. Apart from that, some examples of hash functions will also be provided. A one-way function has the following property; that given any y in the range of f it is infeasible to compute any x which f(x) = y. One example is a sparse polynomial of high degree modulo a large prime. [Trappe and Washington, 2002]. Some examples of hash functions are SNEFRU,N-Hash,MD4,MD5 and SHA. MD5 produces a 128-bits message digest, while SHA's message digest has 160 bits, and so is even more resistant to birthday attacks. [Trappe and Washington, 2002]

**Digital Signature Standard (DDS)**

DSS used the Digital Signature Algorithm (DSA). The DSS/DSA is similar to Schnorr's signature scheme [C.P. Schnorr,1990 ] and ElGamal's signature scheme [T.ELGamal,1985].

**Birthday Attack**

The birthday attack notion is based on the birthday paradox. It says that the probability of two people in a group sharing the same birthday is greater than ½ if the group is chosen at random and has more than 23 members [Han Delf, Helmut Knebl,2002]. The birthday attack is used for a hash function with output twice. So, this attack can be prevented by choosing large prime numbers [Wade Trappe, Lawrence C. Washington, 2002].

## 2.8 Digital Watermarking Model

### 2.8.1. Watermarking Model

Figure 2.2 shows the model of a digital watermarking system [Erkin, Piva, Katzenbeisser, Lagendijk, Shokrollahi, Neven, Barn, 2007]. The inputs of the system is the original host signal **X** and some application dependent to-be-hidden information, here represented as a binary string **B** = [$b1$,$b2$, . . . , $b_L$], with $bi$ taking values in [0, 1]. The embedder inserts the watermark code **B** into the host signal to produce a watermarked signal $\mathbf{X_w}$, usually making use of a secret key sk to control some parameters of the embedding process and allow the recovery of the watermark only to authorized users.

The watermark channel takes into account all processing operations and (intentional or non-intentional) manipulations that the watermarked content may undergo during distribution and use. As a result, the watermarked content $\mathbf{X}_w$ is modified into the “received” version **X**. Based on **X**, either a detector verifies the presence of a specific message given to it as input, thus only answering *yes* or *no*, or a decoder reads the (binary) information conveyed by the watermark. Detectors and decoders may need to know the original content **X** in order to retrieve the hidden information (non-blind detector/decoder), or they do not require the original content (blind or oblivious detector/decoder).

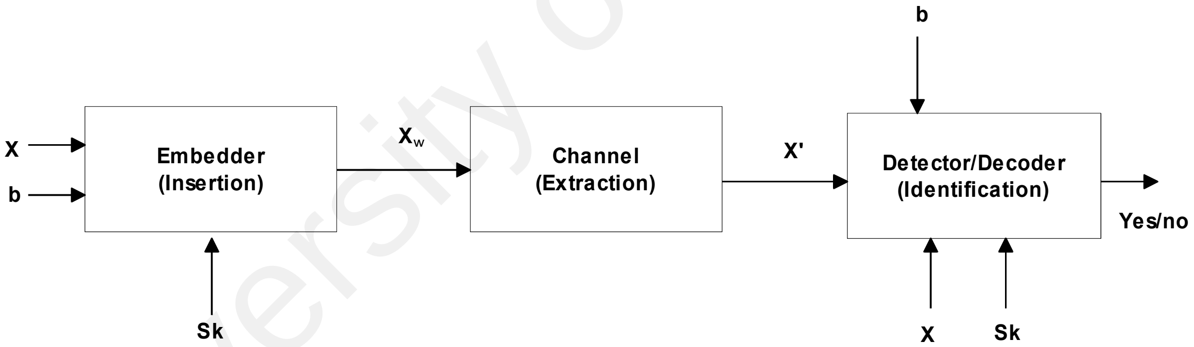


**Figure 2.2: The Model  of Digital Watermarking** [Erkin, Piva, Katzenbeisser, Lagendijk, Shokrollahi, Neven, Barn, 2007]

### 2.8.2 Study on Watermarking Algorithm

Digital watermarking technology processing contains two cores: **watermark embedding algorithm and detection algorithm**. Robustness, imperceptibility, invisibility and security of digital watermarking generally is the focus of the requirements. Watermarking extracting can be regarded simply as the reverse of watermarking embedding.

In general, the digital watermarking algorithm is composed of three parts: **The watermark** (payload), **The encoder** (marking insertion algorithm**), The decoder and comparator (**verification or extraction or detection algorithm)**.**

Each owner has a unique watermark or an owner can also put different watermarks in different objects the marking algorithm incorporates the watermark into the object. The verification algorithm authenticates the object determining both the owner and the integrity of the object.Watermark insertion involves watermark generation and encoding process.

The watermark can be a logo picture, sometimes a binary picture , sometimes a ternary picture ; it can be a bit stream or also an encrypted bit stream etc. The encryption may be in the form of a hash function or encryption using a secret key [Liu and Wen, 2004]. The watermark generation process varies with the owner.

As figure 2.3, in the encoding process both the original data and the payload data are passed through the encoding function. The payload signal and the original host signal now together occupy space, which was previously occupied only by the host signal. For this purpose either the original data is compressed or redundancy in digital content is explored to make space for the payload [Macq and Quisquater, 1995].

Extraction is achieved in two steps. First the watermark or payload is extracted in the decoding process and then the authenticity is established in the comparing process.

The decoding process can be itself performed in two different ways. In one process the presence of the original unwatermarked data is required and other where blind decoding is possible. Figure 2.4 and Figure 2.5 show the two processes. A decoder function takes the test data  whose ownership is to be determined and recovers the payload.

Figure 2.6 illustrates the comparing function. In this process the extracted payload and the original payload are passed through a comparator. The comparator output C is the compared with a threshold and a binary output decision generated. It is 1 if there is a match i.e. $C >= \delta$ and 0 otherwise.A watermark is detectable or extractable to be useful [Graham Shaw (2000)]. Depending on the way the watermark is inserted and depending on the nature of the watermarking algorithm, the method used can involve very distinct approaches. In some watermarking schemes, a watermark can be extracted in its exact form, a procedure we call **watermark extraction**. In other cases, we can detect only whether a specific given watermarking signal is present in an image, a procedure we call **watermark detection**. It should be noted that **watermark extraction can prove ownership** whereas **watermark detection can only verify ownership** [Yu and Lu, 2005].

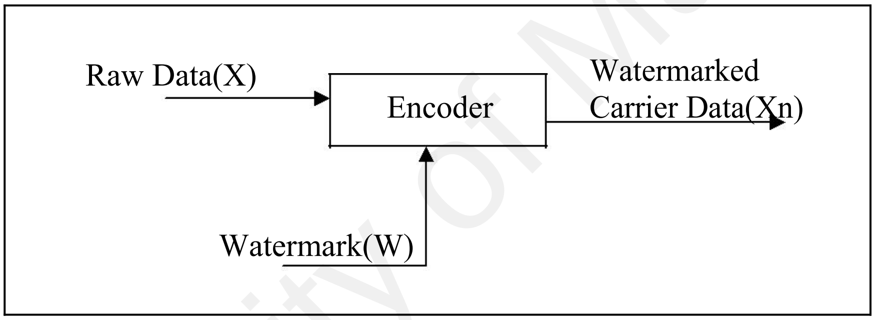


**Figure 2.3 Watermark Encoder (Embedder)**

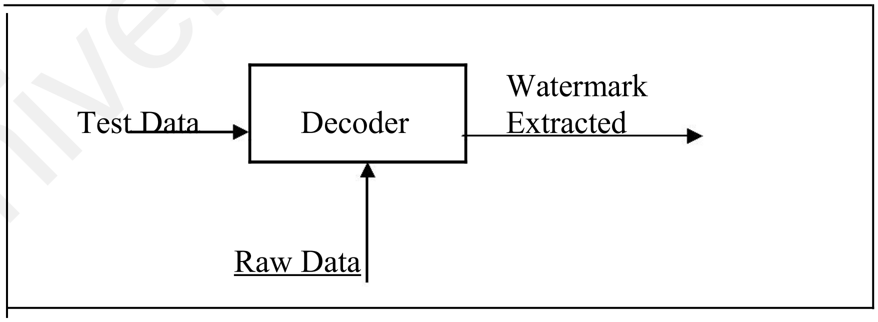


**Figure 2.4 Simple decoding Process**

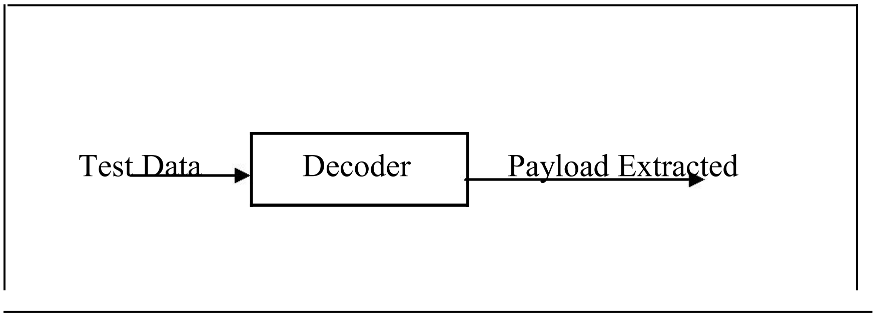


**Figure 2.5 Blind Decoding Process**

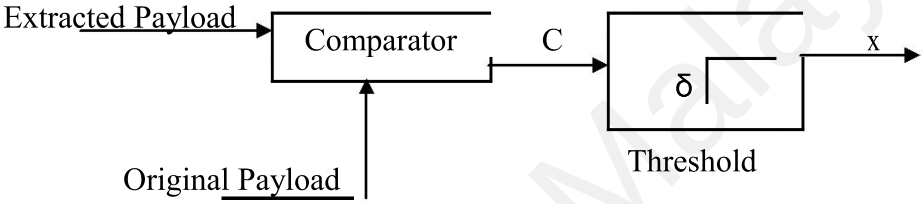


**Figure 2.6 Comparing Process**

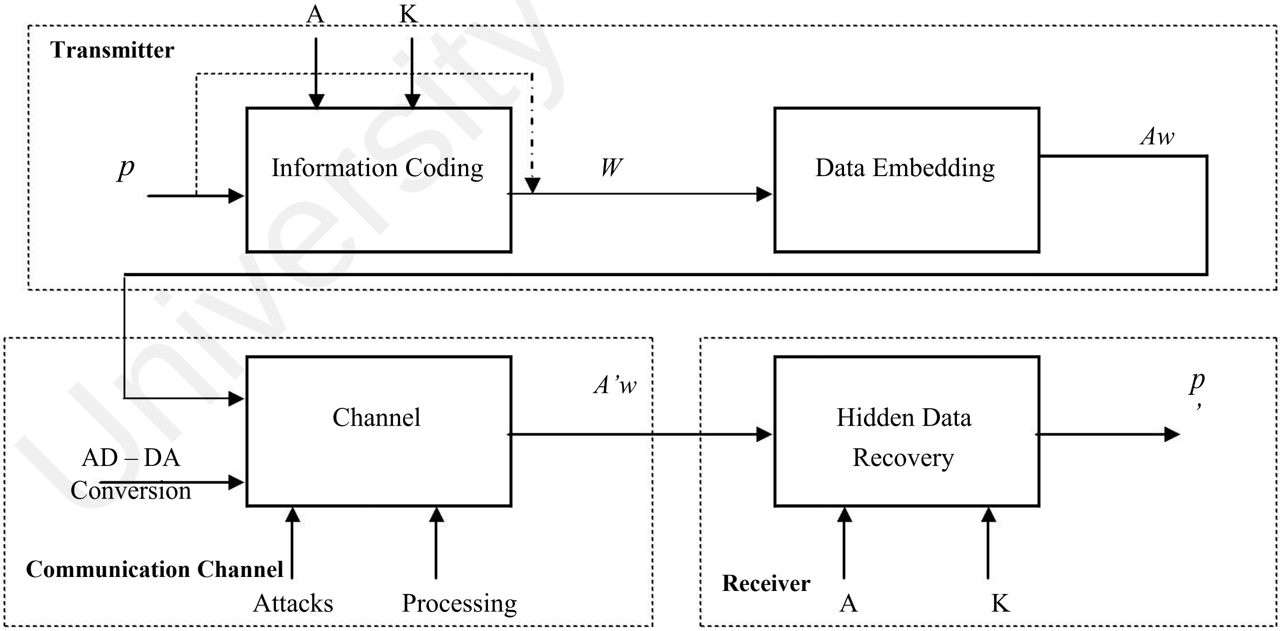


**Figure 2.7 Overall Picture of Data hiding System**

In Figure 2.7 the host data is depicted as asset *'A'*. *'A'* may be an audio file, a still image, video or a combination of audio and video. Throughout the thesis though only still images have been handled. So in the rest of the work *‘A’* refers to still images. The information embedded is depicted as payload *‘p’*. Then *p* is transformed in a watermark signal *‘w’* (optionally *p* = *w*). The embedding module may accept a secret key *‘K’* as an additional input. Watermarked asset *‘Aw’* is formed as a result of this encoding. The key *‘K’* introduces some secrecy within the embedding step. Due to possible attacks *‘Aw’* is transformed to *‘A'w’*. Finally the decoder/detector recovers the hidden information from 'A'w’.

## 2.9 Overview of Text based Watermarking techniques

There are now several types of text based watermarking techniques. Previous works on digital text watermarking can be classified in the following categories; *an image based approach, a syntactic approach and a semantic approach* as shown in Figure 2.8. The description of each category and the work done accordingly is explained in detail the next section.

Text watermarking algorithms are also dependent on text size, its language, rules, grammar, conventions and writing styles. In the past, various text watermarking techniques were proposed. These include text watermarking using text images, synonym based, pre-supposition based, syntactic tree based, noun-verb based, word and sentence based, acronym based, typo error based methods etc. The description of work done in each category of text watermarking is as follows sections:

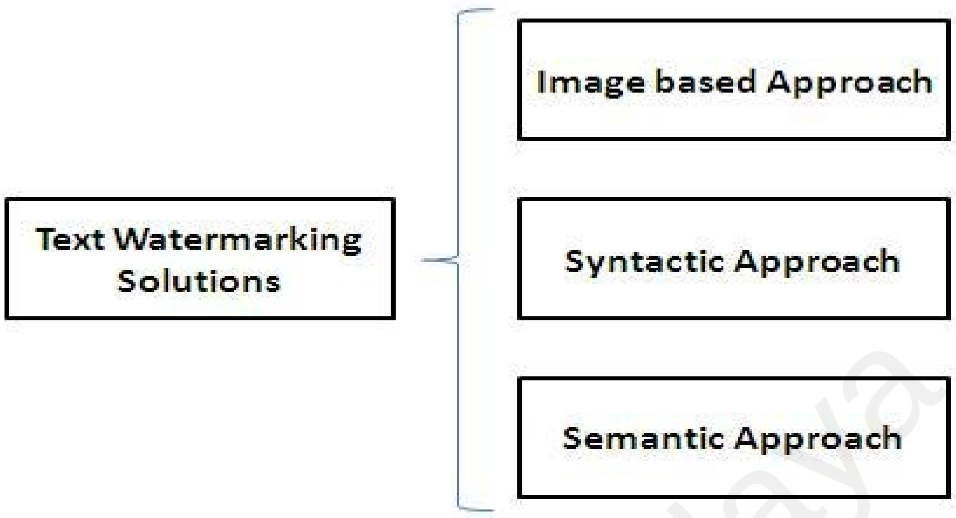
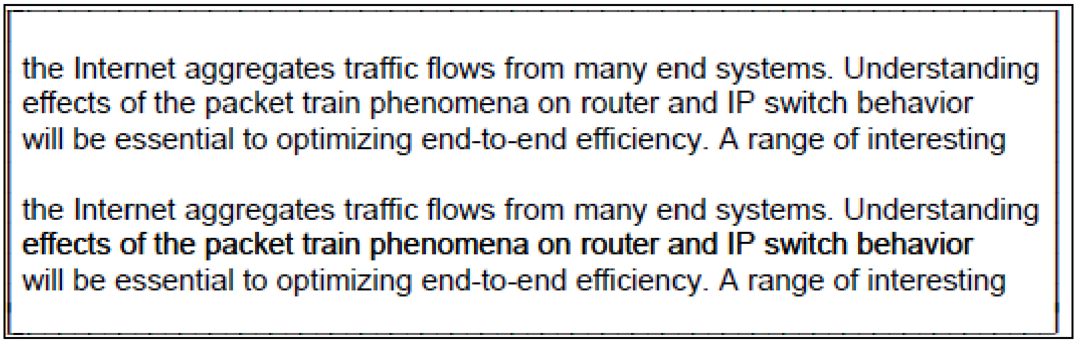


**Figure 2.8: Text Watermarking Solutions [Jalil, Mirza(2009)]**

### 2.9.1 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 a text document by using text image [Brassil,Low,Maxemchuk, O'Gorman,(1999)]. The first method proposed by Brassil 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 Figure 2.9. The detection algorithm is non-blind in which the original document should 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 both in non-blind and blind modes. The third method is the feature coding algorithm in which certain text features are altered to encode watermark bits in the text.



the Internet aggregates traffic flows from many end systems. Understanding
effects of the packet train phenomena on router and IP switch behavior
will be essential to optimizing end-to-end efficiency. A range of interesting

the Internet aggregates traffic flows from many end systems. Understanding
effects of the packet train phenomena on router and IP switch behavior
will be essential to optimizing end-to-end efficiency. A range of interesting

**Figure 2.9: Line Shift Coding [J. T. Brassil, S. Low, and N. F. Maxemchuk(1999)]**

### 2.9.2 Syntactic Techniques

A 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 structures depending on type of language and its conventions. Applying syntactic transformations on a text structure to embed watermark had also been one of the approaches towards text watermarking in the past.

Mikhail. J. Atallah, et al. first proposed the natural language of watermarking scheme using the syntactic structure of text [Atallah,McDonough,Nirenburg, (2001)] 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 Natural Language Processing (NLP) techniques are used to analyze the syntactic and the semantic structure of text while performing any transformations to embed the watermark bits.

Hassan et al. proposed the natural language watermarking algorithm by performing the morphosyntactic alterations to the text [Hassan M. Meral et al.,(2009)]. The text is first transformed into a syntactic tree diagram where the hierarchies and the functional dependencies are made explicit and then watermark is embedded. The watermarking process is shown in Figure 2.10.

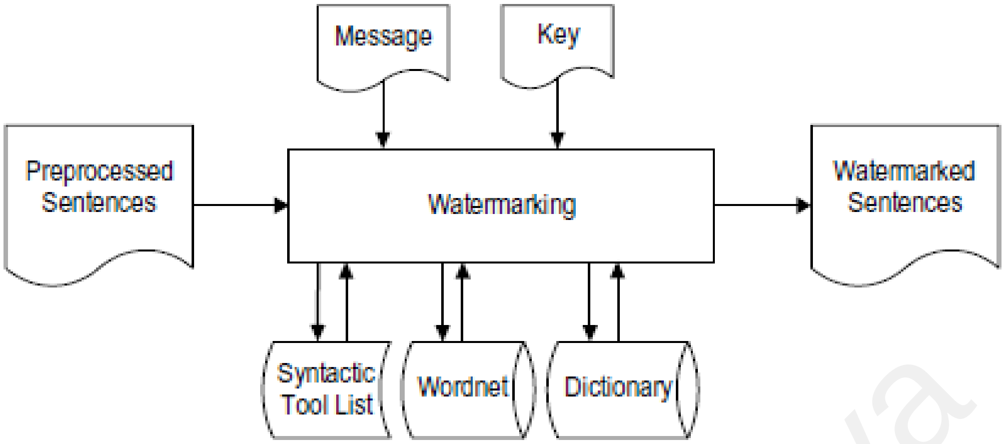


**Figure 2.10: Syntactic Sentence Level Watermarking [Hassan M. Meral et al.(2009)]**

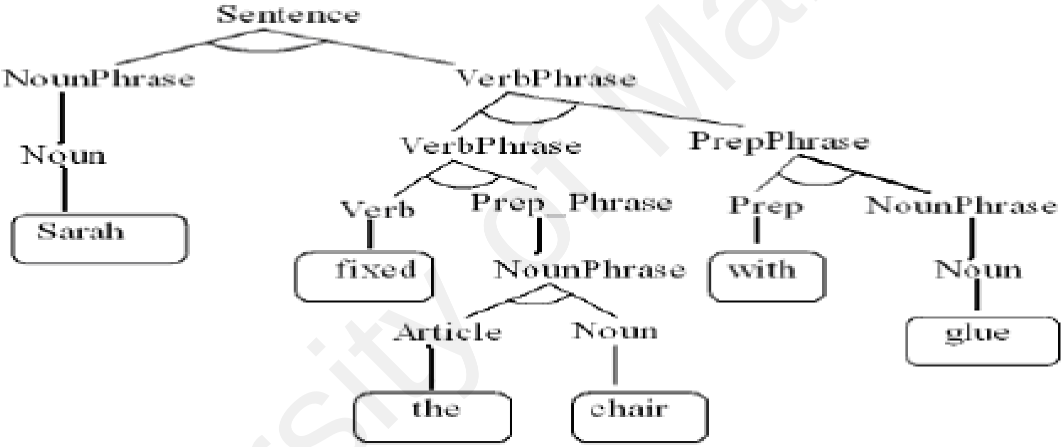


**Figure 2.11: Noun verb based tree of sentence “Sarah fixed the chair with glue”[Xingming Sun, Alex Jessey Asiimwe(2005)]**

### 2.9.3 Semantic Techniques

The semantic watermarking schemes focuses 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 watermarks in the text. Atallah et al. was the first to propose the semantic watermarking schemes in 2000 [Atallah,McDonough,Nirenburg,Raskin(2000)(2002)]. Later, the synonym substitution method was proposed where the watermark is embedded by replacing certain words with their synonyms, without changing the context of text [Topkara,Atallah(2006)]. Xingming, et al. proposed noun-verb based technique for text watermarking [Sun, Asiimwe,(2005)] which exploits nouns and verbs in a sentence parsed with a grammar parser using semantic networks. Figure 2.11 shows the parse tree for noun-verb based transformation.

