

**FRAMEWORK FOR DATA BREACH
DETECTION AND PRIVACY ASSURANCE
OF ELECTRONIC DOCUMENT**



By

Muhammad Sohaib Hassan

Fall 2016-MS (IS) - 00000172753

Supervisor

Dr. Hasan Tahir

Department of Computing

A thesis submitted in partial fulfillment of the requirements for the degree of
Masters of Science in Information Security (MS IS)

In

School of Electrical Engineering and Computer Science,
National University of Sciences and Technology (NUST),
Islamabad, Pakistan.

(July 2021)

Approval

It is certified that the contents and form of the thesis entitled "Framework for Data Breach Detection and Privacy Assurance of Electronic Document" submitted by MUHAMMAD SOHAIB HASSAN have been found satisfactory for the requirement of the degree

Advisor : Dr. Dr Hasan Tahir

Signature:  _____

Date: 23-Jul-2021

Committee Member 1:Dr. Seemab Latif

Signature:  _____

23-Jul-2021

Committee Member 2:Dr. Mehdi Hussain

Signature:  _____

Date: 23-Jul-2021

Signature: _____

Date: _____

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis entitled "Framework for Data Breach Detection and Privacy Assurance of Electronic Document" written by MUHAMMAD SOHAIB HASSAN, (Registration No 00000172753), of SEecs has been vetted by the undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/M Phil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature:  _____

Name of Advisor: Dr. Dr Hasan Tahir _____

Date: 23-Jul-2021 _____

Signature (HOD): _____

Date: _____

Signature (Dean/Principal): _____

Date: _____


Dedication

I dedicate this dissertation to my parents, wife and honorable teachers for their love and affection.

Certificate of Originality

I hereby declare that this submission titled "Framework for Data Breach Detection and Privacy Assurance of Electronic Document" is my own work. To the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics, which has been acknowledged. I also verified the originality of contents through plagiarism software.

Student Name: MUHAMMAD SOHAIB HASSAN

Student Signature:  _____

Acknowledgment

I am highly indebted and prayerful to the magnificent and Merciful Almighty Allah, who bestowed his immense blessings enabling me to undertake and complete the studies reported in this manuscript. Countless salutations be upon our Prophet Muhammad (peace be upon him) who declared it to be an obligatory duty of every Muslim to seek and acquire knowledge.

I am thankful to my beloved parents and lovely wife for their moral and spiritual support, especially the father for his guideline and prays for my success. They have always encouraged me to upgrade my education status. I am profoundly grateful to my supervisor Dr. Hasan Tahir But for his continued interest in planning, execution and successful completion of the project. It is only because of his consistent encouragement, inspiring guidance, dynamic supervision and sympathetic attitude that enabled me to prepare this manuscript. He will remain a great source of inspiration and kindness for me. The heartiest thanks and gratitude are also extended to my other two respected committee members Dr. Seemab Latif and Dr. Mehdi Hussain for their scholarly contribution, valuable suggestions and constructive criticism toward the successful completion of the thesis. I am obliged to all my respectable teachers for sparing their valuable time and sharing the knowledge. I believe that this work would not have been possible without their cooperation and support.

I will also express my gratitude to my mentor and class fellow Mr. Muhammad Zaid who encouraged and supported me to do MS along with my regular job. Thanks to all my friends and colleagues at Pakistan Civil Aviation Authority especially Ahmed Ali Khan for their help and support. Despite all the assistance provided by supervisor, committee members and others, I take the responsibility for any errors and omissions, which may unwittingly remain.

Table of Contents

1	Introduction	1
1.1	Historical Background	1
1.2	Problem Statement	3
1.3	Goal and Objectives	4
1.4	Thesis Motivation	4
1.5	Thesis Organization	5
1.6	Summary	6
2	Literature Review	7
2.1	Data Security and Privacy Concerns	7
2.1.1	Electronic Data	7
2.1.2	Privacy Concern for Data Loss	8
2.1.3	Data Protection Techniques	9
2.1.4	Data Leak Threats	10
2.2	Data Leakage Detection and Prevention	11
2.3	Novel User Level Data Leakage Detection	11
2.4	Guilt Model	12
2.5	Watermarking Technique	13
2.6	Summary	13