## 2.10 Chapter Summary

This chapter has provided the background information on digital watermarking and cryptography; such as overview of watermarking, definition of Random Number Generator, cryptography, the objective of cryptography, overview of cryptography concept, the mathematical of modern cryptography, secret and Public Key Cryptography, and some specific cryptography with text watermark. Additionally, this chapter also provided a discussion on digital watermarking model, overview of text based watermarking, client-side watermarking embedding and secure watermark detection. The succeeding  chapter will described the literature review of this research paper .

# CHAPTER 3: LITERATURE REVIEWS ON RELEVANT WORKS

## Section 3. The Relevant Facts and Basic Notation

In this section, related work in Software (Pseudo) Random Number Generator, Hardware Random Number Generator, the properties of these random number generator, some random number generator related to cryptographic applications, some well known cryptographic applications, practical random number generator in the real world cryptographic applications, encryption package for secure password and message, and security problems in cryptographic applications, will be surveyed accordingly.

Besides that, in the Section 4, a survey on related work in the Digital Watermarking Model, overview of Text-based Watermarking Techniques, Client-side Watermark Embedding, Secure Watermark Detection, future research directions in Digital Watermarking, Pseudo-Random Number Generator (PRNG) Algorithms for Cryptography Applications, future research directions in Pseudo-random number generator (PRNG) will also be carried out respectively.

### 3.1 PRNG and Text -based Watermarking

**PRNG** is cryptographic algorithm used to generate random numbers that cannot be predicted by a hacker. Therefore, PRNG uses an algorithm to generate truly and genuine random numbers that used for stimulation of statistical methods and random processes. Thus, it is appropriate to use a Pseudo Random Bits Generator (PRBG) due to perfect randomness [J.Linn, 1988] for all cryptographic applications. A random number generation performed well in the cryptographic security [B.A.Wichmanna,I.D.Hill.(2006)]. PRNG plays an important role in many application fields and systems namely Statistical Stimulation

System, Military Secret Signal Communications System, Quantum Cryptography, Any Scientific Computation Analysis and many more.

**Text-based Watermarking** is the process of embedding digital document content copyright owner information. Text watermarking aims is embedding confidential information within text itself and protected copyright authorship and authentication [Cox, Miller, Bloom (2002)]. Several digital text based watermarking techniques will describe in the Section 3.7.

### 3.1.1 Software (Pseudo) Random Number Generators and its examples

PRNGs require "seeds" that are bits of data collected from the clock, running processes, status registers, key press timing, and mouse movements which is used as an operand in a mathematical computation to create the pseudo random number. [Durrant (1999)]. Practical PRNGs are designed to be cryptographically secured, and these include [refer to CSPRNG website] the Yarrow algorithms such as UNIX special file /dev/random ,OpenBSD , Mac OS X, CryptGenRandom in Microsoft's Cryptographic API, and the Python function *urandom*.

### 3.1.2 Hardware (Physical) Random Number Generators and its examples

A hardware (Physical) random number generator which generates real random numbers [Durrant (1999)] is also well-known as the **True Random Numbers (TRNG)**. Hardware random number generators are non-deterministic due to they are not susceptible to intrusion or disclosure. The major application is for *encryption*. Some examples of hardware random number generators is the New Zealand Lotteries Commission, and Extra Sensory Perception (ESP)[Robert Davies(2000)]. Another example of the commercial hardware random number generator used is Protego, Hotbit, Sun PCI Crypto Accelerator.

**Table 1: Comparisons between characteristics of PRNG and TRNG**

| | PRNG | TRNG |
|---|---|---|
| **Principle of operation** | Mathematical formulas,<br>algorithms, no entropy | Physical devices,<br>quantum based entropy |
| **Deterministic** | Yes | No |
| **Predictable** | Yes | No |
| **Periodic** | Yes | No |
| **Main field of usage** | Simulation, Cryptography<br>(CSPRNGs) | Security/Cryptography,<br>Gambling |
| **Output type** | Stream of numbers<br>(int, float) | Stream of bits<br>(boolean) |
| **Speed (output bit rate)** | Very fast<br>(aprox. GB/sec) | Slow (KB/sec) /<br>Fast (several MB/sec) |

Random number generators can be classified in two main categories: psudo-random number generators – PRNG (also called these deterministic or software **(Pseudo) Random Number Generators**) and true random number generators – TRNG (also called physical or hardware **(Physical) Random Number Generators**). Table 1 presents a brief comparison between PRNGs and TRNGs. Furthermore, there are also intermediate classes between these two extremes, such as the unpredictable random number generators which rely on the nondeterminism of user-computer interaction or on the complexity of the underlying phenomenon while executing a set of deterministic instructions, or the hybrid generators which combine both TRNGs and PRNGs.

### 3.2 Advantages of PRNG and Watermarking combination

The reason on why the combination of PRNG and Watermarking is used is because it has several advantages. The following are the few advantages of combining Watermarking with PRNG techniques:

- The combination provides a more secured security due to complex algorithms (when the two Watermarking is combined with a PRNG).

- A hacker finds it hard and difficult to predict its public key due to the combination of RSA algorithms and Watermarking algorithms.
- It can protect users and software developers for authentication, hidden data/message and randomized content.

### 3.2.1 The Properties of These Random Number Generator Are a Major Interest for Cryptographic Applications.

Some properties that a sequence of random numbers might have are as follows

- The sequence looks random. It passes the statistical test of randomness.
- The sequence is unpredictable; knowing the algorithm and previous bits, one will not be able guess the next bit, although the sequence can be reproduced. Such sequences of random numbers may be used as key streams for stream ciphers.
- The sequence cannot be reliably reproduced: If the random number generator (RNG) is run twice with the same input(as closely as possible), two different random sequences will be produced or generated. The sequence cannot be compressed. Sequences of such sort might be used to select a secret key, like a large prime.
- Equal probability: An equal probably random number must be tested so that their output shall be statistically acceptable.

## 3.3 Combination of PRNG and Watermarking for Improved Security

The combination of PRNG and Watermarking improves security as pointed below:

- PRNG is complex cryptographic algorithms that is used to generate a random number that cannot be guessed by an attacker, thus, making it a more secured security method for encrypting and hiding text data in digital media. As referred to the Section 6.2, the Modification and Improvement of using a combination of PRNG and Watermarking is further described in detail.
- A hacker finds it hard to predict its public key due to the combination of RSA algorithms and Watermarking algorithms. A maximum to 2048 bit RSA key(Common Cipher strength 128-bit keys) was modified and this proved that it was impossible for a hacker to break its public key.
- It can protect users and software developers for authentication, hidden data/message and randomized content. It is not only encrypts/decrypts data/message in PRNG but also allows no modification/deletion of its content document in Watermarking.
- In Section 5, the proposed implementation model is further described.

**Some Random Number Generator Related To Cryptographic Applications.**

### 3.3.1 Pseudo-Random Number Generator

PRNG produces good RNG output, where random number are unpredicted and uniformly distributed to random variables. PRNGs are e*fficient,*, *deterministic, periodic*. Therefore, PRNGs are suitable for simulation and modeling applications, data encryption and gambling.

### 3.3.2 Lottery Number Generator

According to the Pennsylvania Lottery website, the Pennsylvania Lottery uses an Automated Drawing Machine (ADM) with the utmost confidence and security by lotteries for many years. ADM will select winning tickets at random and not winning number in the Pennsylvania Lottery's Millionaire Raffle drawing.

### 3.3.3 Zero Knowledge Proof

**Zero Knowledge Proofs**:

Goldwasser, Micali and Rackoff(GMR) invented Zero knowledge proofs in 1985. In "The Knowledge Complexity Of Interactive Proof Systems" paper [Goldwasser, S.; Micali, S.; Rackoff, C. (1989)] with applications ranging from practical signature schemes to proving that many NP-complete problems are hard even to approximate. Zero knowledge proofs have found many applications such as authentication systems, cryptographic protocols, E-Voting, etc.

## 3 . 4 Key Researchers in the Combination of PRNG and Watermarking

There are some key published researchers in the combination of PRNG and Watermarking namely Jalil and Mirza's who researched and wrote papers titled "Text Watermarking using Combined Image-Plus-Text Watermark" [Jalil and Mirza(2010)], "A Novel Text Watermarking Algorithm Using Image Watermark "[Jalil, Jaffar, Mirza(2011)], "Text Watermarking by Syntactic Analysis[Kim (2008)], "Natural Language Watermarking using Semantics Substitution for Chinese Text"[Chiang(2004)], "Syntactic Tools for Text Watermarking"[Meral (2007)].

In Chiang's conference paper [Chiang (2004)], he proposed a synonym-based watermarking algorithm, which is context based rather than format. The synonym-based watermarking is suitable for Chinese textual messages. The algorithm was designed to select appropriate candidate terms for textual messages for use in the embedded watermark. The binary tree encoding methodology of the synonym-based watermarking algorithm selects suitable synonyms, and embeds the maximum watermarking bits. The encoding/decoding algorithm of the encryption key based encryption also is proposed, and enables higher watermarking resilience, besides extending the thesaurus to improve the reliability of synonym substitution. However, Chiang's method has a weakness; low data hiding capacity in the structure of sentences.

Therefore, Meral, in his paper [Meral(2007)], proposed syntactic tools for text watermarking that provides a good ground for the syntax-based natural language watermarking, with its relatively free word order possibilities and rich repertoire of morphosyntactic structures. The text-watermarking algorithm works by transforming raw sentences into their treebank representation and then into their syntactic trees. A software tool weaves through the text by checking the applicable tools and randomizing their occurrences. However, Meral's method has weakness too, due to lack of syntactic analysis and performance analysis.

Consequently, Kim's paper ([Kim (2008)] proposed the method of text watermarking for Koreans through syntactic analysis. His proposed method is useful for agglutinative languages such as Korean, Turkish and so forth where the properties of the languages have a syntactic constituent order that is relatively free. His experimental results show a better coverage on the sentences selected for embedding watermark bit which is 75% and information-hiding capacity is 1:14.05, which is worse than that of Chiang's (2004). However, his marked text keeps the same style, and it also shares the same information without, a semantic distortion. However, Kim's method has weakness; and hence cannot be applied to sentences constructed in English because it has no cryptography related issues.

Therefore, in Jalil and Mirza's conference paper [Jalil and Mirza(2010)], both researchers proposed combined image-plus-text watermark to fully protect the text documents. His results demonstrate dispersed tampering attacks on the text. In Jalil, Jaffar and Mirza's Journal paper[Jalil, Jaffar, Mirza(2011)], they proposed embeds the watermark image in the text logically, using embedding algorithm and then extracts it later by using an extraction algorithm. His results demonstrated dispersed attacks on the text. However, his method has weakness, because there is no combined PRNG with the text based watermarking and no cryptography related issues.

All of the above methods have no encryption. Therefore, this research paper proposes to use PRNG combined with text based Watermarking for Cryptography applications, which produces a more secured method for encrypting and hiding text data in digital media. This design methodology is further described in Section 4. Apart from that, Modification and Improvement is described in Section 6.2 on the use of combining PRNG and Watermarking. A comparison of existing text based watermarking is also described along with the proposed work on Table 6.9.

## 3.5 Current Challenges in the Combination of PRNG and Watermarking

The design of the robust watermarking technology is a complex non-trivial problem that requires taking into account many contradictory requirements such as robustness, visibility, capacity and algorithmic complexity. In many practical situations it is often very difficult or sometimes even impossible to satisfy all of requirements simultaneously. There are times when the definition of robustness is not well defined in the context of digital watermarking and that creates a problem, thus causing difficulty in countermeasure design. In order to overcome the mentioned above weakness, the problem needs to be solved in order to achieve the objectives. Even Jalil and Mirza (2010) proposed a novel text watermarking algorithm using combined image-plus-text watermark to fully protect the text documents. However, his method still needs improvement. Therefore, this research paper proposes to use combined PRNG and text based Watermarking for Cryptography application, thus producing a more secured method for encrypting and hiding text data in digital media. This method will be able to overcome Jalil and Mirza's (2010) weakness. The objectives are described in Section 1.2.

### 3.6 Pseudo-random Number Generator Algorithms for Cryptography Applications.

In PRNG algorithms, the **Linear Congruential Generators (LCGs)** is the best-known pseudo-random number generator algorithms. **Lagged Fibonacci Generator** is based on generalization of the Fibonacci sequence and an improvement on the 'standard' linear congruential generator. **Blum Blum Shub** is also a secured PRNG method. All will be further explained in detail in Section 4.7.2.

### 3.6.1 Requirement

- **Efficiency** – It can produce many numbers in a short time span.
- **Deterministic** - A given sequence of numbers can be reproduced.
- **Security** – It is the best met with a well analyzed block cipher in CTR (Counter) mode. Such a PRNG will leak a mere one bit of information after producing $2^{n/2}$ blocks of output, where n is the block size of the cipher in bits.
- **Predictability -** If the algorithm and the seed (i.e. the number that is used to start the generation) are known, then the numbers generated are predictable.

### 3.6.2 Some Background Linear Congruential Generators

A simple exampe of a PRNGs is Linear Congruential Generators (LCG):

$$X_{n+1} = (a * X_n + b) \text{ (mod m)}$$

Once the parameters a, b and m are chosen, the sequence of the generated pseudorandom numbers $(X_n)_{n=>1}$ will depend only on the initial value $X_0$. For example, the Unix rand generator is implemented by a LCG with parameters a = 1103515245, b = 12345 and m = $2^{31}$. The characteristic of PRNGs is that they produce a long sequence of random numbers from a short initial input, the so-called seed, by means of a completely deterministic algorithm. [Andrea Rock (2005)]

### Lagged Fibonacci Generator

Fibonacci sequence is defined as: [Tamer ÖZ (2006)]

$$X_n = X_{n-p+q} + X_{n-p} \bmod m$$

In a 32-bit computer, *m* is usually set to $2^{32}$ for efficiency. The indices *p* and *q* are chosen such that $x^p + x^q + 1$ is a primitive trinomial. The period of such a generator is $2^{31}(2^p - 1)$. This generator is fast and easy-to-implement. The most prominent feature is that a very long period can be achieved by choosing a large *p*. One derivative suggested by Marsaglia is to replace the addition in the formula with multiplication, i.e., $X_n = X_{n-p+q} \times X_{n-p} \bmod m$ [Marsaglia (1984)].

### Blum Blum Shub

Blum Blum Shub (BBS) is a pseudorandom number generator proposed in 1986 by Lenore Blum, Manuel Blum and Michael Shub. [Tamer ÖZ (2006)]

$$X_{n+1} = X_n{}^2 \bmod m$$

where $m$=$pq$, | $p$ | = | $q$ | (i.e. both have equal lengths), p and $q$ are distinct primes of the form 4$x$+3 [Blum (1986)]. With the assumption that the quadratic residuacity problem is intractable, the BBS generator is practically unpredictable. Unlike other deterministic generators, the period of a BBS generator depends on the seed and can only be worked out using an algorithm. Moreover, the BBS generator is much slower than other deterministic generators.

**Mersenne Twister**

Mersenne Twister (MT) is an extension of GFSR generator suggested by Makoto Matsumoto and Takuji Nishimura [Matsumoto (1998]. The general formula is $x_{k+n} = x_{k+m} \oplus ((x_k{}^u \mid x_{k+1}{}^l)\mathbf{A})$, $1 \leq m \leq n$. Let $w$ be the number of bits in a word and $0 \leq r \leq$ $w$-1. $x_k{}^u$ is the upper $w$-$r$ bits of $x_k$ where $x_{k+1}{}^l$ is the lower $r$ bits of $x_{k+1}$. The operator, | , concatenates the two operands to form a word. A is a $w$×$w$ binary matrix. When a vector of bits multiplies with A, say $x$A, the effect is equivalent to (i) first shifting $x$ to the right 1 bit position, (ii) if the rightmost bit of the original $x$ is 1, exclusive-or the shifting result with the last row of A.

### 3.6.3 Design of Pseudo-random number generator

In principle, any method of distinguishing between PRNG outputs and random outputs is an attack; in practice, we care much more about the ability to learn the values of PRNG outputs not seen by the attacker, and to predict or control future outputs. Figure 3.1 for a high-level view of a PRNG.

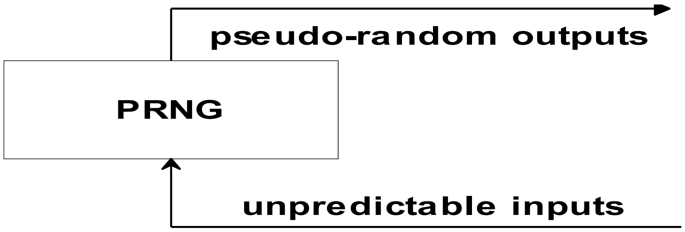


**Figure 3.1: Black-box View of a PRNG[Kelsey, Schneier, Wagner, Hall. (1998)]**

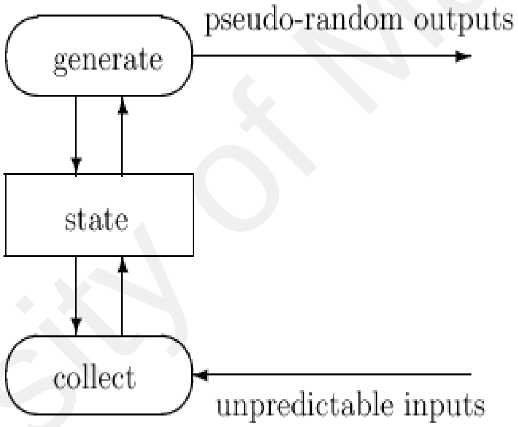


**Figure 3.2: View of internal operations for most PRNGs[Kelsey, Schneier, Wagner, Hall. (1998)]**

Figure 3.2 refines the terminology a bit. A PRNG often starts in a random state and must process many seeds to reach a secure state **S**. Upon request, it must generate outputs that are indistinguishable from random numbers to an attacker who doesn't know and cannot guess **S**. In this, it is very similar to a stream cipher. Additionally, however, a PRNG must be able to alter its secret state by repeatedly processing input values (seed).

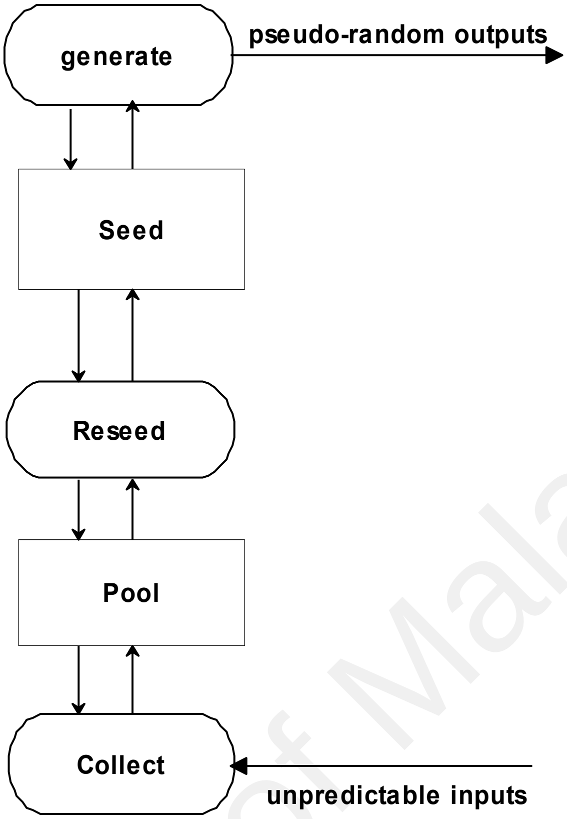


**Figure 3.3: Generalized PRNG, with Periodic Reseeding [Kelsey, Schneier, Wagner, Hall. (1998)]**

Figure 3.3 shows a PRNG with periodic reseeding, Figure 3.3 also depicts a possible architecture for the implementation of catastrophic reseeding. The part of the internal state that is used to generate outputs should be separated from the entropy pool. The generation state should be changed only when enough entropy is collected to resist iterative guessing attacks, according to a conservative estimate.

### 3.6.4 Design Model of Pseudo-Random Number Generator algorithm

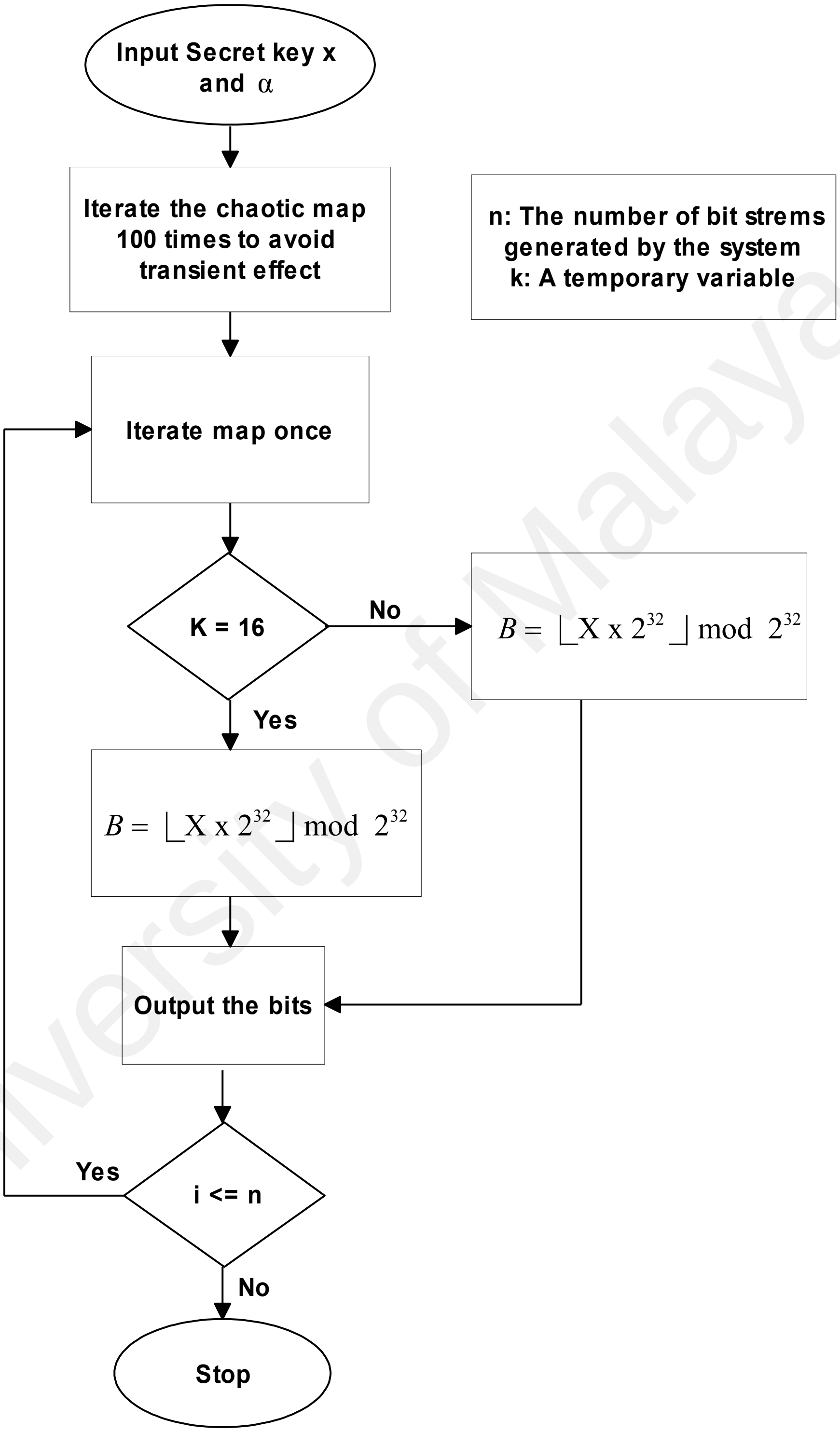

Input Secret key x and α
Iterate the chaotic map 100 times to avoid transient effect
n: The number of bit strems generated by the system
k: A temporary variable
Iterate map once
K = 16
No
$B = \lfloor X \text{ x } 2^{32} \rfloor \text{mod } 2^{32}$
Yes
$B = \lfloor X \text{ x } 2^{32} \rfloor \text{mod } 2^{32}$
Output the bits
Yes
i <= n
No
Stop

**Figure 3.4: Block Diagram of the Proposed PRNG**

**[Behnia,Akhavanb, Akhshani, Samsudin (2011)]**

A good random number generator must have some properties such as good distribution, long period and portability. Random number generators are commonly used in encoding algorithms[Djema, Barbot, Belmouhoub(2009) ; Belmouhoub, Djema, Barbot(2005)]. In the Pseudo-Random Number Generator algorithm[Behnia,Akhavanb, Akhshani, Samsudin (2011)] in figure 3.4, the combination of control parameter (α), initial condition (x) and degree of Chebyshev polynomial (N) of the chaotic system can be used as encoding keys. The parameter N is originally the degree of the Chebyshev polynomials and a change in this parameter would lead to a change in the structure of the whole map and its characteristics, such as chaotic behavior, interval and the attractors. Therefore, the generated sequences by two maps with a single digit difference in their N parameter are completely random in respect to each other. [Behnia,Akhavanb, Akhshani,Samsudin (2011)]

### 3.7 Summary

The literature review described the PRNG term in detail, discussed text based watermark, asserted the use of the PRNG and watermark combination, identified the key researchers in the combination of PRNG and Watermarking, and discussed current challenges in the combination of PRNG and Watermarking. Apart from that, the discussion in the chapter proceeded with a review of digital watermarking model and text watermarking techniques. The following chapter attempts to describe the methodology used in this thesis study.

# CHAPTER 4: DESIGN METHODOLOGY AND DESIGN METHOD, PROPOSED IMPLEMENTATION MODEL

## 4.1 Overview

The purpose of this research is to develop a text based digital watermarking that can be synthesized by using the Pseudo-Random Number Generator (PRNG) from 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 from 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.

### 4.1.1 Considerations

Most of the research work that has been done in this research field looks at increasing the output speed in computing the cryptographic results in a TRNG platform [D.Eastlake, S.Crocker and J.Schiller, 1994]. As more literature is read, it is notable that other written researches relates to work in [Jiezhao Peng, Qi Wu, 2009] and [Zunera Jalil and Anwar M. Mirza, 2011]. These two research papers discusses 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 this experimental research and experimental on design and implementation of text based watermarking combined with cryptographic techniques.

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

- Good security because of its complex algorithms, when Watermarking is combined with PRNG
- It causes a hacker to face difficulty in predicting its public key due to the combination of RSA and Watermarking algorithms.
- It is reusable and can be an updated source code whenever necessary.
- It can protect its user and software developer for authentication and data.

### 4.1.2 Design Method

**Design Tool**

RSA Key generator testing file consist of three JavaScript files as below:

- **Multiple-precision library (BigInt.js):** a suite of routines for performing multiple-precision arithmetic in JavaScript.
- **Modular reduction library(Barrett.js):** a class for performing Barrett modular reduction computations in JavaScript.
- **RSA library(RSA.js)**: suite of routines for performing RSA public-key computations in JavaScript.

The RSA key generator application is written in Delphi 8(Object Pascal language: See Figure 7.2). The user interface of RSA Key Generator Testing is also created by using a HTML Editor with integrated JAVA Scripts. (See Figure 7.1). RSA Key Generator is created for generating a new RSA key for public key and private key (See Figure 7.2). Apart from that open source text based digital watermarking algorithm is used as an experimental tool with JAVA to embed and extract algorithm for watermarking. (See Figure 7.3). The user interface of the Digital Watermarking tool is also based on an open source watermarking which is combined with PRNG.

This part of the manuscript describes the design method and experimental setups used for a Text-based Digital Watermarking Techniques; for the purpose of this PhD thesis as well as for the experimental results. Chapter 4 describes the automatic design flow for PRNG which is used to design a text based watermarking algorithm and Pseudo-Random Number Generator (PRNG) algorithms as mentioned in the thesis' manuscript. In the next chapter; Chapter 5 discusses in detail the architecture, methods, and experiments for a fully functional Text based Digital Watermarking Techniques embedded in Pseudo-Random Number Generator (PRNG) algorithms for Cryptography applications. This PRNG had been tested to be partially functional, and further tests are will be on Text based Digital Watermarking. A proposal is proposed on using Text-based Watermarking techniques by using Pseudo-Random Number Generator (PRNG) algorithms for Cryptography applications.

### 4.2 Proposed Model:

The implementation model is shown below in Figure 4.1. Although  the primary emphasis is on the evolution of a new text-based watermarking technique but the aim of the research is also on  providing a blended solution comprising of RSA-based encryption and text based digital watermarking ,dully added by compression before encryption.

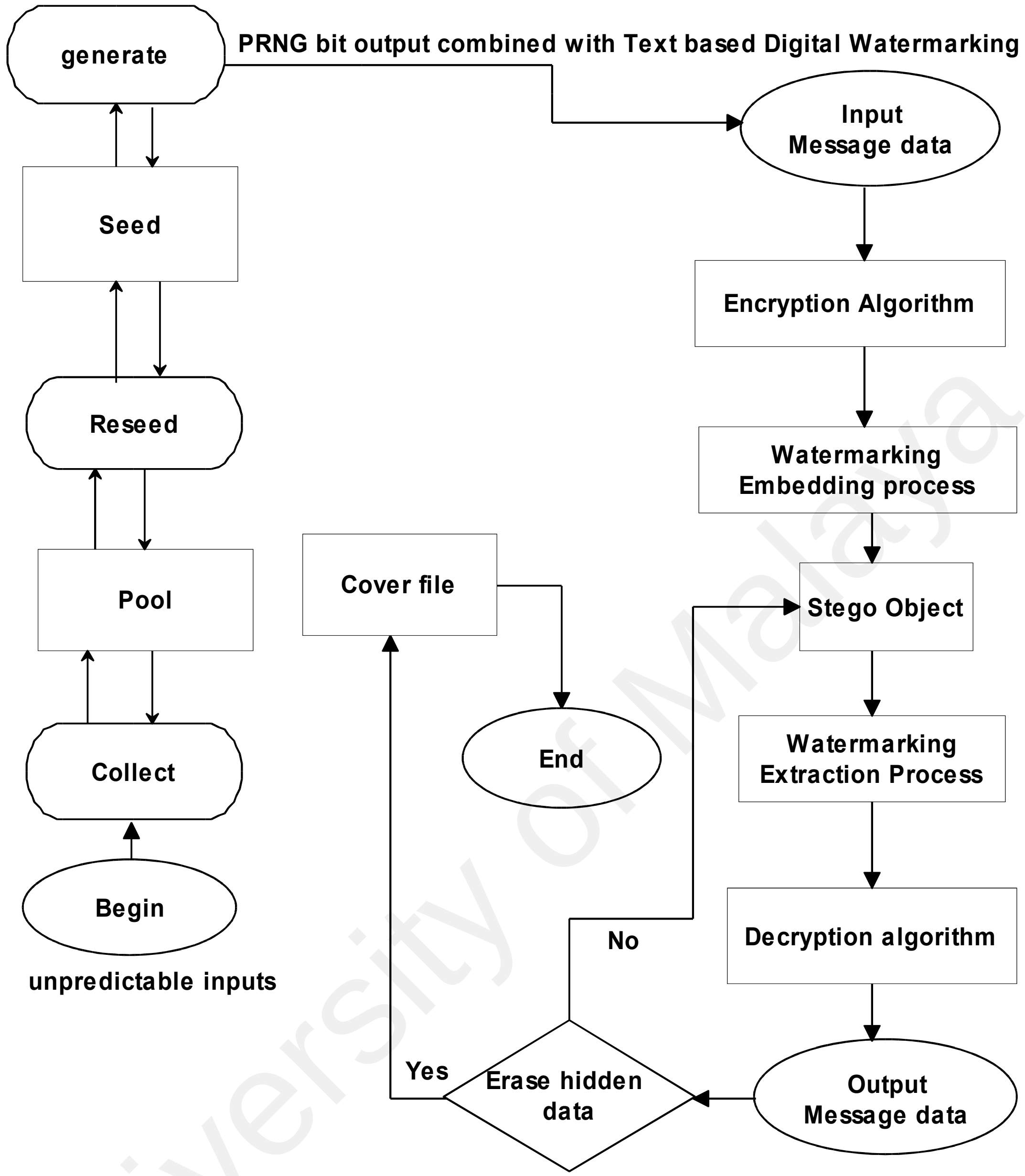


Figure 4.1: The proposed model of the Text-based Digital Watermarking Techniques embedded in Pseudo-Random Number Generator (PRNG) algorithms for Cryptography applications.

The encryption scheme which is based on RSA for PRNG in Figure 4.1. The above left part is Generalized PRNG with a periodic reseeding [Kelsey, Schneier, Wagner, Hall. (1998)], which is mentioned in Figure 3.11; a PRNG with a periodic reseeding that depicts a possible architecture for implementing catastrophic reseeding. The internal state part which is used to generate outputs, should be separated from the entropy pool. The generation state should be changed only when enough entropy has been collected to resist iterative guessing attacks, in accordance to a conservative estimate.

The above right part is the proposed implementation model of Text-based Digital Watermarking Techniques [Brassil,Low,Maxemchuk(1999)(2000)].Although the primary emphasis is on the evolution of a new Text-based Digital Watermarking Technique but the aim of the research is also on providing a blended solution comprising of Encryption and Text based Digital Watermarking, which is dully added by a PRNG algorithm before encryption. In order to focus more on the subject, a closest, possible, feasible and secured digital content solution is proposed with the combination of hidden data-plus cover (known as stego object) that holds the hidden information upon watermarking extraction process. No message data can be erased upon doing the RSA decryption algorithms. The Pseudo-Random Number Generator (PRNG) algorithms will be used for embedding the best available compression with cryptographic hash function (SHA-2) and encryption algorithms/techniques (RSA encryption/decryption method).

Some solutions to this problem are proposed and the research contributions are listed as below:

- A better understanding of Pseudo Random number primitives will make it easier to design and use PRNGs securely.
- To implement text based watermark, demonstrates the complexity in making a tradeoff between robustness and imperceptibility.
- To propose a new text based watermarking method by using Pseudo-Random Number Generator (PRNG) for Cryptography estimates robustness and

imperceptibility requirements at embedding stage, and actual robustness at the decoding stage.

### 4.3 Strength of the Proposed System

**Advantages and Disadvantages:** Table 4.1 summarizes the advantages and disadvantages of the existing text based data hiding methods and proposed enhancement.

**Table 4.1 – Advantages and Disadvantages of EXISTING and ENHANCED METHOD(s)**

| **Techniques** | **Advantages** | **Disadvantages** |
|---|---|---|
| Image-based Techniques | Variant data hiding | Eye catching |
| Syntactic Techniques | Difficult to detect | Loose format if its save as text |
| Semantic Techniques(Proposal work) | Faster Speed | Language specific |

As seen in Table 4.1, the advantages of existing and enhanced methods are explained.

For image-based techniques, Brassil proposed three methods; the first method is the line-shift algorithm which moves a line upwards or downwards (left or right) depending on the binary signal (watermark) which needs to be inserted as shown in Figure 3.6[Brassil,Low,Maxemchuk, O'Gorman(1995),(1999)]. The second method was the word-shift algorithm which moves words horizontally thus, expanding spaces to embed the watermark. The algorithm can operate both in non-blind and blind modes. The third method is the feature coding algorithm method in which certain text features are altered to encode watermark bits in the text. Therefore, as per-discussed, there are variant data hiding methods. However, due to Line Shift algorithms, its disadvantages are rather obvious.

As for syntactic techniques, Mikhail. J. Atallah, et al. first, proposed the natural language watermarking scheme to use the syntactic structure of a text [Atallah,McDonough, Nirenburg, Raskin(2000),(2001)] 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 the watermark. The Natural Language Processing (NLP) techniques are used to analyze the syntactic and the semantic structure of text while performing any transformations to embed the watermark bits.

Hassan et al. proposed on using the natural language watermarking algorithm by performing the morphosyntactic alterations to the text [Hassan M. Meral et al.,(2009)]. The text is first transformed into a syntactic tree diagram where the hierarchies and the functional dependencies are made explicit and watermark is embedded. The watermarking process is shown in Figure 3.7.Its advantages show that it is difficult to detect absence of the original text. The disadvantage is that it loses its format if it is saved as a text due to the syntactic techniques.

As for semantic techniques, 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 structures, grammatical rules, etc. are exploited to insert watermark in the text. Xingming, et al. proposed a noun-verb based technique for text watermarking [Xingming Sun, Alex Jessey Asiimwe,(2005)] which exploits nouns and verbs in a sentence parsed with a grammar parser using semantic networks. Figure 3.8 shows the parse tree for noun-verb based transformation. And so, its advantage is its fast speed for using the semantic structure of text to embed the watermark. The disadvantage is that English language's specific properties and text contents, like verbs, nouns, prepositions, words spelling, acronyms, sentence structures, grammatical rules, etc. are exploited to insert watermark in the text.

**Table 4.2 – Advantages and Disadvantages of Proposed Enhanced Algorithms**

| **Advantages** | **Disadvantages** |
|---|---|
| Performance speed | fractionally slow than existing techniques |
| Reliability | Uncertain runtime error |
| Simplicity | Less complicate in watermarking processing |

The table above discusses the measured speed, reliability and simplicity of the scheme along with tests on the robustness of watermarking scheme.

In Table 4.2, the proposed enhanced algorithms are explained, emphasising on the advantage of performance speed for faster watermarking embedment. The disadvantage is fractionally slow than existing techniques due to runtime error due to unexpected occurrence. The second advantage is the reliability due to expected embedded and detected watermarking.

However, its disadvantage remains because of uncertain runtime error due to unexpected occurrence. The third advantage is the simplicity which means that it is an easy to use application. However, the disadvantage is that it can be a more or less complicated in watermarking processing.

Due to the RSA key generator generating a key to be placed in this public key in digital based watermarking, so the robustness of the watermarking scheme will not allow nor permit a hacker to do a prediction on this key and there will be no deletion and no modification can performed because require authentication for key.

According to Zunera Jalil and Anwar M. Mirza, in their research paper titled "A Review of Digital Watermarking Techniques for Text Documents"[Jalil, Mirza(2009)], digital text watermarking can be classified in the following categories; an image based approach, a syntactic approach and a semantic approach as shown in Figure 3.5: Text Watermarking solutions [Jalil,Mirza(2009)]. All text based watermarking techniques make use of insertion and/or substitution methods for hiding secret bits by changing a single bit of a byte results in a different ASCII code that may or may not have any relevancy with the word/phrase/sentence. Our proposed solution will be based on any or both of the afore mentioned techniques.

Text watermarking methods for the English language text proposed so far; lack robustness, integrity, accuracy, and generality. Also, the amount of work done on text watermarking is very limited and specific, to date. Text watermarking algorithms using binary text image are not robust against reproduction attacks and have limited applicability. Similarly, text watermarking using text syntactic and semantic structure is not robust against attacks, with limited applicability and usability. The previous techniques are computationally expensive and non robust. Text encountering massive insertion, deletion and reordering attacks need to be protected, and efficient text watermarking algorithms are required. [Jalil,Jaffar,Mirza(2011)].

# CHAPTER 5: IMPLEMENTATION OF RSA KEY GENERATOR AND TEXT BASED WATERMARKING.

In this section, two major parts of implementation are described: The implementation of RSA Key generator and Text-based Watermarking implementation. This chapter aims to describe (a) Overview the RSA cryptosystem, RSA encryption, RSA decryption, example of decryption, breaking the RSA cryptosystem, and the current applications of RSA cryptosystem (b) RSA based PRNG ; (c) RSA Key generator algorithms description, and its modification and improvement ; and (d) Text-based watermarking, and its modification and improvement. Finally, a summary of the chapter is presented.

## 5.1 RSA cryptosystem

RSA[William Stallings,2010,] cryptosystem is widely used in both data encryption and digital signature and authentication. Three researchers at M.I.T. (Ron Rivest, Adi Shamir and Les Adleman) introduced a public key cryptosystem in 1976. Prior to this, only the first private key cryptosystems had been used.

The RSA cryptosystem is based on the modular exponentiation modulo tof he product of 2 large primes. Each individual has an encrypting key consisting of a modulus n = pq, where p & q are large primes, possibly with 200 digits each, with an exponent e that is relatively prime to (p-1)(q-1). To produce a usable key, 2 large primes must be found (this can be done quickly on a computer using probabilistic primality tests). However, the product of these primes n = pq, with approximately 400 digits, cannot be factored in a reasonable length of time[R.Rivest, A.Shmir and L.M. Adleman (1978)].

### 5.2 RSA Encryption

In the RSA encryption method, messages are translated into sequences of integers. This can be done by translating each letter into an integer, as it is done with the Caesar cipher. These integers are grouped together to form larger integers, each representing a block of letters. The encryption proceeds by transforming the integer M, representing the plaintext (the original message) to an integer C, representing the ciphertext (the encrypted message) using the function.[RSA Encryption Tutorial]

$C = M^e$ **mod** n

Below, an example is shown using RSA encryption, anf for practical reasons small primes are used p and q, rather than primes with 100 or more digits. Hence, this cipher is obviously not secured.<no plagiarism>

2 primes were selected, $p = 43$ & $q = 59$ so that $n = 43 \cdot 59 = 2537$, and with $e = 13$.

gcd (e,(p-1)(q-1) = gcd(13,42.58) = 1 (gcd = greatest common divisor)

Taking on the hypothetical message STOP, the letters will be converted into their numerical equivalents (position in the alphabet-1) and then grouped into blocks of 4.