TABLE OF CONTENTS

3	Research Methodology	15
3.1	Identify a General Research Question.....	16
3.2	Preliminary Research on Problem Statement.....	17
3.3	Steps of Research Process.....	17
3.3.1	How to Ensure Data Privacy and Protection?.....	18
3.3.2	How to Ensure Data Authentication?.....	19
3.3.3	How to Identify Source of Data Breach.....	20
3.4	Define the Exact Research Question.....	22
3.5	Detailed Research to Answer the Exact Research Question.....	23
3.5.1	How to Ensure Data Privacy and Protection?.....	24
3.5.2	Types of Watermarking.....	25
3.6	Present Solution to the Research Question.....	30
3.7	Summary.....	31
4	IMPLEMENTATION	32
4.1	Proposed Framework.....	32
4.2	Generator Module.....	32
4.2.1	Watermark Code Generator.....	33
4.2.2	Watermark Code Embedding.....	35
4.2.3	Data Security Module.....	35
4.2.4	Hash Module.....	35
4.3	Extraction Module.....	36
4.3.1	Watermark Code Extraction.....	37

TABLE OF CONTENTS

4.3.2	Guilt Agent Detection.....	37
4.4	Proposed Application.....	38
4.4.1	Application Login.....	39
4.4.2	Admin Login.....	39
4.5	Summary.....	41
5	Conclusion and Future Work.....	42
5.1	Conclusion.....	42
5.2	Future Work.....	43
	Bibliography.....	44

List of Figures

Figure 2-1: Classification of Data Leak Threats.....	10
Figure 3-1: Research Methodology.....	16
Figure 3-2: Public Key Encryption Schemes.....	19
Figure 3-3: Digital Signature.....	20
Figure 3- 4: Simple Non-Repudiation Scheme.....	22
Figure 3-5: Information Hiding Techniques.....	24
Figure 3- 6: Types of Watermarking.....	25
Figure 3-7: Classification of Text Watermarking.....	27
Figure 4-1: Processes involved in Generator Module.....	33
Figure 4-2: Watermark Generator.....	34
Figure 4-3: Text Watermark Embedding.....	35
Figure 4-4: Watermark Code Securing.....	36
Figure 4-4: Watermark Code Securing.....	36
Figure 4-4: Watermark Code Securing.....	37
Figure 4-7: Proposed Application.....	38
Figure 4-8: User Login/Update Form.....	39
Figure 4-9: User Login.....	40

List of Tables

Table 3-1: Solution to Minimum Requirements of Research Question.....	30
Table 4-1: Random Number, White Space and Bit Value Mapping	34
Table 4-2: Manage Application User Forms	40

Abstract

Organizations are handling huge amounts of data on daily basis. This data includes passenger name records (PNR), patient's medical information, financial statements, State level secrets i.e. watch list of terrorist/spy, exit control list, etc. In case secret data is leaked, it could cause huge losses in term of legal proceedings, financial penalties and reputation loss. As same data is being shared to and processed by multiple organizations, it very important to identify exact party from where data has been breached so that responsibility of data loss can be fixed.

Public key encryption can be helpful to map identity of any party (originator or recipient) to the encrypted data, but once data is decrypted, the identity is lost. Digital watermarking is a technique, which embeds hidden code or secret information inside data in decrypted or readable form. However, watermarking alone cannot guarantee non-repudiation as watermarked data is visible to other parties as well i.e. originator. If both techniques are used i.e. public key encryption and text watermarking through a controlled framework or flow, data can be securely shared to multiple parties with an embedded unique identity (watermark code).

In this project, a complete framework is proposed to provide data integrity, confidentiality and non-repudiation of delivery of data, which is being shared, to multiple parties of the system. In case data is leaked the watermark code can be extracted to identify exact location or party of data breach

Chapter 1

Introduction

This introductory chapter includes context and basic information to develop understanding about this research work. The process and terminologies followed are briefly discussed to build essential knowledge base for both novice and expert audience. It also includes the motivation behind this study, and research questions and problem statement to be addressed. Goals and objectives, intended audience, and scope of this work are also incorporated. At the end of this chapter, the organization of this research project is included. Following sections enclose these contents as listed below:

- Section 1.1 Historical Background
- Section 1.2 Problem Statement
- Section 1.3 Goal and Objectives
- Section 1.4 Thesis Motivation
- Section 1.5 Thesis Organization
- Section 1.6 Summary

1.1 Historical Background

With the ceaseless growth of the internet and advancement of online services, the transmission and sharing of digital media are very easy. Sharing of digital media among various organizations and agents of a company is of prime importance and