1819 1415 = ST OP

Each block will be encrypted using the mapping:

$C = M^{13}$ **mod** 2537

Computations using modular multiplication show that $1819^{13}$ **mod** 2537 = 2081, and $1415^{13}$ **mod** 2537 = 2182. The encrypted message is thus, 2081 2182.

### 5.3 RSA Decryption

The plaintext message can be quickly recovered when the decryption key d, an inverse of e modulo (p-1)(q-1) is known. (Such an inverse exists since gcd(e,(p-1)(q-1))=1). To see this, note that if d e ≡ 1 (mod (p-1)(q-1)), there is an integer k such that d e = 1 + k(p-1)(q-1). It follows that: [[RSA Encryption Tutorial]

$C^{d} = (M^{e})^{d} = M^{de} = M^{1+k\,(p-1)(q-1)}$

By Fermat's theorem (assuming that gcd(M,p) = gcd(M,q) = 1, which holds except in rare cases, it follows that $M^{p-1} \equiv 1$ (mod p) and $M^{q-1} \equiv 1$ (mod q), consequently.

$C^{d} = M \cdot (M^{p-1)\,k\,(q-1)} \equiv M \cdot 1 \equiv M$ (mod p)

and:-

$C^{d} = M \cdot (M^{q-1)\,k\,(p-1)} \equiv M \cdot 1 \equiv M$ (mod q)

Since gcd(p,q) = 1, it follows that:-

$C^{d} \equiv M$ (mod pq)

### 5.4 Example Decryption

Using the simple cipher above the message 0981 0461 is received and the next step of decrypting it is taken. [RSA Encryption Tutorial]

n = 43 · 59 and e (exponent) = 13, we can work out that d = 937 is an inverse of 13 modulo 42 · 58 = 2436. We therefore use 937 as our decryption exponent, therefore.

$P = C^{937}$ **mod** 2537

The fast modular exponentiation (an algorithm) is used to compute $0981^{937}$ **mod** 2537, = 0704 and $0461^{937}$ **mod** 2537 = 1115. A quick translation reveals that this message was HELP.

### 5.5 Breaking the RSA Cryptosystem

There are a few possible methods of "breaking" the RSA cryptosystem. When an attacker discovers that the private key corresponds to a given public key; the attacker would then be able to read all messages encrypted and make signatures as it needs to factor the public modulus, *n*, into its two prime factors, p and q. From *p*, *q*, and *e*, the public exponent, the attacker can easily obtain access to *d*, and get the private key. The hard part is factoring n; the security of RSA depends on the hard problem of factoring the large integer *n*. It should be noted that the hardware improvements alone will not weaken the RSA cryptosystem, as long as appropriate key lengths are used.[RSA Laboratories 3.13]

Another way to break the RSA cryptosystem is to find a technique to compute *e*th roots mod *n*. Since $c = m^e$ mod *n*, the eth root of *c* mod *n* is the message *m*. The attacker can recover encrypted messages and make signatures even without finding the private key.[RSA Laboratories 3.13] This attack is not known to be equivalent to factoring. No general methods are currently known that an attempt to break the RSA cryptosystem in this way. However, in special cases where multiple related messages are encrypted with the same small exponent, it may be possible to recover the messages.

### 5.6 The Current Applications of RSA Cryptosystem

The RSA cryptosystem is currently used in secured telephone lines, on ethernet network cards, on smart cards and secure Internet communications, including S/MIME, SSL, SSH, TSL, IPSEC and S/WAN. However, the US Government Cryptography Export/Import Laws does not allow all US commercial software companies (well-known companies such as RSA security, EMC, Microsoft Cryptography, VeriSign, Check Point Software and etc)  to export Cryptography open source coding due to hacker or Cyber attack.

### 5.7 RSA based PRNG

RSA based PRNG is based on public-key cryptosystem RSA

RSA key pair generation algorithm are as below [Hani, Wen, Paniandi, 2006]:

1. Generate 2 primes, *p* and *q* randomly.

2. Calculate $M = p.q$, and $\Phi(M) = (p\text{-}1).(q\text{-}1)$.

3. Generate E that fulfills both the conditions $1< E < \Phi(\mathrm{M})$, and $gcd\ (\Phi(M)\ ,\ E) = 1$.

4. Calculate $D = E^{-1}\ mod\ \Phi(M)$.

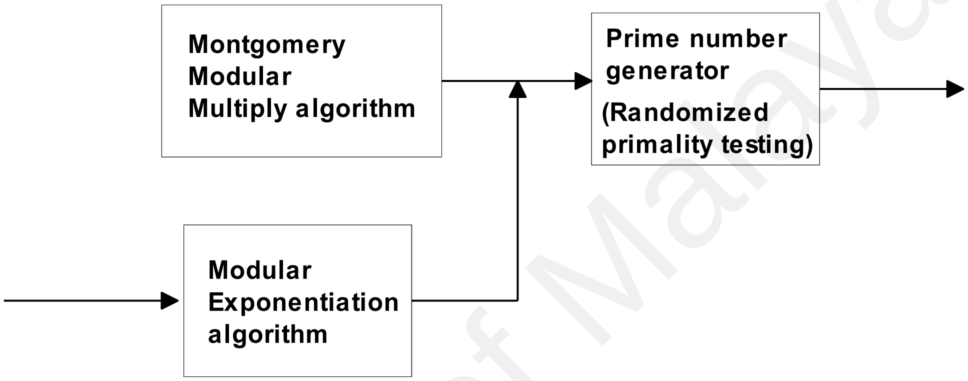

Montgomery Modular Multiply algorithm
Prime number generator (Randomized primality testing)
Modular Exponentiation algorithm

Figure 5.1: Structure of RSA Cryptosystem Algorithm

In order to implement RSA, the following is needed :(refer to Figure 5.1)

- Multiple precision arithmetic
- Pseudo Random Number Generator (PRNG)
- Prime number generator

The difficulty of implementation greatly depends on the targeted platform, application usage and the quantity of the tools needed to implement from scratch.

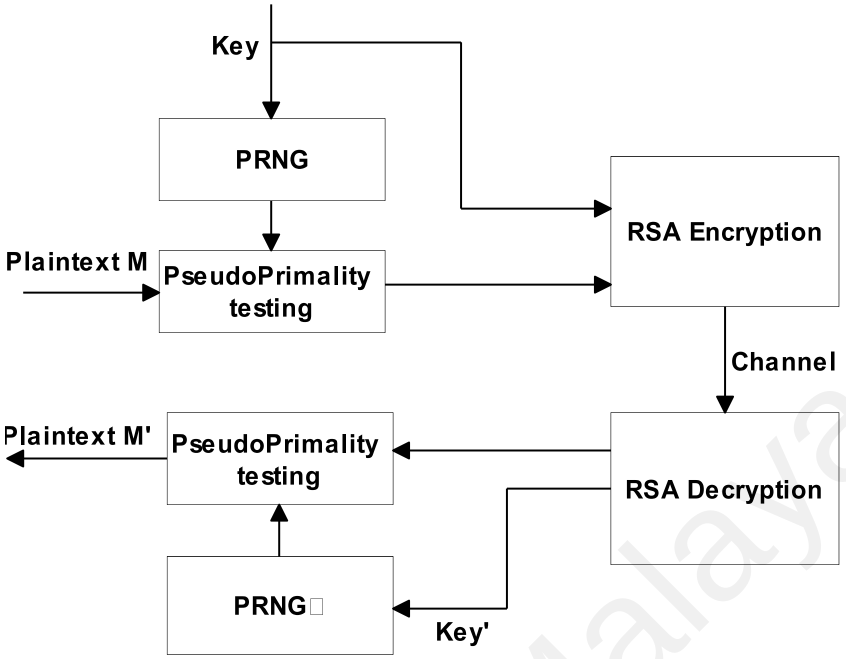


Figure 5.2: The Encryption Scheme based on RSA for PRNG [Fei Xiang, Shui-Sheng Qiu, Jie-Xin Pu, 2008]

The algorithm of the encryption is as follows:[Fei Xiang, Shui-Sheng Qiu, Jie-Xin Pu, 2008]

(1) Plaintext $M$ is encrypted by the PRNG bit sequence with *key*,

which produces $a$. Ciphertext feedback is introduce in encryption:

$x_i = x_i \oplus a_{i-1}$, $a = M \oplus x$;

(2) *key* and $a$ are encrypted by RSA with PU$\{e, n\}$, which produces ciphertext $C$:

$C = \mathrm{RSA}(a \cup key, \{e, n\})$;

(3) Ciphertext $C$ is transferred through a channel, and the receiver gets $C'$ :

$C' = C$;

(4) The receiver decrypts $C'$ by RSA with PR$\{d, n\}$ and gets $a'$ and $key'$ :

$a' \cup key' = \mathrm{RSA}^{-1}(C', \{d, n\})$;

(5) Recovered plaintext $M'$ is obtained by decrypting $a'$ using $x'$ with $key'$:

$x'_i = x'_i \oplus a'_{i-1}$, $M' = x' \oplus a'$.

Examples of PRNGs designed for cryptographic applications include MD5 Random and SHA1 Random (which respectively hold 128 bits and 160 bits of state) in the BSAFETM,3 cryptographic toolkit.[Benjamin Jun and Paul Kocher (1999)] The Intel Security Driver uses a SHA-1 mixer with 512 bits of state as illustrated in Fgure 5.3.

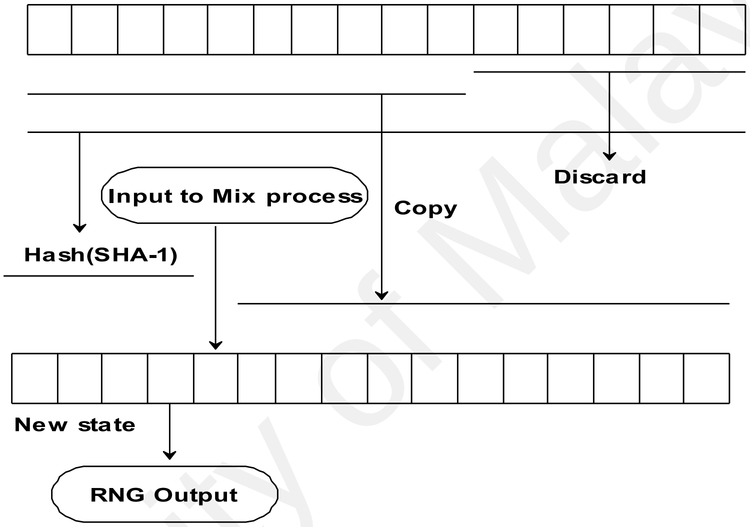


Figure 5.3: SHA-1 Mixer Design[Benjamin Jun and Paul Kocher (1999)]

## 5.8 RSA Key Generator Algorithm Description

Since this research concentrates mainly on watermarking, a simple encryption called RSA algorithm is used for encryption. After embedding the watermark into the text document and generating the watermark key, the text is encrypted using RSA encryption algorithm as described below to produce the cipher text. The reverse operation of encryption is called decryption which coverts the cipher text (the encrypted information) back to the plain text.

The RSA algorithm is named after Ron Rivest, Adi Shamir and Len Adleman, who first invented it in 1977. The RSA Algorithm is based on integer factorization assumption. The

RSA algorithm can be used for both public key encryption and digital signatures. It consists of the following procedures: key generation, encryption, decryption, and verification [Jiezhao Peng, Qi Wu. (2009)]. The proposed algorithm in this particular research , works in four phases are: the preliminary, the encryption and the extraction phase.

### 5.8.1 Preliminary Phase: *Key Generation*

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

The above mentioned procedure is shown in Figure 5.4:

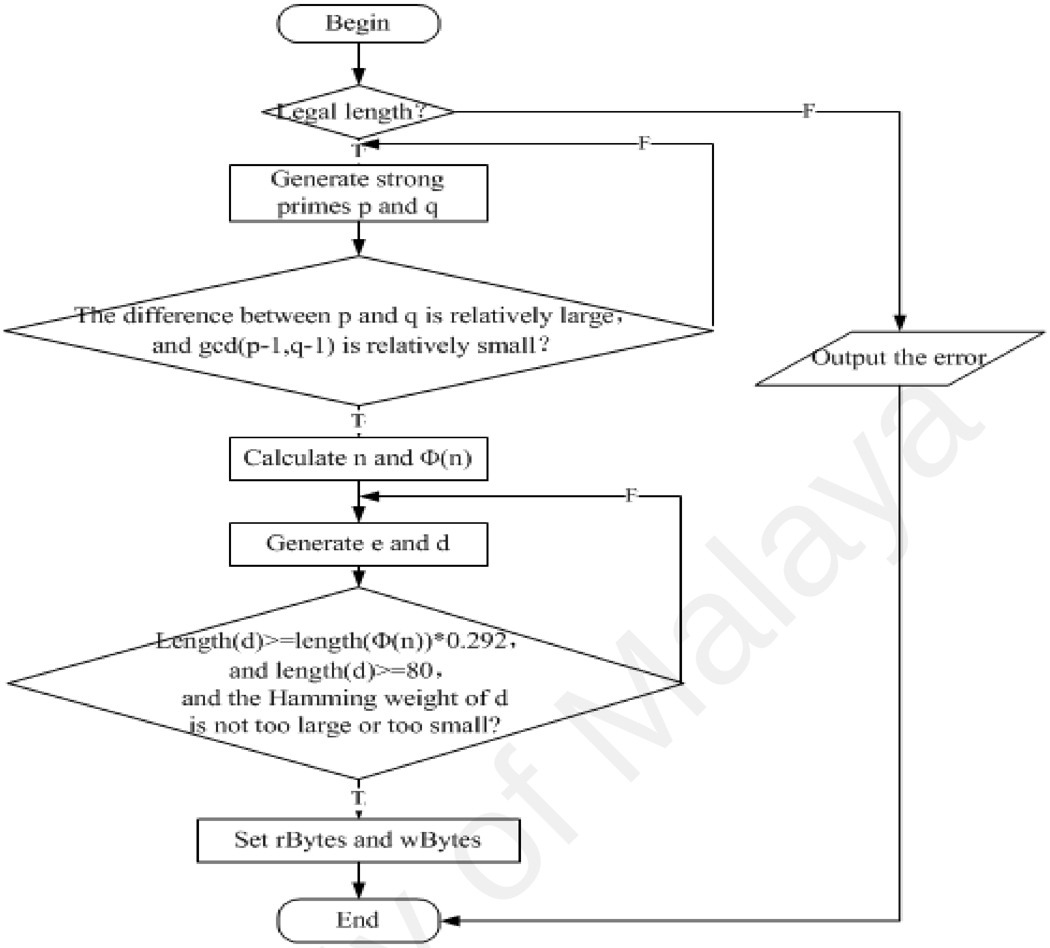

Begin
Legal length?
F
T
F
Generate strong primes p and q
The difference between p and q is relatively large, and gcd(p-1,q-1) is relatively small?
Output the error
T
Calculate n and Φ(n)
F
Generate e and d
Length(d)>=length(Φ(n))*0.292, and length(d)>=80, and the Hamming weight of d is not too large or too small?
T
Set rBytes and wBytes
End

Figure 5.4: Parameter Generation [Jiezhao Peng, Qi Wu. (2009)]

Give an example for parameters[Jiezhao Peng, Qi Wu. (2009)]:

p=4848131534065617704831441154117937596257062704169843615742581605921814337 8033

q=3947214584243518483662769569930888621088126772384588606275075191290250922697

n=1913661549759470865725258788301165799905246212975135116429079490613381754119086234952951471818129175072712173541473862206769308518920791871803187309 15001

$\Phi(n)$=19136615497594708657252587883011657999052462129751351164290794906133817541185619496537024748628094032616010708956372746686 28478

27128378295929810336614272

e=600420657137630040815032606994418133674056871928886888149768593273453456

7936900348437725917082464406498026526880563549640087746076

2851305905458633952215

d=145858088538511692944348530356221967550907794800365755916669785226001513

18234707695148915321044883143761074691747322842082780104219022567976200286

2271335

Here, p and q are strong primes which can withstand the attack by the multiplication of small primes[Zhang.(2003)]. The difference between p and q is relatively large, which can withstand the square attack. The gcd(p-1,q-1) is relatively small , thus it can avoid the iteration attack. Length(d)>=length(φ (n))*0.292 can resist a low private exponent attack(Wiener attack requires that length(d)>= length( φ (n))*0.25, while Boneh-Durfee attack requires the length(d)>=length(φ (n))*0.292). It is enough for length (d) to be bigger than 80 in order to withstand an exhaust attack, for $2^{80}$ rounds exhaust is complex enough. In order to withstand a combination attack, the Hamming weight of d should not be too large or too small. If there are too many or too few 1s in d, one could exhaust every possible position for 0 or 1, and then get the correct d. Hence, according to the bit-length of d, it should be guaranteed that the Hamming weight of d could make the number of combinations more than $2^{80}$.

5.8.2 Encryption Phase: *Encryption Procedure*

1. Partition the message m to groups $m_i$, i=1,2,…, | $m_i$ |= |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

### 5.8.3 Decryption Phase: *Decryption Procedure*

1. Partition c to $c_i$, $i=1,2,\ldots, |c_i| = |n|-1$;

2. Decrypt each $c_i$ : $mi = c_i^d \bmod n$

3. Connect each $m_i$ and recover the plain text m.

### 5.8.4 Modification and Improvement of RSA Key Generator

The modification and improvement of the proposed RSA Key Generator algorithms as described below:

1) RSA based PRNG Algorithm

The researchers believe that the RSA based PRNG is better than chaotic PRNG because Chaotic PRNG faces difficulty when combined with Text-based Watermarking. Chaotic PRNG needs to be generated using logistic maps. Moreover, this logistic mapping has serious security lacks, which makes it inadequate for cryptography applications [Arroyo 2008]. Instead of using Chaotic PRNG, the RSA based PRNG uses the RSA cryptosystem, which it is easy to multiply two large prime numbers, but difficult to factorize the product. The researchers modified the RSA key pair (refer to Figure 6.2) upon repeated testing of the “RSA Key generator testing file” for generating a new key. Then, the public key in placed in the text based watermarking pool, as shown in Figure 6.3. The RSA based PRNG is placed into data to protect it from a hacker so that it difficult for him to predict or crack its public key.

Apart from that, the researchers can also modify it by replacing the chaotic map equation as below:

$$\tilde{x}_{n+1}(x,\alpha) = \frac{1}{\alpha^2}\tan^2(N\ \arctan\sqrt{x_n})$$

(Refer to [Behnia,Akhavanb, Akhshani, Samsudin (2011)] is used to design a new PRNG algorithm. However, this research paper does not cover chaotic function theory. So this is assumption of the new method is possible area that our future research can be extended to for further studies in advance mathematics in chaotic functions theory with an improved watermarking method based on double random phase encoding technique[Behnia,Akhavanb, Akhshani, Samsudin (2011)].

According to [Shen (2009)], the RSA system which is implemented with Java language provides a lot of class library such as RSACore class including publicRSA, privateRSA etc. This system can run on all platforms and provide its application in electronic commerce. Therefore, RSA based PRNG algorithm can be used to link to function call or looping command for editing and modifying RSACore class and speed up the time of encryption and decryption.

2) Improvement of RSA based PRNG Algorithm

Due to RSA based PRNG encryption and decryption, the PRNG random bit sequence made and strengthened its the security in protecting data, by maximally modifying to 2048-bits RSA key(Common Cipher strength 128-bit keys). Although it is more secured for data protection, but the speed of encryption and decryption will be slower because it will take a longer time for processing the RSA Key.

However, the speed of digital signature verification and encryption will increase once the bit length of E is modified.

## 5.9 TEXT WATERMARKING ALGORITHM DESCRIPTION

The proposed algorithm works in two phases which are: the embedding, and the extraction phase.

5.9.1 Embedding Algorithm: Watermark Embedding

The modification and improvement of the proposed Text- based Watermarking algorithms is shown below:

1). The researchers (we) made some modifications to [Jalil,Anwar,Mirza,Iqbal (2010)], They proposed a Zero-Watermarking algorithm for authentication and copyright protection of text document. They used a combined image-plus-Text watermark method without PRNG, different from our proposed method with PRNG. The watermarking processes involve 2 steps; watermark embedding and watermarking extraction. Watermark embedding is performed by the original author and the extraction process is later done by the certifying authority on behalf of the author.

A. Watermark embedding algorithm

The embedding algorithm embeds the watermark in the text logically and generates a watermark key based on the maximum occurring first letter of double letter words in a text. The watermark embedding process is shown in Figure 5.5

1. Input W, GS, Pr and T.
2. Make partitions of T based on Pr.

```
3. Make groups of text based on GS, where Number of groups = No. of
   partitions/GS
4. Count occurrence of double letters in each group and find the second largest
   occurring double letter in each group.
5. Populate 2nd Largest Occurring Letter (OL) list for each group.
6. Merge WT and WTxt to get W.
7. Generate watermark key using steps 8, 9, and 10.
8. While (j<watermark_length) repeat step 9 to 10.
9. if (wj € MOL list)
            Key (i) =0, key (i+1) = groupnumber (2 Maximum Occurring Letter)
   else
            Key (i) =1, Key (i+1) = (wj+k) MOD 26,
            where k is in $Z_{26}$ and $Z_{26}$ represents 26 alphabets (a- z)
10. Increment i.
11. Output Key.
```

W: watermark, WTxt: text watermark, GS: Group size, Pr: Partition, T: text file, WT: text watermark

The embedding algorithm of the first text watermark (WTxt) read text file is to alphabetical watermark (WA). The preposition (Pr) and group size (GS) input by the user are used to form partitions and groups. Next, the occurrence of the first letter in each of the double letter word is counted in each group and the maximum occurring first letter (MOFL) is formed. The key generator generates the watermark key by using the watermark letters and MOFL list as stated in the embedding algorithm. This watermark key is then, registered with the certification Authority(CA) along with the text watermark, current date and time.

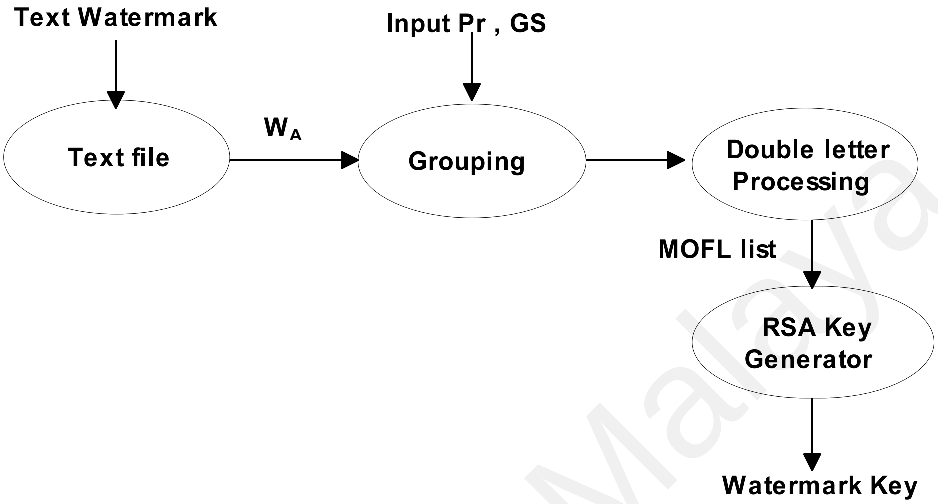


Figure 5.5: Watermark Embedding

A. Watermark Extraction Algorithm

The algorithm which extracts the watermark from the text is known as the extraction algorithm. The algorithm takes the watermark key as input and extracts the text watermark from the text. The algorithm is kept with the CA that uses it to resolve copyright conflicts, if any. The watermark extraction process is shown in Figure 5.6

1. Input Key and T.

2. Read Pr from Key and set counter = 1.

3. Make partitions of T based on Pr.

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

```
5. Count occurrence of double letters in each group and find the second largest
   occurring double letter in each group.
6. Populate 2Largest OL (Occurring Letter) list in each group.
7. Extract watermark from text using steps 8 to 11 with the help of Key.
8. L=length (Key), I=6
9. While (I<L) repeat 10 to 11
10. If (Key(I) equals 0) then
        W (I) =groupnumber(2Largest OL)
    else W(I)= Key(I+1) i.e. cipher letter.
11. Increment I by 1.
12. Obtain $W_A$
12. Output WTxt.
```

In the extraction algorithm, the text is partitioned using preposition (Pr) from the watermark key (WK). After this step, partitions are merged to form text groups based on group size. The occurrence of first letter in the double letter word in each group is analyzed and the maximum occurring first letter (MOFL) in each group is identified. The contents of watermark (WK) are then used to obtain the alphabetical watermark (WA) from the text. The alphabetical watermark is then converted to the original Text Watermark by using the watermark generator, as shown below in step 8 to 10 of the algorithm.

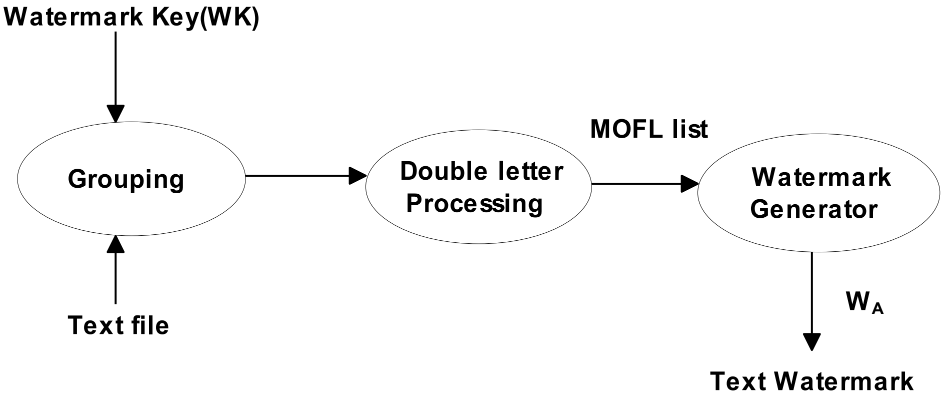


Figure 5.6: Watermark Extraction

2) Improvement of Text-based Watermarking Algorithms

The parameters retained for the experiments are as follows:

i) Group Size

To achieve better success rate, an optimum group size is required in different text categories. The possible group sizes (gpsize) are shown below:

$$\text{gpsize} \in [2,i] \text{ where } i = TL/2$$

Where, TL represents the text length measured using sentences as units. Factors of text length can also be used as group size. However, the researchers (we) used 2, 6, 12, 55, 110, 550, and 1100 as possible group sizes for ( our) experiment. The researchers' (our) experiment does not use mass groups of size with a maximum of sentences less than 2000.

$$\text{gpsize} \in \text{factor}(TL)$$

Where, TL represents the text length and factors is a function which gives factors of TL.

ii) Semantic Approach Text- based Watermark

Experiments show that the semantic approach on text based watermark can improve the information hiding capacity due to the modifying of meaning of sentences. It is possible to speed or fasten the search of the word tree by using the semantic structures of a text to embed the watermark.

iii) Accuracy percentage

The accuracy percentage represents how accurately watermark is retrieved from the watermark embedded text. It depends on a text length and watermark length. It is represented below:

$$\text{Accuracy} = WL / TL * 100\ \%$$

Where, TL represents a text length measured using sentence as a unit, and WL indicates the watermark length.

For that reason, the above three parameters' value will affect the improvement performance of text based watermarking (refer to table 6.3 - table 6.6). Therefore, our experiment and proposed text based watermarking method shows better performance when compared to previous approaches in term of capacity, robustness and security (refer to Table 6.6).

## 5.10 CHAPTER SUMMARY

As referred to 3.7.3, the semantic watermarking techniques are used for the proposed implementation of text based watermarking. 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. was the first to propose the semantic watermarking schemes in 2000 [Atallah,McDonough,Nirenburg,Raskin(2000)(2002)]. Later, the synonym substitution method was proposed in which the watermark is embedded by replacing certain words with their synonyms without changing the context of text [Topkara,Atallah(2006)]. Xingming, et al. proposed noun-verb based technique for text watermarking [Sun, Asiimwe,(2005)]. Figure 3.8 shows the parse tree for noun-verb based transformation. The advantages of the semantic method in text based watermarking techniques is that it a) analyzes text meaning and performs transformation. b) improves the information hiding capacity of English text by modifying the granularity of meaning of individual term/sentence. c) speeds the search of the word tree faster by 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.

However, thesemantic approach has some weakness due to the sensitivity of legal documents. The reason is that quotations and poetry could not make random semantic transformations. In this chapter, the implementation of RSA key watermarking and text based watermarking is described, including in the modification and improvement of RSA based PRNG algorithm and Text based Watermarking algorithms in detail. Digital text documents being an important medium of information exchange requires secure data protection.

# CHAPTER 6. EXPERIMENTAL RESULTS AND DISCUSSIONS

## 6.1 Simulation Results

The following sequence of steps identifies the RSA Key Generator adapted in this work:.

1. Identifying and defining the problems.
2. An algorithm 1**: Multiple-precision Library ([1]igInt.js)** is a suite of routines for performing multiple-precision arithmetic in JavaScript.
3. An algorithm 2: **Modular reduction Library ([2]Barrett.js)** is a class for performing Barrett modular reduction computations in JavaScript.
4. An algorithm 3: **RSA Library ([3]RSA.js)** is a suite of routines for performing RSA public-key computations in JavaScript. It links algorithm 1 and algorithm 2 into algorithm 3 in 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 reimplementation of the Multiple-Precision Library and the Miller-Rabin Test for primality. Keys are generated using Algorithm 8.1 from the Handbook of Applied Cryptography [Menezes, Oorschot, Vanstone (1996)][Schneier (1996)].

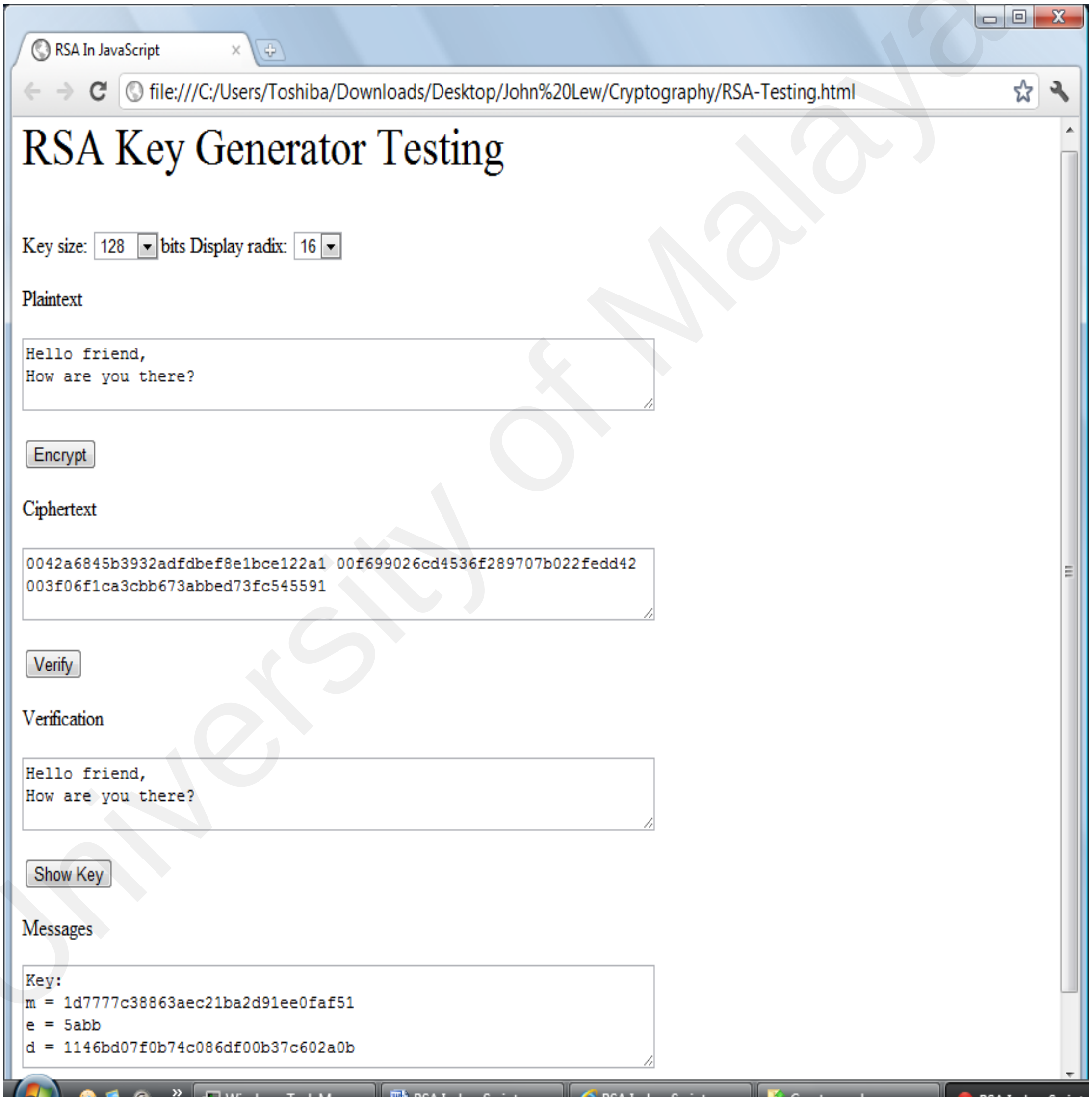
The following procedures are taken for compiling and testing the RSA Key Generator:

1. Before testing the"RSA Key Generator Testing File", the researcher made sure that all files (BigInt.js,Barrett.js, RSA.js) are included in the same directory.

[1] Refer to www.ohdave.com/rsa/BigInt.js
[2] Refer to www.ohdave.com/rsa/Barrett.js
[3] Refer to www.ohdave.com/RSA.js

2. The “RSA-testing” file is opened [Appendix B - RSA-Testing.htm] using the Internet Explorer browser (such as Microsoft IE9, Google Chrome, Firefox).

3. The file is tested accordingly as indicated below:


RSA In JavaScript
file:///C:/Users/Toshiba/Downloads/Desktop/John%20Lew/Cryptography/RSA-Testing.html
Encrypt
Verify
Show Key

RSA Key Generator Testing

Key size: 128 bits Display radix: 16

Plaintext

```
Hello friend,
How are you there?
```

Ciphertext

```
0042a6845b3932adfdbef8e1bce122a1 00f699026cd4536f289707b022fedd42
003f06f1ca3cbb673abbed73fc545591
```

Verification

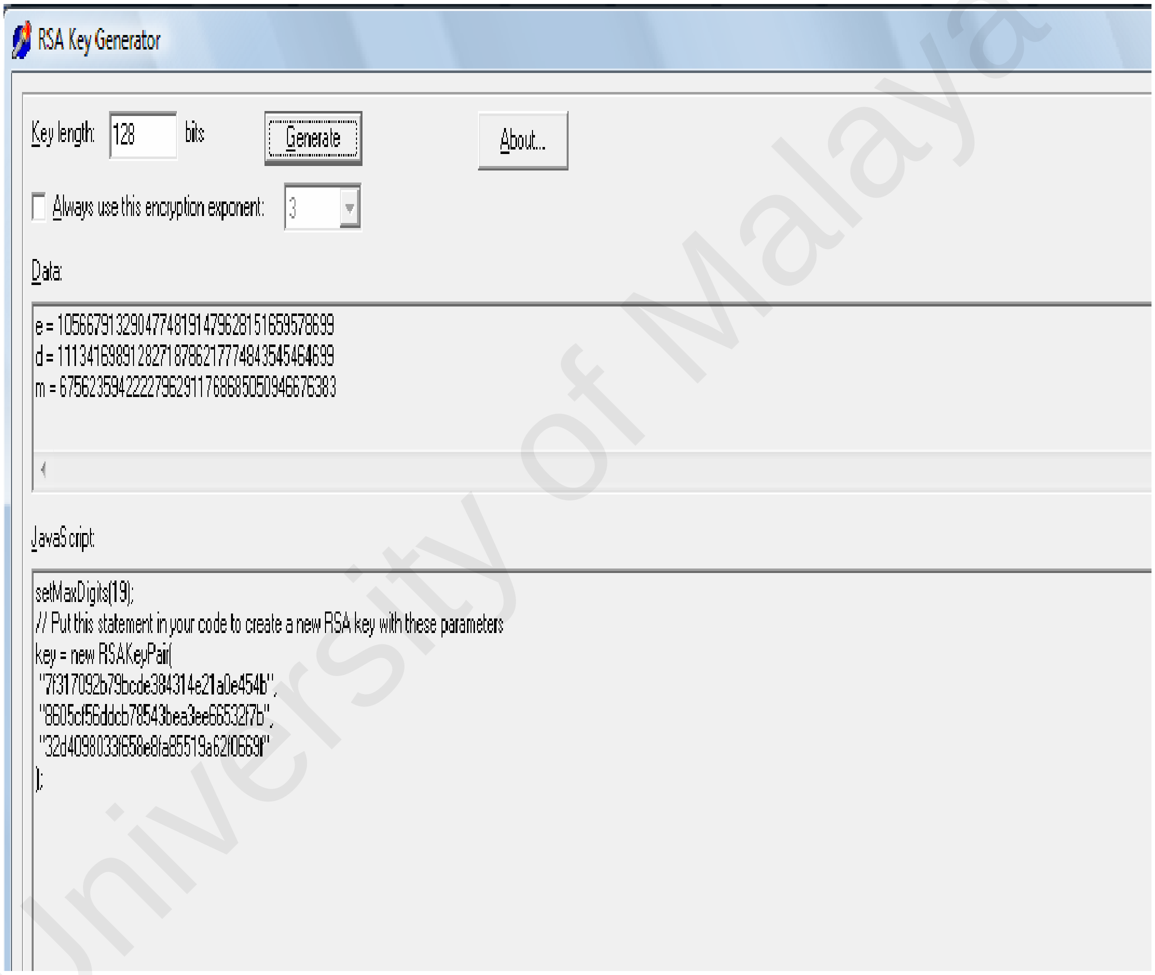
```
Hello friend,
How are you there?
```

Messages

```
Key:
m = 1d7777c38863aec21ba2d91ee0faf51
e = 5abb
d = 1146bd07f0b74c086df00b37c602a0b
```

Figure 6.1: RSA Key Generator Testing

4. Borland Delphi 7 or 8 Edition is used for compiling the RSA Key Generator source code. All source code files are ran and compiled (Appendix A - Main.pas, BigMath.pas, RSA.pas) before it executing the RSA Key Generator (main program).
5. After running the "RSA Key Generator" execution file, the file is tested as indicated below:



Key length: 128 bits

□ Always use this encryption exponent: 3

Data:

```
e = 10566791329047748191479628151659578699
d = 11134169891282718786217774843545464699
m = 67562359422227962911768685050946676383
```

JavaScript:

```
setMaxDigits(19);
// Put this statement in your code to create a new RSA key with these parameters
key = new RSAKeyPair(
 "7f317092b79bcde384314e21a0e454b",
 "8605cf56ddcb78543bea3ee66532f7b",
 "32d4098033f658e8fa85519a62f0669f"
);
```

Figure 6.2: RSA Key Generator Testing Generating for New Key

6. In order to secure the RSA Key Generator Testing (Figure 6.1), the researcher edited this source code by using the Notepad or HTML editor. The following JavaScript lines are added as below:

```
setMaxDigits(19);
// Put this statement in your code to create a new RSA key with these parameters
key = new RSAKeyPair(
"3",
"18e78c00696284b3923a6d7401aee5b",
"255b52009e13c7108d7b57f0a764527"
 );
```

7. The test is repeated  on the “RSA Key Generator Testing File” until completion.

The above mentioned is the implementation of the RSA Key Generator in the PRNG’s application. On the other hand, according to “A Novel Dynamic Model of Pseudo Random Number Generator” [Behnia,Akhavanb, Akhshani, Samsudin (2011)] author,in his paper chaotic map equation.(Refer to below equation) is used to design a new PRNG algorithm.

$$\tilde{x}_{n+1}(x,\alpha) = \frac{1}{\alpha^{2}} \tan^{2} (N \arctan \sqrt{x_{n}})$$

[Behnia,Akhavanb, Akhshani, Samsudin (2011)]

Our future research can be extended from the above for further studies in advance mathematics in Chaotic Functions Theory, an improved watermarking method based on double random phase encoding technique [Behnia,Akhavanb, Akhshani, Samsudin (2011)} can be utilised or used.

Furthermore, according to the “Robust Hash Functions for Digital Watermarking”[Fridrich, Goljan (2000)] author , she explained  that the generation of the digital watermark from a pseudo-random Gaussian sequence. The PRNG then generates a pseudo-random sequence $\xi^{(i)}$ of a desired length (determined by the particular watermarking technique).

$$\xi^{(i)} = PRNG(K \oplus B \oplus i \oplus b_{\pi_1(i)} \oplus b_{\pi_2(i)} \oplus \ldots \oplus b_{\pi_q(i)}$$ [Fridrich, Goljan (2000)]

In the expression above the symbol $\oplus$ denotes concatenation. The final Gaussian sequence $\eta \in N(0,1)$ is obtained by summing up $\xi^{(i)}$ for all i and normalizing:

$$\eta = \sqrt{{}^{3}\!/_{N} \sum_{i=1}^{N} \xi^{(i)}}$$ [Fridrich, Goljan (2000)]

Our future research can extend the above for further studies on double hash functions for text based digital watermarking.

## 6.2 Further Modification and Improvement

In this section, futher modification and improvement is listed below:

1) The modification and improvement of PRNG based on an 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 the performance of the RSA Key Generator Testing, a new RSA public key and private key was generated. A new RSA public key and private key is changed always, repeatedly after running and conducting the RSA Key Generator Testing.

ii) Due to the speed performance constraint, a maximum to 2048-bits RSA key is modified (Common Cipher strength 128-bit keys), so that it can be a more secured data protection.

An elaborate explanation is given in Section 5.8.4 on the modification and improvement of our proposed RSA Key Generator Algorithms.

iii) Due to the RSA based PRNG encryption and decryption, the PRNG random bit sequence strengthened the security of protecting data. However, this research paper does not cover the Chaotic Function Theory. So, this is the assumption of the other new method that can be used to design a new PRNG algorithm.