CHAPTER 1. INTRODUCTION

high value because a large number of users from varied disciplines are using the data for multiple tasks [1]. Digital media could be in the form of image, text, audio, or video [2] but the text is one of the most widely used digital media on the Internet [3]. The data is collected, processed, interpreted and used for different purposes. It may include Personally Identifiable Information (PII) of clients like name, cell phone number, passport number, gender, date of birth[4] or organization-specific data like financial planning, forecasting, business strategy, trade secrets, etc. The IT experts have shown their concern about the security of data [5]. Cybersecurity, just like any other information technology field, poses a great challenge to data sharing requirements as far as data integrity, confidentiality, authentication, and non-repudiation are concerned [6]. Whenever the data is needed to be share, the most important concerns include data privacy and data security. It is a common misconception that data privacy and data security are the same things, instead data privacy is mainly focused on how data is collected, shared and processed whereas data security is concerned with the protection of data from external hackers or malicious insiders [7].

During the data sharing process, the most important concern of the data owners (originator and recipient parties) regarding data privacy is who have processed or accessed their data through the course of its transmission [8]. Another important requirement is the assurance of data origin or guarantee of data delivery (non-repudiation) which provides sending and receiving party with the ability to prove that data has been shared by a particular user and received by a specific user so that neither party can deny the exchange of data at any given point in time [9]. When it comes to the aviation industry both requirements of data sharing becomes a more challenging issue as aviation data includes flight information, passenger's personally identifiable information, and secret/classified information of different border control authorities [10]. Furthermore, if you look at the stakeholders of aviation data, the data sharing process becomes an even more sensitive issue as aviation data is being shared across multiple stakeholders from different organizations of various countries e.g. airlines, ground handlers, airport authorities, immigration authorities, security

CHAPTER 1. INTRODUCTION

agencies, national and international intelligence departments, etc. Similarly, data privacy and data security with non-repudiation assurance is an essential requirement of everyone in this era of digitization and the internet.

Data security can be achieved using different encryption techniques but establishing a non-repudiation claim is not a simple task [11]. Even encryption with the public key of interested parties of the system is not enough because identity is lost once data is decrypted or converted to plaintext form. Another challenge with non-repudiation claims is that the same data is available on the originator side as well. There are techniques, which provide non-repudiation options. Digital Signatures are attached to electronic documents to verify their authenticity (integrity plus sender's identity) [12], whereas watermarking (data hiding technique) is used to identify ownership of the digital assets by embedding a unique identity (watermark code) inside original data [13]. However, digital signatures and watermarks, cannot provide data security. Hence, alone we could have the data security but not non-repudiation and vice versa. Therefore, it is important to develop a solution that could ensure the non-repudiation of data ownership along with data security.

1.2 Problem Statement

Data is the most important and critical asset of the present digital world. Once data is leaked, it could cause short-term or long-term consequences but their outcome can be very detrimental like legal proceedings, monetary damage, loss of trust and reputation. So first place it is important to protect data (data security) but if data is breaches identification of the guilty party (legitimate owner or receiver of the data) who is responsible for a data breach is essentially important so that the responsibility of data abuse can be fixed. The existing data leakage detection models are not advanced enough to identify the guilty party with an assurance to fix claims of non-repudiation of delivery while ensuring confidentiality and integrity of data [14]. Furthermore, the detection process of the guilty party should be quick and simple instead of long and complex forensic investigation, which usually requires access

and detailed inspection of the IT systems of all parties. Public key encryption can be used to secure data and bind user identity before sharing data but once data is decrypted, the identity is lost [11]. Similarly, text watermarking can bind user identity to an electronic document but data security and non-repudiation cannot be achieved.

1.3 Goal and Objectives

Design a framework for data leakage detection models by incorporating additional controls to establish the claims of non-reputation against the guilty party (an agent who has received the data from the owner and is responsible for data breach) while ensuring data confidentiality and integrity. Digital text watermarking will be used to embed a secret identity (watermark code) inside the data requested by an agent. The watermark code will be extracted from the leaked file to identify the suspected guilty agent. To prove the non-repudiation of delivery (NRD) claim against the suspected guilty agent, the data will be shared after encryption using the public key of the requestor (respective agents). Encryption with public key provides data confidentiality and endorsement to NRD. As there is a requirement to keep watermark code and shared file in a database (for comparison and identification of suspected guilt agent), the database security will be ensured by using hash values of the sensitive data. Hashing of watermark code and shared data will also be used to provide data integrity check. To test the functionalities and design of the proposed system, a dedicated application will be developed. The main function of proposed application are:

- Users upload plaintext word documents and get the watermarked encrypted file for each intended recipient with guaranteed data privacy, data authentication, and non-repudiation assurance.
- Upload a leaked electronic document (word file) and identify the exact source/recipient of the leaked file.