2) The Modification and Improvement of Text-Based Watermarking which is based on an open source Watermarking algorithms is identified below:

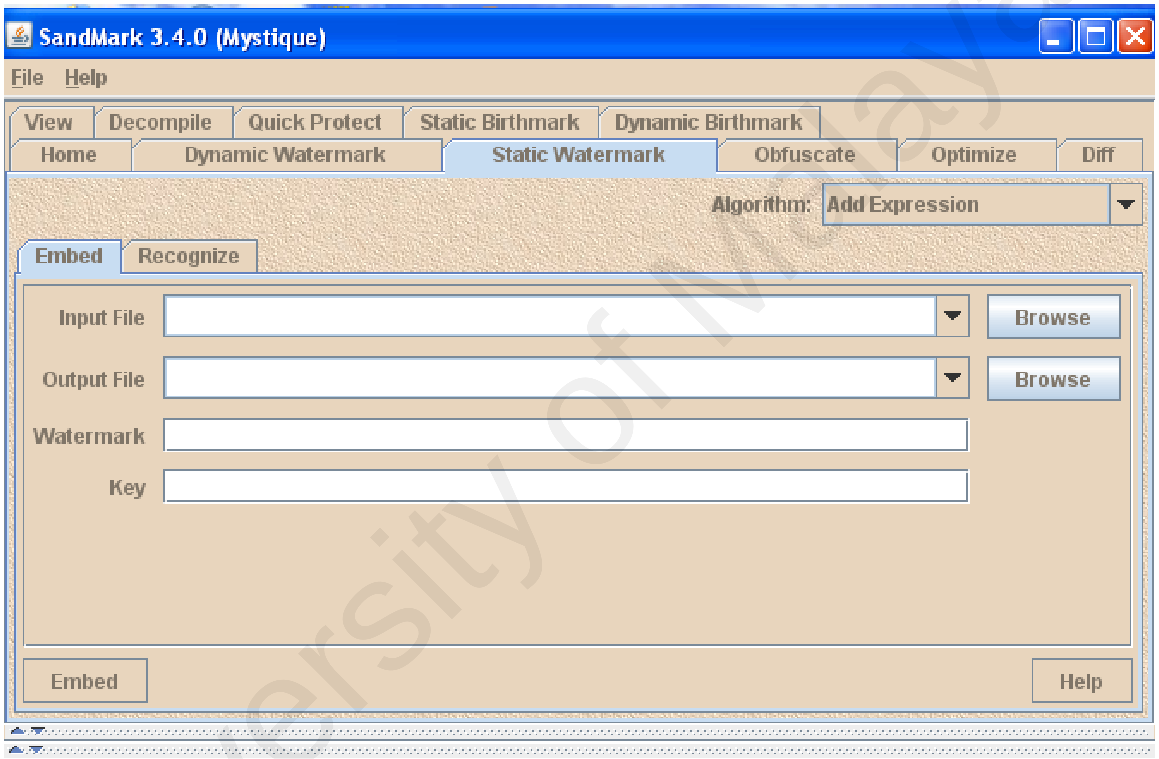
i) Semantic techniques are used for Text Based Watermarking techniques. The semantic watermarking schemes focus on using the semantic structure of texts to embed the watermark, for a faster speed and accuracy in the embedded text contents.

ii) We edited the open source watermarking algorithms and added up the RSA based PRNG algorithms; embedded with Text Based Watermarking algorithms. Apart from the researchers, no other individual knew that it was a combined PRNG random bit sequence embedded with Text Watermarking Algorithms. Further in depth explanation is described in Section 5.9.1 on the Modification and improvement of our proposed Text-based Watermarking Algorithms.

iii)After compiling the RSA Key Generator Testing, the Public Key and Private Key are generated . Then, the Public Key is placed in the Text Based Watermarking pool as identified in Figure 6.3. Its invisibility and erased hidden data, due to the PRNG random bit sequence causes a hacker to face difficulty in predicting or cracking its Public Key.

## 6.3 Implementation Result

This paper developed an experimental tool with JAVA on windows according to the above embedding and extracting algorithm for watermarking. Its interface is shown below in Figure 6.3.



Algorithm: Add Expression

Figure 6.3: Digital Watermarking Tool

The performance of the algorithms under both Text with/without PRNG is evaluated for means of random bit sequence for random insertion, deletion and reordering of word characters to and from the text. The average accuracy of the extracted watermark under text with/without PRNG is shown in Table 6.2 and Figure 6.4.

Table 6.1: Text Categories [Jalil, Jaffar,Mirza.(2011)]

| | 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 the dataset designed and used in [Reuter Corpus]. Each text samples are divided into four categories based on the text length, which is identified in Table 6.1[Jalil, Jaffar, Mirza.(2011)]. Each text category contains 5 samples. These samples have been randomly selected from Reuter's Corpus, e-books, newspapers, articles, research papers, novels, web contents, reports, and so forth.

Table 6.2: Accuracy of Extracted Watermark Under Text With/Without PRNG.

| Text category | Accuracy of Extracted Watermark | |
|---|---|---|
| | Image-plus-Text without PRNG [Jalil,Mirza(2010)] | 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 technique is used as the text based watermarking technique. The semantic watermarking scheme focuses on using the semantic structures of texts to embed the watermark, for faster speed and accuracy, in the text contents, like verbs, nouns, prepositions, words spelling, acronyms, sentence structure, and grammatical rules. Due to the RSA based PRNG, textual watermarking is more secured and sensitive to tampering attacks compared to texts with PRNG. In Section 6, Sub-Section 6.2, we mentioned further modification and improvement of PRNG based on an open source RSA algorithm and Text-Based Watermarking that affected the improvement. This is shown in Table 6.2.

The most important difference is that Jalil and Mirza[Jalil,Mirza (2010)] measured the performance using Image-plus-Text-based Digital Watermarking without PRNG. Comparatively, we measured the performance using Text-Based Digital Watermarking with PRNG.

The results are given in Table 6.2 which shows better accuracy of the extracted watermark due to Text-Based Watermarking combined with PRNG. The accuracy of the extracted watermark under text with PRNG exceeded 97%. Textual watermarking is more secured and sensitive to tampering attacks than text with PRNG. Hence, the accuracy of texts without PRNG is lesser than texts with PRNG. Experiments were also performed on text with/without PRNG on all text samples and the percentage's accuracy of extracted watermark or text with/without PRNG is shown in Table 6.2 and Figure 6.4.

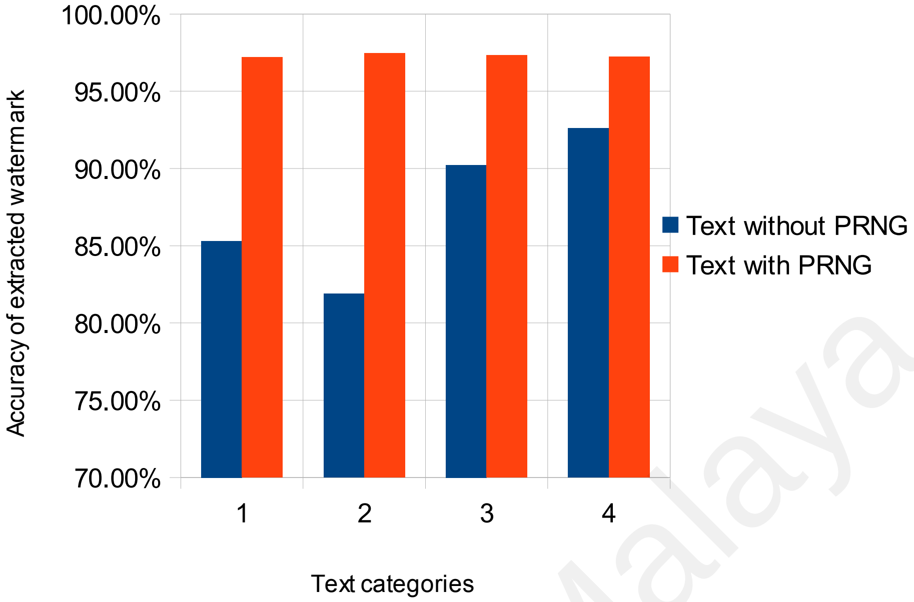


Figure 6.4, the accuracy of the extracted watermark under Text with/without PRNG

The accuracy of extracted watermark exceeds 85%. Textual watermarking is more sensitive to tampering attacks than text with PRNG. Hence, the accuracy of texts without PRNG is lesser than texts with PRNG. However, the combined accuracy is about 97%. Textual watermarking is more secured and sensitive to tampering attacks than texts with PRNG. Hence, the accuracy of texts without PRNG is lesser than texts with PRNG. Our performance results show better accuracy of the extracted watermark due to Text Based Watermarking combined with PRNG. Experiments were also performed on text with/without PRNG on all text samples and the percentages of accuracy of the extracted watermark for texts with/without PRNG is shown in Table 6.2 and Figure 6.4.

### 6.4. Performance Analysis

Table 6.3: Performances of Proposed System

| | Mi-Young Kim [Kim(2008)] | Our system |
|---|---|---|
| The number of Sentences | 2080 | 2820 |
| Average number of words/sentences | 15.67 | 18.25 |
| Sentences selected for embedding watermarking bit string(%) | 75.00% | 85.00% |
| Unsuitable sentences among marked sentences | 29.29% | 25.65% |
| Unsuitable sentences among non-transformed sentences | 14.81% | 13.25% |

As shown in Table 6.3, the researchers (we) used 2820 sentences with the average number of words or sentences being 18.25. The sentences selected for embedding watermarking bit are 85% higher compared to Mi-Young Kim's [Kim(2008)] because of its fast speed of embedding watermarking bit string. The percentage rate of unsuitable sentences among marked sentences is 25.65% and non-transformed sentences rate is 13.25%

Table 6.4: Comparison of Performance for Coverage and Information Hiding Capacity

| | Y.L Chiang[Chiang(2004)] | Mi-Young Kim[Kim(2008)] | Our system |
|---|---|---|---|
| Coverage about the sentences selected for embedding watermarking bit (%) | 6.70% | 75.00% | 85.00% |
| Relation of information-hiding capacity | 01:06.60 | 01:14.05 | 01:05.50 |

In Table 6.4, the total coverage of sentences selected for embedding watermark bit is 85% which is higher than the coverage of Mi-Young Kim's[Kim(2008)]. The relation of information-hiding capacity is 1:5.5 which means that for every 5.5 terms of text, one bit of watermark can be a hidden data. It is higher than that of Mi-Young Kim's [Kim(2008)] and Y.L.Chiang's [Chiang(2004)].

Table 6.5: Comparison of Performance for Average Edit

| | H.M.Meral[Meral(2007)] | Mi-Young Kim[Kim(2008)] | Our system |
|---|---|---|---|
| Average edit(%) | 55.90% | 25.67% | 22.50% |

In Table 6.5, the average edit rate is 22.5% which shows a better result than that of H.M.Meral's [Meral(2007)] and Mi-Young Kim's[Kim(2008)] due to the lesser time in the average edit.

Table 6.6: Comparison Between Our Texts Based Watermarking Method and Previous Approaches in Term of Capacity, Robustness and Security.

| Approach | Speed | Robustness | Imperceptibility |
|---|---|---|---|
| Our proposed in this work<br>(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 robustness | Slight visible |
| Syntactic Techniques | Moderate speed | Not robustness | Slight visible |

In Table 6.6, our proposed work also considered the general comparisons (speed, robustness and imperceptibility) with the different text-based watermarking techniques that were previously explained in Section 2.1-2.3. Based on our observation, the semantics technique is considered as more suitable technique compared to the other two approach techniques because of high speed performance, robust, and invisibility for hidden data.

**Comparative Results**

Table 6.7: Summary Comparison of PRNG Generator With Proposed RSA Key Generator

|  | /dev/random | Yarrow | BBS | HAVEGE | AES | Proposed RSA Key Generator |
|---|---|---|---|---|---|---|
| Cryptographic Primitives used | SHA-1 or MD5 | SHA-1 3DES | None | None | AES | RSA based PRNG |
| Strength | 1024 bits | 160 bits | $\leq \log_2 \Phi$ (N) | Size of table + thousands of volatile states | Counter mode: 256 bits. PRNG | PRNG mode: 2048 bits |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | mode: 128 bits | |
| Statistical Test Result available | DIEHARD battery | None | NIST test suite | NIST test suite, DIEHARD battery | Overlapping serial test, gambling test. | Passed NIST test suite( refer to table 6.10) |
| Speed | 8-12K bits/s | No result | 3 bits/s | HAVEG: 8k-16K bits/s<br>HAVEG: 280 Mbits/s | 426 Mbits/s | 3.5 –5 bits/s |
| Portability | Part of Lunix Kernel | Yes | Yes | Yes | Yes | Yes |
| Pros<br>• security<br>• Algorithm<br>• Key length | Secure<br>Complex<br>Large | Medium<br>Complex<br>Large | High<br>Complex<br>Large | High<br>Complex<br>Large | High<br>Complex<br>Large | Very High<br>Much Complex<br>Large enough |
| Cons<br>• Feasible<br>• Quality assurance | Slight<br>Applicable | Medium<br>Slight | Medium<br>Medium | Slight<br>Medium | Slight<br>Medium | Yes<br>Applicable |
| Complexity<br>• Space<br>• Time<br>• Design<br>• Searching | 0(n)<br>0(n)<br>-<br>- | 0(n)<br>0(n)<br>-<br>- | 0(n)<br>0(n)<br>-<br>- | 0(n)<br>0(n)<br>-<br>- | 0(n)<br>0(n)<br>-<br>- | 0(n)<br>0(m)<br>-<br>- |

There are different complexities of PRNG that enlisted in Table 6.7. It is important to consider that the space generated an n-cell Pseudo-Random bit sequence is equal to ‘$2^n$’. Therefore, the space consumed for the RSA based PRNG, is ‘O(n)’. Time required m-length for the RSA key generator is [($2^n = 2^m * (2^{n-m})$ for $n \geq 1$ and m=1,2,3…(n-1)], is ‘O(m)’. where ‘m’ < ‘n’. Hence, the advantages and cost effectiveness of the RSA based PRNG are reported in Table 6.7.

As seen in Table 6.7, the proposed RSA Key Generator using the RSA based PRNG with a strong strength PRNG mode of 2048 bits key size, also passes all NIST tests as reported in Table 6.8, with a speed of about 3.5-5 bits/second. It is shown that our proposed RSA key generator has better security performance compared to other PRNG generators.

NIST SP 800-22 is a standard published by U.S. National Institute of Standards and Technology and include a suite of 16 tests to determine the quality of PRNGs. In the final revised edition (NIST S.P. 800-22, 2010), the Lempel-Ziv test was dropped, remaining only 15 randomness statistical tests: More detail about test descriptions and mathematical computations refer to “*A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications”*. NIST Special Publication. 800-22, Revision 1a.[NIST 2010]**.** Table 6.8 illustrates the step by step process that is followed in the evaluation of a single binary sequence.

**Table 6.8: Evaluation Procedure For A Single Binary Sequence**

| Step By Step Process | Comments |
|---|---|
| 1. State your null hypothesis. | Assume that the binary sequence is random. |
| 2. Compute a sequence test statistic. | Testing is carried out at the bit level. |
| 3. Compute the P-value. | P-value $\varepsilon$ [0, 1]. |
| 4. Compare the P-value to $\alpha$. | Fix $\alpha$, where $\alpha$ $\varepsilon$ (0.001, 0.01]. *Success* is declared whenever P-value $\geq$ $\alpha$; otherwise, *failure* is declared. |

Table 6.9 describes the general characteristics of each of the statistical tests.

**Table 6.9: Characteristics of the NIST Statistical Tests**

| No. | NIST Statistical Test & Purpose | Defect Detected |
|---|---|---|
| 1 | ***Frequency Test :-*** The purpose of this test is to calculate the proportion of the number of "1" and "0" within the given sequence. For a proportion of approximately ½ from the total number of the elements for the two values, the sequence is considered random. | Too many zeroes or ones. |
| 2 | ***Frequency Test within a Block*** *:-*This test is realized to compute the proportion of "1" within *M*-bit block sequences. For a frequency of *M*/2 for the ones in an *M*-bit block, the sequence is considered random. The $\chi_2$ distribution is considered for statistical test | The longest run of ones within M-bit blocks. |
| 3 | ***Runs Test*** :- This test is focused on computing the total number of runs of uninterrupted identical bits. A *k* length run of ones is bounded before and after with zeros. It also determines the oscillation between "1" and "0" values and the total number of runs, $V_{n(\text{obs})}$, for both zeros and ones. | Large (small) total number of runs indicates that the oscillation[1] in the bit stream is too fast (too slow). |
| 4 | ***Test of the Longest Run of ones in a Block*** *:-* This test is focused on the longest run of "1" values in a *M*-bits block. Its purpose is to calculate if the longest run of ones from *ε* is consistent with the length of the one expected in a random sequence. | Deviation of the distribution of long runs of ones. |
| 5 | ***Binary Matrix Rank Test*** *:-* This test calculates the rank of sub-matrices of *ε*. Linear dependence among fixed length substrings of *ε* is checked. Here $\chi_2$(obs) is a measure of how ranks of various orders matches the expected values for randomness characteristics. | Deviation of the rank distribution from a corresponding random sequence, due to periodicity[2] |
| 6 | ***Discrete Fourier Transform Test*** *:-*This test seeks to detect repetitive patterns, using DFT (Discrete Fourier Transform) applied to the *ε* sequence. It detects the deviations from the assumption of randomness for bit patterns. *d* represents the normalized difference between expected and observed number of frequency components over 95% thresholds. | Periodic features in the bit stream. |
| 7 | ***Non-overlapping Template Matching Test****:-*This test aims to detect generators that provide aperiodic patterns. An *m*-bit length sequence is used to determine *m*-bit patterns. *B* is a string of length *m.* | Too many occurrences of non-periodic templates. |
| 8 | ***Overlapping Template Matching Test :-*** This test counts | Too many occurrences of m-bit |

[1] *Oscillation* refers to abrupt changes between runs of zeroes or runs of ones

[2] *Periodicity* refers to sub-sequences that repeat.

| | | |
|---|---|---|
| | the number of occurrences of previously specified strings. As in the non-overlapping template matching test it uses an *m*-bit window in order to find an *m*-bit pattern. It aims to detect generators that provide aperiodic patterns. An *m*-bit length sequence is used to determine *m*-bit patterns. *B* is a string of length *m*. | runs of ones. |
| 9 | ***Universal Test* (*Maurer's Statistical Test*) :-** This test detects if a sequence can be significantly compressed without information alteration. A non-random sequence can be significantly compressed. For *L* being the length of each block, $f_n$ represents the sum of $\log_2$ distances between *L*-bit templates. | Compressibility[3](regularity). |
| 10 | ***Linear Complexity Test:-***This test focuses on the length of an LFSR (Linear Feedback Shift Register). A short LFSR implies non-randomness. Here, $\chi_2$(obs) is a measure of how the observed number of occurrences of LFSRs fixed length matches an expected number. | Deviation from the distribution of the linear complexity[4] for finite length (sub)strings. |
| 11 | ***Serial Test :-*** The serial test determines the number of occurrences of $2_m$ *m*-bit patterns and their chance to appear across the entire sequence $\varepsilon$ .<br>For $m = 1$, this test is similar to frequency test. | Non-uniform distribution of m-length words. Similar to Approximate Entropy. |
| 12 | ***Aproximate Entropy Test :-*** The approximate entropy test compares the frequency of two consecutive blocks with lengths *m* and $m + 1$, against an expected value for a random sequence $\varepsilon$ . | Non-uniform distribution of m-length words. Small values of ApEn(m) imply strong regularity. |
| 13 | ***Cumulative Sums Test :-***This test is focused on the maximal excursion of the random walk defined by the cumulative sum of modified digits in the sequence and determines if they are too small or too large for random sequences . | Too many zeroes or ones at the beginning of the sequence. |
| 14 | ***Random Excursion Test:-***The random excursion test is focused on the number of cycles with *K* visits in a cumulative random walk and determines the number of visits to a state in a cycle . | Deviation from the distribution of the number of visits of a random walk[5] to a certain state. |
| 15 | ***Random Excursion Variant Test :-*** This test focuses on the total number of times that a state occurs in a random walk and detects the deviations from an expected number of occurrences. It contains eighteen subtests with individual conclusions. | Deviation from the distribution of the total number of visits (*across many random walks*) to a certain state. |

[3] *Compressibility* refers to the existence of a sub-sequence that represents the entire sequence.

[4] *Linear complexity* is the length of the shortest linear feedback shift register that generates the sequence.

[5] A 1-D *random walk* is a sequence of steps, each of whose characteristics is determined by chance.

Table 6.10: Result of NIST Test Suite on the Proposed RSA based PRNG

| **No** | **NIST Statistical Test** | **P-Value** | **Proportion (success ratio)** | **Pass/Fail** |
|---|---|---|---|---|
| 1 | The Frequency (Monobit) Test | 0.5712 | 95/100 | Pass |
| 2 | Frequency Test within a Block | 0.5813 | 96/100 | Pass |
| 3 | The Runs Test | 0.6235 | 92/100 | Pass |
| 4 | Tests for the Longest-Run-of-Ones in a Block | 0.7516 | 96/100 | Pass |
| 5 | The Binary Matrix Rank Test | 0.7833 | 97/100 | Pass |
| 6 | The Discrete Fourier Transform (Spectral) Test | 0.8812 | 98/100 | Pass |
| 7 | The Non-overlapping Template Matching Test | 0.8367 | 98/100 | Pass |
| 8 | The Overlapping Template Matching Test | 0.9561 | 99/100 | Pass |
| 9 | Maurer's "Universal Statistical" Test | 0.9374 | 95/100 | Pass |
| 10 | The Linear Complexity Test | 0.9877 | 99/100 | Pass |
| 11 | The Serial Test | 0.5321 | 91/100 | Pass |
| 12 | The Approximate Entropy Test | 0.5751 | 92/100 | Pass |
| 13 | The Cumulative Sums (Cusums) Test | 0.5853 | 92/100 | Pass |
| 14 | The Random Excursions Test | 0.6371 | 93/100 | Pass |
| 15 | The Random Excursions Variant Test | 0.7813 | 96/100 | Pass |

A statistical test is formulated to test a specific null hypothesis (H0). For the purpose of this document, the null hypothesis under test is that the sequence being tested is random.Each test is based on a calculated test statistic value, whichis a function of the data. If the test statistic value is $S$ and the critical value is $t$, then the Type I error probability is $P(S > t \,||\, Ho$ is true$)$ = $P($reject $Ho \,|0$ is true$)$, and the Type II error probability is $P(S \leq t \,||\, H0$ is false$)$ = $P($accept $H0 \,|\, H0$ is false$)$. The test statistic is used to calculate a *P-value* that summarizes the strength of the evidence against the null hypothesis. For these tests, each *P-value* is the probability that a perfect random number generator would have produced a sequence less random than the sequence that was tested, given the kind of non-randomness assessed by the test. If a *P-value* for a test is determined to be equal to 1, then the sequence appears to have perfect randomness. A *P-value* of zero indicates that the sequence appears to be completely non-random. A significance level (α) can be chosen for the tests. If *P-value* ≥ α, then the null hypothesis is accepted; i.e., the sequence appears to be random. If *P-value* < α*,* then the null hypothesis is rejected; i.e., the sequence appears to be non-random. The parameter α denotes the probability of the Type I error. Typically, α is chosen in the range [0.001, 0.01]. The minimum pass rate for each statistical test with the exception of the random excursion (variant) test is approximately = 96 for a sample size = 100 binary sequences. Therefore, we can conclude that the data file sequence is random.

As Table 6.10, the proposed RSA based PRNG have passed all NIST Statistical Tests as stated above. The table shows that our proposed RSA based PRNG has the better security performance, excellent pseudo randomness, and its security shows that it can be utilized to generate pseudorandom bit sequences.compared to other PRNG generators.

Table 6.11: Comparison of Text Based Watermarking and Our Proposed Work.

| **Proposed Scheme** | **Text watermark language** | **Text based Watermarking techniques** | **Data hiding speed** | **Visibility** |
|---|---|---|---|---|
| [Chiang 2004] | Chinese language | Semantic substitution | Low speed | Slightly visible |
| [Meral 2007] | Turkish language | Syntax based natural language | Low speed | Slightly visible |
| [Gutub 2008] | Arabic language | New "kashida" method | Low speed | Invisible in code |
| Kim[2008] | Korean language | Syntactic Analysis | Medium speed | Slightly visible |
| Jalil,Mirza(2010) | English language | Image-plus-text watermark: | Medium speed | Slightly visible |
| | | | | Invisible and |

| | | | | |
|---|---|---|---|---|
| Our proposed work | English language | Semantic approach | High speed | combined PRNG with text based watermark |

We compared our proposed work with other text based watermarking works. It concludes that our proposed work has faster speed of data hiding and is invisible compared to others. (Refer to the Table 6.11)

## CHAPTER 7. CONCLUSION

### 7.1 Objectives Achieved

This research paper have achieved the following an objectives as below:

- To develop new systems based on PRNG combined with Text based Watermarking techniques.(refer to figure 6.1, figure 6.2, figure 6.3) We had developed the key generator as illustrated in Fig.6.1.Fig.6.2 which shows a random key generated by the PRNG. In Fig.6.3, user interface of the watermarking tool is presented.
- To investigate our implementation, results that show that better accuracy of extracted watermark(97% as refer to table 6.2 and figure 6.4) and PRNG random bit sequence made it strengthen the security of protecting data. Every time the RSA key Generator Testing is compiled (refer to figure 6.2), it generated a new Public key and Private key. Then, the public key is placed in text based watermarking pool as shown in figure 6.3. Invisibility and erased hidden data due to PRNG random bit sequence. It causes makes a hacker hard to have difficulity to predict or crack its public key. So these random key generated by the PRNG has good random value.

- To apply an Open source PRNG Key generator(PRNG algorithms) with combined with text based watermarking techniques in terms of efficiency for performance speed, robustness and security(refer to table 6.6).
- To evaluate implementation results for text based watermarking upon completing PRNG Key generator testing output results. In Table 6.7, our results are compared with PRNG generator with proposed RSA key generator. Our proposed RSA key generator using RSA based PRNG with strong strength PRNG mode 2048 bits key size, also passes all NIST tests as refer to table 6.10 with a speed of about 3.5-5 bits/second. It is shown that our proposed RSA key generator has better security performance that of the other compared PRNG generator.

Based on the comparative analysis result, the researchers (we) combined PRNG with Text-based Watermarking to produce a double security on data security so that it becomes difficult to crack this data information. As mentioned in the above section on the achieved objectives, this hybrid scheme fulfils copyright protection, digital content documents protection and content authentication.Besides that, according to [Manmeet Kaur, Kamna Mahajan (2015)], the authors cited the (our) proposed method by stating that "RSA Key was generated and made public in text watermarking. This technique can be used in the information hiding area using cloud computing". On the other hand, according to [Awrad Mohammed Ali, Neslisah Torosdagli, Josiah Wong (2016)], the authors mentioned about "the use of one-way functions allows pseudorandom generators to create long random-like strings from short truly random inputs so that secure keys can be shorter than the messages they encrypt. Such pseudorandom stretches are provably secure against polynomial-time adversaries due to proofs of unpredictability for each bit of the stretch." According to [Mina Mishra, V.H. Mankar

(2015)], the authors proposed that “PRNG encryption algorithm have shown strength against linear, differential and statistical attacks. Modified RNG possess good plaintext and key sensitivity property.” According to Jean-Christophe Zapalowicz (2015) in his doctoral thesis, he mentions on the security of the pseudorandom number generators and implementations of RSA signature schemes protects it from being attacked by a hacker.

## 7.2 Contributions

To our knowledge, this thesis makes the following contributions:

My contribution work does in Cryptography area as follows:

- A better understanding of Pseudo Random number primitives will make it easier to design and use PRNGs securely. We developed new system based on PRNG combined with text based digital watermarking techniques. Thus, this thesis is intended to provide a reference finding for newcomer security designers and to promote more activities in these security issues related to combined PRNG with text based digital watermarking techniques for Cryptographic application..
- The Modification and improvement of PRNG based on Open source RSA Algorithms. We modified a maximum to 2048-bits RSA key size(Common Cipher strength 128-bit keys). So it will be a more secured data protection  Also, every time RSA key Generator Testing was compiled(refer to  figure 6.2), it generated a new Public key and Private key. Then public key is placed in text based watermarking pool as shown in Figure 6.3, with invisibility and erase hidden data due to PRNG random bit sequence.

My contribution work in Watermarking area is listed as below:

- Better understanding of concepts and methods of Text based Watermarking combined with PRNG for Cryptography application. Our thesis helps researchers to understand our implementation results which shows that better accuracy of extract watermark and PRNG random bit sequence strengthens the security of protecting data. Thus, this thesis is intended to provide a reference finding for newcomer security designers and to promote more activities in these security issues related to combined PRNG with text based digital watermarking techniques for Cryptographic application..

- The Modification and improvement of Text based Watermarking based on Open source Watermarking Algorithms. We used semantic techniques for Text based Watermarking techniques. The semantic watermarking schemes focused on using the semantic structure of text to embed the watermark, with faster speed and accuracy embedded the text contents. We edited Open Source Watermarking Algorithms, by adding up RSA based PRNG algorithms embedded with text based watermarking algorithms. So nobody knows that it is combined PRNG random bit sequence embedded with text watermarking algorithms

- Every time the RSA key Generator Testing is compiled, it generate a new Public key and Private key. Then, the public key is placed in text based watermarking pool as shown in figure 6.3. Due to PRNG random bit sequence, it makes invisibility and erases hidden data. Also it makes a hacker to have difficulity time to predict or crack its public key due to the combination of open source RSA algorithms and watermarking algorithms.

This thesis is the first attempt to combine the PRNG with Text Based Watermarking. This attempt was taken due to the potential advantages of the combination which produces a more secured method for encrypting and hiding text data in digital media.

This research paper proposed the method of using the combination of PRNG with Text Based Watermarking for Cryptography application. The reason why this research paper developed this new method is similar to the aims and objectives of this thesis paper.(Refer to Section 1.2) The list of advantages found compared to the old methods  are also mentioned in the Table 6.7 , which depicts the summarised comparison of  the PRNG Generator with the proposed RSA Key Generator. Additionally, Table 6.11 presented the comparison chart of the text based watermarking with the proposed work.

Due to cost implementation compared to other PRNG methods, and among different PRNG Generators in Table 6.7 is was decided that the proposed RSA Key Generator is the best option. Thus, the proposed methodology is good for generating a random sequence in solving actual world digital copyright problems such as bitcoin, online transaction, cloud data security.

## 7.3 Future Research

This thesis proposed a Text based Watermarking algorithm based on Pseudo-Random Number Generator(PRNG) for cryptography application. The thesis reported the experimental results that shows that the 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 thesis has large information embedding capacity, which can be used for secret communication of confidential information.

The work of this thesis will given for further research work direction as the following.

- Our future research can be extended for further studies in advance mathematics in Chaotic Functions Theory combined with Text Based Watermarking.[Behnia,Akhavanb, Akhshani, Samsudin (2011)].
- Further improvement in the watermarking method based on the double random phase encoding technique[Behnia,Akhavanb, Akhshani, Samsudin (2011)] that combines with PRNG.
- Our future research can extend further studies for double hash function for text based digital watermarking[Fridrich, Goljan (2000)] that combined with PRNG.
- Our future research can be extended for further studies in combining secure Multi-Stage Pseudo-Random Number Generator with Digital Watermarking[Abdulameer K. Husain, Osama Al-Haj Hassa, Adi A. Maaita, Hamza A. A. Al_Sewadi (2015)]

# References

A.O.L. Atkin and F.Morain,(1988). Elliptic Curves and Primality Proving, *Mathematics of Computation.* 29-68.

A Shamir. (1979). How to Share a Secret. *Communications of ACM.* 612-613.

A. Renyi. (1970). Probability Theory, Amsterdam: North-Holland.

A. Lemma, S. Katzenbeisser, M. Celik, and M. van der Veen. *Secure watermark embedding through partial encryption*.(2006). Paper presented at the Proceedings of the 5th International Workshop on Digital Watermarking (IWDW '06). Jeju Island, Korea.

A.Fiat, A.Shamir. (1987). *How to prove yourself: Practical solutions to identification and signature problems*: Springer-Verlag.

A.Menezes, P.C.Van Oorshot, S.A. Vantone. (1996). *Handbook of Applied Cryptography.* New York, CRC-Press.

Alfred W.Aresenault and Sean Turner. (2000). Internet X.509 Public Key Infrastructure Roadmap.

Andrea Rock (2005). Pseudorandom Number Generators for Cryptographic Applications. PhD Thesis.

Awrad Mohammed Ali, Neslisah Torosdagli, Josiah Wong (2016). Security at Its Finest: Overview of Cryptography Mechanisms. *ACM Transactions on Embedded Computing Systems*, Vol.5, A1-A10.

B.A.Wichmanna, I.D. Hill.(2006).Generating good pseudo-random numbers, *Computational Statistics & Data Analysis*, 1614 – 1622

Bellare, Mihir; Rogaway. Phillip.(2005). Introduction. Introduction to Modern Cryptography. (21 September 2005). pp. 10.

Benjamin Jun and Paul Kocher.(1999). *Analysis of the Intel True Random Number Generator*, Cryptography Research.

Bruce Schneier.(1996). *Applied Cryptography: Protocol, Algorithms and Source Code in C,* Second edition, John Wiley& Sons, 3th edition.

Barker E.B. (2000). *A Statistical suite for random and pseudorandom number generators for cryptography application*, ITL Bullentin.

Blake, G. Seroussi, N. Smart.(1999). *Elliptic Curve in Cryptography*. Cambridge University Press.

B.M.Macq, J.J.Quisquate (1995). Cryptography for Digital TV Broadcasting. Proc. Of the IEEE, 83(6), 944-957.

Bruce Schneier. (1996). *Applied Cryptography*. John Wiley & Sons.

Baker, W.1(991). I*ntroduction to the Analysis of the Data Encryption Standard(DES*). Loguna Hills, CA:Aegean Park Press.

B.Schneier. (1995). *Email Security (with PGP and PEM).* New York City: John Wisely & Son.

Belmouhoub, M. Djema, J.P. Barbot. (2005). Observability quadratic normal forms for discrete-time systems. *IEEE Trans. on Automatic Control,* 50(7), 1031–1038.

Bharti Kashyap, K. J. Satao.(2015). A Review on Multi-Biometric Cryptosystem for Information Security. I*nternational Journal of Advanced Research in Computer and Communication Engineering*, 346-351.

Blum L., Blum, M, and Shub, M. (1986). *A simple unpredictable pseudo-random number generator*," *SIAM J. Comput.,* 15(2), 364-83.

Chaitin,G.J. (1975). Randomness and mathematical proof. *Scientific American*, 232, 47-52

Charles Wright (2004), So You Need a Random Number Generator, Emporia State University, research projects. Retrieved August, 2013, from http://brannigan.emporia.edu/projects/hardwareRNG/TRNG/Wright.pdf

Chia-Mu Yu and Chun-Shien Lu.(2005). Robust Non-Interactive Zero-Knowledge Watermarking Scheme Against Cheating Prover. Paper presented at the Proceedings of ACM Portal MM-SEC'05, New York, New York, USA.

Chichester. (2003). *Cryptography and public key infrastructure on the internet*, West Sussex, England: Wiley.

Compagner. (1991). A:Definition of randomness, *America Journal of Physic*,59,700-705.

C.E. Shannon. (1948). A Mathematical theory of communication. *Bell System Journal*, 27:379-423, 623-656.

C.E. Shannon. (1949). Communication theory of secrecy systems. *Bell System Journal*, 28:656-715.

C.P. Schnorr. (1990). Efficient Identification and Signature for Smart Cards. *Advances in Cryptography*: Springer-Verlag.

C.H. Benett, G.Brassard and A.K. Ekert. (1992). Quantum Cryptography. *Scientific American*, 26-33.

Carl M.Ellision, Bill Frantz, Butler Lampson, Ron Rivest, Brian M.Thomas and Tatu Ylonen. (1998). *SPKI Certificate Theory*: Internet Draft.

C.F.Gauss. Disquisitiones Arithmeticae, 1966. Yale University Press, New Haven, English edition.

Charles P.Pfleeger. (1997). *Security in Computing*: Prentice Hall PTR.

Chen, Jiageng; Miyaji, Atsuko; Su, Chunhua.(2014). *Distributed Pseudo-Random Number Generation and Its Application to Cloud Database*. Lecture Notes in Computer Science, 8434: 373-387. Springer-Verlag

D.R. Stinson.(1995) *Cryptography – Theory and Practice*. New York: CRC-Press.

D.Eastlake, S.Crocker and J.Schiller. (1994), *RFC 1750: Randomness Recommendation for Security:* Internet Activies Board.

Dilip Kumar Sharma, Vinay Kumar Pathak and G. P. Sahu. (2007). Digital watermarking for secure E-Government framework. Computer Society Of India;pp182-191.

Dorothy Elizabeth Robling Denning. (1983). *Cryptographic and Data Security*: Addision-Wesley.

David Chaum,A.Fiat and M.Naor.(1992). *Untraceable electronic cash. In Advance in Cryptology.* Lecture Notes in Computer Science 576: Springer-Verlag, 324-337.

D.Eastlake, S.Crocker and J.Schiller. (1994). *RFC 1750: Randomness Recommendation for Security:* Internet Activies Board.

Davis, D., Ihaka, R., Fenstermacher. (1994). *Cryptographic Randomness from Air Turbulence in Disk Drives,* LNCS, Vol. 839: Springer.

D. Kundur and K. Karthik.(2004). Video fingerprinting and encryption principles for digital rights management. *Proceedings of the IEEE*, 92(6), 918–932.

D.Arroyo,G.Alvarez, and V.Fernandez. (2008). On the inadequacy of the logistic map for cryptography applications. *Comite Organizador de la X RECSI,* 1:77-82.

Ellison, Carl. (1998).Cryptographic Random Numbers. Draft P1363 Appendix E. Retrieved August 2013,from http://www.clark.net/pub/cme/P1363/ranno.html .

E. Rieffel and W.Polak. (1998). An Introduction to Quantum Computing for Non-Physicists. Retrieved August 2013, from xxx.lanl.gov/abs/quant-ph/9809016.

ECRYPT II Yearly Report on Algorithms and Keysizes. Retrieved August 2013, from, http://www.ecrypt.eu.org/documents/D.SPA.13.pdf.

Ellison, C., Cryptographic Random Numbers. Retrieved August 2013, from http://www.clark.net/pub/cme/P1363/ranno.html.

Eastlake, Crocker, and Schiller. (1994). *RFC 1750: Randomness Recommendations for Security*. IETF Network Working Group..

Federal Information Processing Standards Publication 180-1. (1995). *Secure Hash Standard.* U.S. Department of Commerce/NIST, Springfield, VA: NTIS.

F.J. Mac Williams and N.J.A. Sloane. (1977). The Theory of Error-Correcting Codes: North-Holland.

Fei Xiang, Shui-Sheng Qiu, Jie-Xin Pu. (2008). A New Pseudo-Random Number Generator with Application in RSA, Paper presented at the ICCS 2008, 11th IEEE Singapore International Conference on , Communication Systems.

G.H. Hardy, E.N.Wright.(1979). *An Introduction to the theory of Numbers*, 5nd edition,Oxford University Press.

G.S.Verman.(1926).Cipher printing telegraphy systems for secret wire and radio telegraphic communications, *Journal of American Institute for Electrical Engineer*, 45;109-115.

G.S. Vernam: Secret signaling systems, U.S. Patent #1,310,719,1919.

Goldreich,O,1999:Modern Cryptography, Probabilistic Proofs and Pseudorandomness. :Springer Verlag.

Goldwasser, S.; Micali, S.; Rackoff, C. (1989). "The knowledge complexity of interactive proof systems". SIAM Journal on Computing (Philadelphia: Society for Industrial and Applied Mathematics) 18 (1): 186–208.

Graham Shaw (2000). Digital Document Integrity. Paper presented at the ACM Multimedia Workshop Marina Del Rey CA USA.

Han Delf, Helmut Knebl, (2002). *Introduction to Cryptographic: Principles and Applications,* Springer-Verleg Berlin Heidelberg.

H.E.ROSE.(1994). *A Course in Number Theory.* 2rd edition, Oxford:Clarendon Press.

H.Riesel. (1994). *Prime Numbers and Computer Method for Factorization*. Boston,Basel: Birkhauser.

H.W. Lenstra, Jr. (1987). Factoring Integers with Elliptic Curves. *Annals of Mathematics,* 649-673.

H.Bauer.(1996). *Probability Theory*, Berlin: de Gruyter.

H.Gordon. (1997). *Discrete Probabilty*, Heidelberg, New York: Springer-Verlag.

Haraid Niederreiter,(2002). *Coding Theory and Cryptology,* Singapore University Press and World Scientific Publisher Co. Pte. Ltd.

Horowitz, P., Hill, W. (1980). *The Art of Electronics*, Cambridge University Press.

Hamed khataeimaragheh and Hassan Rashid. (2010).A novel watermarking scheme for detecting and recovering distortions in database tables. *International Journal of Database Management Systems ( IJDMS ).,* 2(3).

Hassan M. Meral et al. (2009). Natural language watermarking via morphosyntactic alterations. *Computer Speech and Language*, 23, 107-125.

Hasan M. Meral, et al. , Syntactic tools for text watermarking. (2007). Paper presented at the Proceedings of 19th SPIE Electronic Imaging Conf. 6505: Security, Steganography, and Watermarking of Multimedia Contents.

H.M.Meral, E.Sevinc, E.Unkar, B.Sankur,A.S.Ozsoy, T.Gungor. (2007). Syntactic tools for Text Watermarking. Paper presented at the Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents,.

I. J. Cox, M. L. Miller, and J. A. Bloom. (2002). *Digital* Watermarking: Morgan Kaufmann Publishers.

I.J. Cox, M.L. Miller, J.A. Bloom, J. Fridrich, T. Kalker (2008). *Digital Watermarking and steganography,* Seattle, Washington, USA., Morgan Kaufmann publishers.

I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon. (1997). Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, 6(12), 1673–1687.

I. Niven, H.S.Zuckerman and H.L. Montgomery. (1991). *An Introduction to the theory of Number*, New York, 5th. Edition, John Wiley.

I. Belmouhoub, M. Djema, J.P. Barbot.(2005). Observability quadratic normal forms for discrete-time systems, *IEEE Trans. on Automatic Control,* 50(7), 1031–1038.

John Kelsey, Bruce Schneier, David Wagner, Chris Hall. (1998) *Cryptanalytic Attacks on Pseudorandom Number Generators*, Paper presented at the Proceedings of 5th International Workshop Proceedings, FSE 1998, Paris, France.