1.4 Thesis Motivation

To ensure security in Information Technology services is the need of the hour.

CHAPTER 1. INTRODUCTION

Specifically, when it comes to data, which is always a critical asset for any organization, the value of ensuring security becomes even more important. An organization would never want to lose its propriety information without its consent. It is paramount to keep a strict check on the sharing of sensitive data within or outside of the organization, which can only be achieved by following the standards, which can ensure the traceability of the data shared by any specific person or entity.

The motivation for this project comes mostly from the fact that most organizations these days are trying to adopt various methodologies to protect their proprietary information. They also want to affix responsibility in case of any data leak. Another motivational factor for this project is also a fact, which relates to legal frameworks of data leak. In case of data leak, responsibilities can be affix as per law only in case when the data leak has been proven via well-established standards or some recognized technique, which is proven flawless and is as per the recommendation of IT security best practices, otherwise legal actions cannot be taken for the data leak. Hence, considering the importance of data protection and traceability of data sharing for the organization, it becomes highly well in demand topic, which should be focused and requires a lot of attention in this cyber security era. A lot of work has been done on the encryption and authentication schemes, which provide best authenticity and encryption of the data while shared on the unsecure mediums. These techniques are highly effective, but only during the transmission of the data over the unsecure mediums. These techniques lack the methodologies and techniques to handle the data specifically when it comes to traceability of the data sharing which motivates to work on these gaps and propose solution under the well-established security standards of public and private key model. In case of public and private key model, non-repudiation is always ensured as the sender always sends the data after signing the document or any electronic transaction with his private key, which cannot be denied by the sender at any given point in time and is always traceable as well. Public private key pair is mathematically proven secure algorithm, which can be relied upon easily for ensuring the data security and for the traceability of the data sharing.

1.5 Thesis Organization

The thesis is divided into different chapters for better understanding and semantics. Following is a brief description of the chapters.

- Chapter 1 provides the important basic concepts related to the problem and historical background of the problem
- Chapter 2 explains the previous work done by researchers in this domain.
- Chapter 3 describes the methodology adopted for undertaking the research
- Chapter 4 explains the implementation details of the solution.
- Chapter 5 consists of the results obtained after solution implementation
- Chapter 6 concludes the thesis and presents directions for future work

1.6 Summary

This chapter introduced the basic concepts and core research area on which this study further builds. In this chapter, we explained the importance of digital data and how it is critical for companies to share data among different agents or organizations. The importance of data privacy and data security was explained along with the impact of data breaches. In this chapter, we study some of the biggest data breaches of the last decade and establish that how devastating the consequences of data leakage could be in terms of financial penalties, legal proceedings, reputation damage, loss of trust, etc. Therefore, it is necessary to ensure privacy and protection of data. It is also explained that is important to find out the exact party (guilty agent) who is the last owner of leaked data so that the responsibility of data breach can be fixed with respect to legal proceedings or monetary penalties. Finally, this chapter discussed some of the existing techniques and algorithms to prove data security and non-repudiation.

Chapter 2

Literature Review

This chapter discusses the related work and terminologies. The related work is the research carried out by different researchers over the years, which relates to the work done in this thesis and contributed towards making a new solution. The focus is to present prominent and latest researches related to the work.

2.1 Data Security and Privacy Concerns

In today's world of automation and digitization, Data is one of the most important assets. A lot of research has been done regarding securing electronic data and ensuring its privacy. Due to reliable and efficient transmission mode, data sharing in electronic form is considered to be the most efficient way of data transmission. However, with the ease that data sharing in electronic form brings, it is also prone to several threats, which may challenge its effectiveness. The data shared in the electronic form, should be reliable and should be exchanged between the intended parties. Another major concern, as far as the non-repudiation is concerned, the data shared should be always traceable in order to ensure data integrity and privacy.