J Menezes, PC van Oorschot, and SA Vanstone.(1996), *Handbook of Applied Cryptography*: CRC Press

Jiezhao Peng, Qi Wu. (2009), *Research and Implementation of RSA Algorithm in Java*, Paper presented at the Proceedings of International Conference on Management of e-Commerce and e-Government.

John Viega,Gary McGraw. (2001). *Building Secure Software: How to avoid Security Problems the right way:* Addison-wesley.

J.H Van Lint.(1992). *Introduction to Coding Theory*: Springer-Verlag.

Jozef Gruska. (1999). *Quantum Computing*: Mc-Graw-Hill Publisher(UK).

J.Linn (1987). *Privacy Enhancement for Internet Electronic Mail: Part 1 – message Encipherment and Authentication Procedures*: RFC 989.

J.Linn. (1988). *Privacy Enhancement for Internet Electronic Mail: Part 1 – message Encipherment and Authentication Procedures*: RFC 1040..

Jennewein, T., Achleitner U., Weihs G., Weinfurter H., and Zeilinger A.(2000). A Fast and Compact Quantum Random Number Generator, *Review of Scientific Instruments*, 71, 1675-1680.

J. T. Brassil, S. Low, and N. F. Maxemchuk. (1999). *Copyright Protection for the Electronic Distribution of Text Documents*. Paper presented at the Proceedings of Proceedings of the IEEE, 87(7), 1181-1196.

J. Crowcroft, C. Perkins, and I. Brown, *A method and apparatus for generating multiple watermarked copies of an information signal*, WO Patent No. 00/56059, 2000.

J. J. Eggers, J. K. Su, and B. Girod. (2000). P*ublic key watermarking by eigenvectors of linear transforms*, Paper presented at the Proceedings of the European Signal Processing Conference (EUSIPCO ’00), Tampere, Finland.

J. J. Eggers, J. K. Su, and B. Girod. (2000). *Asymmetric watermarking schemes*, Paper presented at the Proceedings of the Sicherheit in Mediendaten,GMD Jahrestagung, Berlin, Germany.

John Kelsey , Bruce Schneier , David Wagner. (1998). *Chris Hall, Cryptanalytic Attacks on Pseudorandom Number Generators*, Fast Software Encryption, Paper presented at the Proceedings Fifth International Proceedings.

J. L. Massey. (1998). *An Introduction to Contemporary Cryptology*, Paper presented at the Proceedings of the IEEE, 1988, 76(5), 533-549.

Jean-Christophe Zapalowicz (2015). *Security of the pseudorandom number generators and implementations of public key signature schemes*. Cryptography and Security. Doctoral thesis, Universite Rennes 1, 2015.

Jaseena K.U. , Anita John. (2011). Text Watermarking using Combined Image and Text for Authentication and Protection, *International Journal of Computer Applications* (0975 – 8887), 20(4).

Knuth,D.E. (1997).*The Art Computer Programming:* Addsion-Wesley, 3rd.edition.

K.Ireland, M.I Rosen. (1982). *A Classical Introduction to Modern Number Theory,* Berlin,Heidelberg, New York, Springer-Verlag.

K.H.Rosen.(2000). *Elementary Number Theory and its Applications*, 4th edition,Reading,MA:Addision-Wesley.

Kazimerz Alster, Jerzy Urbanowecz, Hugh C.William.(2001). *Public-Key cryptography and computational Number Theory.* Paper presented at the Proceedings of International Conference stefn Banach International Mathematical, Center Warsaw, Poland.

Khan Farhan Rafat , Muhammad Sher. (2009). *Digital Steganography for ASCII Text Documents.* Paper presented at the Proceedings of the 7th International Conference on Frontiers of Information Technology, *FIT'09*, CIIT, Abbottabad, Pakistan.

L'Ecuyer. (2004). *Random number generation, Draft for a chapter of the forthcoming Handbook of Computation Statistics,J.E Gentle, W.Haerdle and Y.Mor*i: Springer-Verlag.

Luby,M. (1996). *Pseudorandomness and Cryptographic Applications*: Princeton Computer Notes.

Landau, S. (2004). *Polynomials in the Nation's Service: Using Algebra to Design the Advance Encryption Standard*: American Mathematical Monthly.

Lingfeng Liu, Suoxia Miao, Hanping Hu, Yashuang Deng (2015). Pseudorandom bit generator based on nonstationary logistic maps. *IET Information Security,*1-8.

Liu Yongliang , Wen Gao (2004). Secure Watermark Verification Scheme. Paper presented at the Proceedings of the IEEE International Conference on Multimedia and Expo (ICME).

Manpreet Kaur, Sonika Jindal, Sunny Behal. (2012), A Study Of Digital Image Watermarking. *International Journal of Research in Engineering & Applied Sciences*, 2(2).

Maurer,U. (1992) A Universal Statistical Test for Random Bit Generators, *Journal of Cryptology*, 5, 89-105.

Marsaglia, G.. (1984). A current View of Random Number Generators, Keynote Address, *Computer Science and Statistics: 16th Symposium on the Interface*, Atlanta.

M. Khalil, A. Z. Shameri, and W. S. Chong, (2000). *Pipeline Implementation of Secure Hash Algorithm SHA-1 for Cryptographic application in Network Security.* Paper presented at the Proceedings of in National Conference on Telecommunication Technology, Johor Bahru, Malaysia.

Mohamed Khalil Hani, Hau Yuan Wen, Arul Paniandi. (2006). Design and implementation of a private and public key crypto processor for next-generation it security applications. *Malaysian Journal of Computer Science*, 19(1), 2006.

Matsumoto, M., and Nishimura, T., 1998, Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Trans. Model. Comput. Simul. 8*,(1), 3-30.

N.Robbin. (1993). *Beginning Number Theory*: Wm C.Brown, Dubuque, Iowa.

N.Koblitz. (1987). Elliptic Curve Cryptography. *Mathematics of Computation*, 203-209.

Miroslav Knezevic, Frederik Vercauteren, Ingrid Verbauwhede. (2010). Faster Interleaved Modular Multiplication Based on Barrett and Montgomery Reduction Methods, *IEEE TRANSACTIONS ON COMPUTERS*, 59(12).

M.Mignotte. (1983). *How to share a Secret. Cryptography,* Lecture notes in Computer Science, Springer-Verlag, 371-375.

M. J. Atallah, C. McDonough, S. Nirenburg, and V. Raskin. (2000). *Natural Language Processing for Information Assurance and Security: An Overview and Implementations.* Paper presented at the Proceedings of 9th ACM/SIGSAC New Security Paradigms Workshop, Cork, Ireland.

M. J. Atallah, V. Raskin, M. C. Crogan, C. F. Hempelmann, F. Kerschbaum, D. Mohamed, and S.Naik. (2001). *Natural language watermarking: Design, analysis, and a proof-of-concept implementation.* Paper presented at the Proceedings of the Fourth Information HidingWorkshop, LNCS 2137, Pittsburgh, PA.

M. Atallah, V. Raskin, C. F. Hempelmann, M. Karahan, R. Sion, U. Topkara, and K. E. Triezenberg. (2002). *Natural Language Watermarking and Tamperproofing.* Paper presented at the Proceedings of Fifth Information Hiding Workshop, *LNCS* 2578, Noordwijkerhout, The Netherlands.

M. Topkara, C. M. Taskiran, and E. Delp. (2005). *Natural language watermarking*. Paper presented at the Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents VII.

M.A. Jafarizadeh, S. Behnia.(2002). Hierarchy of chaotic maps with an invariant measure and their compositions, *J. Non. Math. Phys,*. 9, 26–41.

M. Barni and F. Bartolini.(2004). Data hiding for fighting piracy, *IEEE Signal Processing Magazine*, 21( 2), 28–39.

Mi-Young Kim. (2008). Paper presented at the Proceedings of the 12th WSEAS International Conference on Computer.

Manmeet Kaur, Kamna Mahajan. (2015). An Existential Review on Text Watermarking Techniques. *International Journal of Computer Applications,* 29-32.

Mina Mishra, V.H. Mankar. (2015). Text Encryption Algorithms based on Pseudo Random Number Generator. *International Journal of Computer Applications*, 111, 1-6.

M Djema, J.P. Barbot, I. Belmouhoub. (2009). Discrete time normal form for left invertibility problem, *European Journal of Control* ,15(2), 194–204.

NIST (2010). Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. NIST Special Publ. 800-22, Revision 1a, April 2010

O.Foster. (1996). *Algorithmische Zahlentheorie, Braunschweigh, Wiesbaden*: Vieweg.

Oded Goldreich. (2001). *Foundations of Cryptography: Basic Tools*, Cambridge University Press.

Parijat Naik (2002), Analysis of Random Number Generators, research project. Retrieved September,2013,from http://cs.ucsb.edu/~koc/ec/project/2002/naikpa.pdf

Peter Shor. (1994). *Algorithm for Quantum Computation: discrete logarithms and factoring*. Paper presented at the Proceedings of the Thirty-Fifth Annual Symposium on the Foundation of Computer Science.

Pradeep Kaur , Pankaj Bhambri (2015), To Design an Algorithm for Text Watermarking, *The SIJ Transactions on Computer Science Engineering & its Applications (CSEA)*, 3(5), 62-67.

Qingkuan Dong, Wenxiu Ding, Lili Wei. (2014). Improvement and optimized implementation of cryptoGPS protocol for low-cost radio-frequency identification authentication. *Security and Communication Networks*. *8*, 1474–1484.

Ravi Sharma (2012), Digital Watermarking Article, paper presentation.Retrieved September,2013,from http://www.slideshare.net/ravi33s/watermark-12641562 .

Robert Davies. (2000), Hardware random number generators, Statistics Research Associates Limited.. Retrieved Semptember 2013, from http://www.robertnz.net/hwrng.htm

R.B.Ash. (1995). *Information Theory*, New York: John Wiley & Son.

R.G.Gallager. (1969). I*nformation Theory and Reliable Communication*: Wiley.

R.W. Hamming.(1995). *Coding and Information Theory*, 2nd edition, Englewood Cliff,NJ, Prentice Hall.

Richard P. Feynman. (1995). *Feynman Lecturers on Computation*: Addison-Wesley.

R.Rivest, A.Shmir and L.M. Adleman. (1978). A method for obtaining digital signature and public key cryptosystems. *Communication of the ACM*, 21(2), 120-126.

Ron Rivest and Butler Lampson. (1996). S*DSI – a simple Distributed Security Infrastructure*: Technical report.

Raymond Hill. (1986). *A First Course in Coding Theory*, Oxford University Press.

Ronald L.Rivest, Cryptogaphy, In.Van Leeuwen. (1990). *Handbook of Theoretical Computer Science*, chapter 13, 718.

R.A Gangoli, D.Ylvsaker. (1967). *Discrete Probability.* New York, Harcourt,Brace & World

R. Parviainen and P. Parnes. (2001). *Large scale distributed watermarking of multicast media through encryption.* Paper presented at the Proceedings of the IFIP TC6/TC11 International Conference on Communications and Multimedia Security Issues of the New Century, Darmstadt, Germany.

RSA Encryption – Tutorial. Retrieved October 2013, from http://www.woodmann.com/crackz/Tutorials/Rsa.htm

RSA Laboratories 3.13. Retrieved October 2013, from http://www.rsa.com/rsalabs/node.asp?id=2216

Reuter Corpus [Online]. Retrieved October 2013, from http://about.reuters.com/researchandstandards/corpus/index.asp

R.Chandramouli, Nasir Memon, Majid Rabbani.(2000). Digital Watermarking. Retrieved November 2013, from http://www.vis.uky.%7Echeung/courses/ee639_fall04/readings/intro_watermark.pdf

S. P. Mohanty.(1999). Digital Watermarking: A Tutorial Review. Report, Dept. of Electrical Engineering, Indian Institute of Science, Bangalore, India.Retrieved November,from http://www.cs.unt.edu/~smohanty/research/Reports/MohantyWatermarkingSurvey1999.pdf

SSH Communications Security (2004).Retrieved November, from http://www.ssh.fi/support/cryptography/algorithms/random.html

Scott Durrant (1999), *Random Numbers in Data Security Systems*: Intel Corporation.

S.Brands. (1994). *Untraceable off-line cash in wallets with observers. Advance in Cryptology – CRYPTO'93*, Lecture Notes in Computer Science 773, Springer-Verlag, 302 -318.

Schneier,B. (1996). *Applied Cryptography , New York,* Wiley,Second edition.

Secure Hash Standard, National Institute of Standards and Technology, 17 April 1995.

Samuel S.Wagstaff Jr. (2003). *Cryptanalysis of Number Theoretic Ciphers*: Chapman & Hall/CRC.

S.W. Golomb, R.E. Peile , R.A. Scholtz. (1994). *Basic Concept in Information Theory and Coding*. New York, Plenum Press.

S.Wicker.(1995). *Error Control Systems for Digital Communication and Storage*: Prentice Hall.

Song Y.Yan.(2002). *Number Theory for Computing*: Springer-Verlag Berlin Heidelberg.

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 M.

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

S. Behnia, A. Akhshani, H. Mahmodi, A. Akhavan.(2008). A novel algorithm for image encryption based on mixture of chaotic maps, Chaos, *Solitons & Fractals*,35, 408–419.

Shafi Goldwasser, Silvio Micali, and Charles Rackoff. (1985). *The Knowledge complexity of interactive proof-systems*. Paper presented at the Proceedings of 17th ACM Symposium on the Theory of Computation, Providence, Rhode Island.

Seyed Morteza Hosseini, Hossein Karimi, Majid Vafaei Jahan. (2014). Generating pseudo-random numbers by combining two systems with complex behaviors. *Journal of Information Security and Applications*,19, 149-162.

Stefanie Falkner, Peter Kieseberg, Dimitris E. Simos, Christina Traxler, and Edgar Weipp. (2014). *E-voting Authentication with QR-codes. Human Aspects of Information Security, Privacy, and Trust*, 8533, LNCS, 149-159.

Suganya Ranganathan1 , Ahamed Johnsha Ali, Kathirvel.K & Mohan Kumar.(2010). Combined Text Watermarking. *International Journal of Computer Science and Information Technologies*, (5), 414-416

*Secure Hash Standard*, National Institute of Standards and Technology, 17 April 1995.

Tamer OZ (2006), Lecturer notes, Random Number Generators.

T.M. Cover, J.A. Thomas, 1992: Elements of Information Theory. New York: John Wiley & Sons.

T.ELGamal,1985 “Public Key Cryptosystem and signature Scheme based on Discrete Logarithms”, IEEE Transaction on Information Theory, 496-472.

T.Okamoto and K.Ohta,1992, “Universal electronic cash”, Advance in Cryptology-CRYPTO’91, Lecture Notes in Computer Science 576, Springer-Verlag, pp324-337.

Tony Warnock. (1987). *Los Alamos Science*, Special issue 1987, Monte Carlo, 137-141.

T. Furon and P. Duhamel.(2000). An asymmetric public detection watermarking technique. Paper presented at the Proceedings of the 3rd International Workshop on Information Hiding (IH ’99), vol. 1768, Lecture Notes in Computer Science, Dresden, Germany.

U. Topkara, M. Topkara, M. J. Atallah.(2006). *The Hiding Virtues of Ambiguity: Quantifiably Resilient Watermarking of Natural Language Text through Synonym Substitutions.* Paper presented at the Proceedings of ACM Multimedia and Security Conference, Geneva.

V.Miller. (1986). *Use of Elliptic Curve in Cryptography. Advances in Cryptology, CRYPTO’85,* Lecturer Notes in Computer Science, Springer Verlag, 417-426.

William Stallings.(2002). *Network security essential: Applications and Standards*: Prentice Hall.

William Stallings.(2010). *Cryptography and Network Security*: *Principles and Practices*: Prentice Hall.

Wade Trappe, Lawrence C. Washington. (2002). *Introduction to Cryptographic with Coding Theory:* Prentice Hall.

W.Diffie, M.E. Hellamn. (1976). New Directions in Cryptography. *IEEE Transactions on Information Theory*, IT-22,644-654.

W.Feller. (1957). *An Introduction to Probability Theory and its Applications* :John Wiley, New York.

Xianhong Zhang. (2003). *Theory and techniques of digital signature*. Beijing: Mechanic Industry Press.

Xinmin Zhou, Sichun Wang, Shuchu Xiong, Jianping Yu. (2010). Attack Model and Performance Evaluation of Text Digital Watermarking, *Journal of Computers*, 5(12), 1933-1941.

Xingming Sun, Alex Jessey Asiimwe.(2005). *Noun-Verb Based Technique of Text Watermarking Using Recursive Decent Semantic Net Parsers. Lecture Notes in Computer Science (LNCS) 3612:* 958-961, Springer Press.

Xianhong Zhang. (2003).*Theory and techniques of digital signature*. Beijing: Mechanic Industry Press.

Yanqun Zhang.(2009).Digital Watermarking Technology: A Review, ETP .Paper presented at the Proceedings of International Conference on Future Computer and Communication.

Y.L. Chiang, L.P.Chiang,W.T.Hsieh,W.C. Chen. (2004). *Natural language watermarking using semantics substitution for Chinese text*, Lecture Notes in Computer Science*,* 129-140..

Zunera Jalil and Anwar M. Mirza.(2010). Text Watermarking Using Combined Image-plus-Text Watermark, Paper presented at the Proceedings of Second International Workshop on Education Technology and Computer Science.

Zunera Jalil, M. Arfan Jaffar and Anwar M. Mirza. (2011), A Novel Text Watermarking Algorithm Using Image Watermark. I*nternational Journal of Innovative Computing*, Information and Control 7(3).

Zunera Jalil and Anwar M. Mirza. (2009). A Review of Digital Watermarking Techniques for Text Documents. Paper presented at the Proceedings of International Conference on Information and Multimedia Technology.

Zunera Jalil, Anwar M. Mirza and Tahir Iqbal.(2010). *A Zero-Watermarking Algorithm for Text Documents based on Structural Components.* Paper presented at the Proceedings of International Conference on Information and Emerging Technologies (ICIET ).

Zekeriya Erkin,Alessandro Piva,Stefan Katzenbeisser,R. L. Lagendijk, Jamshid Shokrollahi, Gregory Neven, Mauro Barni.(2007). Protection and Retrieval of EncryptedMultimedia Content: When Cryptography Meets Signal Processing, *EURASIP Journal on Information Security*, Hindawi Publishing Corporation.

# LIST OF PUBLICATIONS AND PAPERS PRESENTED

Following are the list of publications produced during the Ph.D research studies.

## International Journal Submitted Papers

Chee Hon Lew, Chaw Seng Woo, The Implementation of RSA based on PRNG for Cryptography Application, *Pensee Journal*, 75(11), 2013. (ISSN: 0031-4773, Indexed by Thomas ISI: SSCI, accepted)

Chee Hon Lew, Chaw Seng Woo, Design and Implementation of Pseudo-Random Number Generator (PRNG) Combined with Text based Watermarking for Cryptography Application, *CMES-Computer Modeling in Engineering & Science*, 2017. (Indexed by Thomas ISI: SCIE, accepted).

Chee Hon Lew, Chaw Seng Woo, Practical Application of RSA based PRNG Method For Cryptography Application, *Rairo-Theoretical Informatics and Applications*, 2017. (submitted, under review)

Chee Hon Lew, Chaw Seng Woo, "Practical Application of RSA based PRNG Applied to Text based Watermarking for Cryptography Application, *Turkish Journal of Electrical Engineering & Computer Science* , 2017. (submitted, under review)

## International Conference Publications

Chee Hon Lew, Chaw Seng Woo, Liang Shing Ng, *Using Combined Text based Watermarking with Pseudo-Random Number Generator for Cryptography Application.* Paper presented at the Proceedings of The International Conference on Electrical Engineering and Computer Science (ICEECS-2012), Singapore, 5th. December 2012. (Accepted, ISBN: 978-93-82208-45-7, Published by IRNet International Pte. Ltd.)

Chee Hon Lew, Chaw Seng Woo, *The Implementation of integrating Text based Watermarking and Pseudo-Random Number Generator(PRNG) for Cryptography Application.* Paper presented at the Proceedings of International Conference on Mathematical Science and Statistics(ICMSS'13), February 5-7,2013, Kuala Lumpur(Accepted, ISI /SCOPUS Cited Publication)

Chee Hon Lew, Chaw Seng Woo, *Using Combined Pseudo-Random Number Generator and Text based Watermarking for Cryptography Application.* Paper presented at the Proceedings of The International MultiConference of Engineers and Computer Scientists 2013, Hong Kong, 13-15 March, 2013.(pp. 214-219, ISI Proceedings/SCOPUS Indexed)

Chee Hon Lew, Chaw Seng Woo, *Design and Implementation of Text based Digital Watermarking Combined with Pseudo-Random Number Generator(PRNG) for Cryptography.* Paper presented at the Proceedings of 12th WSEAS International Conference on APPLIED COMPUTER and APPLIED COMPUTATIONAL SCIENCE(ACACOS'13), Kuala Lumpur, during April 2-4, 2013.(pp. 212-218, ISI/SCOPUS Cited Publication)

**National Conference seminar**

Chee Hon Lew, *Design and Implementation of Digital Watermarking by using Pseudo-Random Generator for Cryptography Application*, PhD Symposium 2011, 18-22 July 2011, UNITEN Putrajaya Campus, Selangor