2.1.1 Electronic Data

Electronic devices, which are used to receive, store, process, or transmit information, are called digital devices which include personal computers, cellphone, sport and entertainment systems, location tracking systems, and so forth. People use such

CHAPTER 2. LITERATURE REVIEW

devices for personal use cases in their daily life, local as well as global communication. Data saved in these gadgets is of “digital shape” this is stored in the form of binary units of zeros “zero” and ones “1”. Information in different media types which includes audio, video, text and images are stored as basic units on the storage of these digital tools. A user can read these patterns in understandable format through respective piece of code (software) [15].

These digital devices especially computers and cell phones have provided organizations with the ability to complete business processes such as supply chain management, marketing, forecasting, finance, and complying with industry standards and government regulations [16]. Businesses are adopting technology innovations to drive efficiencies within business operations to increase the value [17]. However, despite the positive effects technology has on businesses, IT services have introduced data security and privacy issues [1]. As data slip through, exposure of confidential, sensitive, or protected information to an unauthorized person. Simply when the digital data (files) are viewed and/or shared without permission is called data breach [18].

2.1.2 Privacy Concern for Data Loss

A Data breach is the exposure of secret or confidential information to unauthorized parties either intentionally to unintentionally. In the era of the digital and electronic world, data is the most valuable and critical asset of any organization. Data leakage could cause a serious impact on the financial position and market reputation of any organization. Detection and prevention of data breaches have become one of the most challenging concerns for all organizations and states across the world as data is growing exponentially. Despite the huge efforts of researchers on safeguarding sensitive information from being leaked, it remains an active research problem.

This review is useful for the intended audience as it includes threats, recent incidents and various protection and prevention techniques regarding data breaches. In 2014, a data breach exposed the account information of 500 million users of Yahoo [19], the largest data breach of that time. The confidential information lost in this theft includes names, date of births, contact numbers, email addresses, a hash of passwords and even security questions and answers. 2015 was a banner year of data breaches in the healthcare industry and 98% of those breaches were caused by IT

CHAPTER 2. LITERATURE REVIEW

hacking incidents [20]. In the same year, network hacking of Medical Informatics Engineering Company exposed the record of 3.9 million patients [21]. The theft information includes social security numbers, diagnoses, and other personal information like name, date of birth, cellphone numbers and mailing addresses. In late 2016, personnel information of around 57 million users and 600,000 drivers of Uber application were stolen [22]. Additionally, the hackers were able to access Uber's Amazon Web Service and GitHub account as well. In 2017, a U.S. credit reporting agency "Equifax" leaked the data of 145.5 million American consumers, which includes much critical information like names, Social Security numbers, addresses, driving license numbers, etc. British Airways was fined \$26m in 2018 for a data breach of more than 0.4 million customers [22]. The same year also witnessed a breach of 4 million passenger's data of Cathay Airlines [23]. In Sep 2019, usernames, email IDs, and Facebook IDs of 218 million customers of a Mobile gaming company "Zynga" was breached by hackers [23]. April 2020 witnessed another data breach of over 500,000 Zoom user account information [24]. The breached accounts were found on the dark web for sale with login credentials, passwords, personal meeting URLs of accounts. So data privacy and protection is one of the most important concerns of the present world. It is evident from the above literature review that data breaches affected almost all industries from healthcare to transport to the communication industry. It is very important for researchers to develop frameworks, algorithms or techniques, which could detect and prevent data breaches quickly and accurately instead of complex and time taking forensic investigations.

2.1.3 Data Protection Techniques

Data protection is not a new concept. People are protecting their secret data or information for years e.g. the first recorded use of the steganography term (information hiding technique) was in 1499 by Johannes Trithemius [25]. Now there are a number of techniques, which can be used to protect data, which includes:

- **Encryption:** It is the process of transforming information (plaintext) to an unreadable form (ciphertext)
- **Watermarking:** This is a technique to hide information in the digital assets such that it remains unnoticeable or unidentifiable to everyone except the originator

or intended recipient

- **Authentication:** Authentication enable the system to identify a legitimate user by verifying credentials.
- **Data masking:** Data is hidden by obscuring letters and numbers with proxy characters i.e. data is present behind the masking. The data is changed to original form when authorized users access the data.

2.1.4 Data Leak Threats

Data leakage threats can be classified in a number of ways, which is mainly based on their causes or hacker role. Data leakage can be classified as intentional threats or inadvertent threats [26]. In addition, there is another way is to classify data threats based on the party who is responsible for data leakage that is an insider to outsider threats. As shown in Figure 2-1, intentional data leakage can be either caused by external or internal threats/hackers.

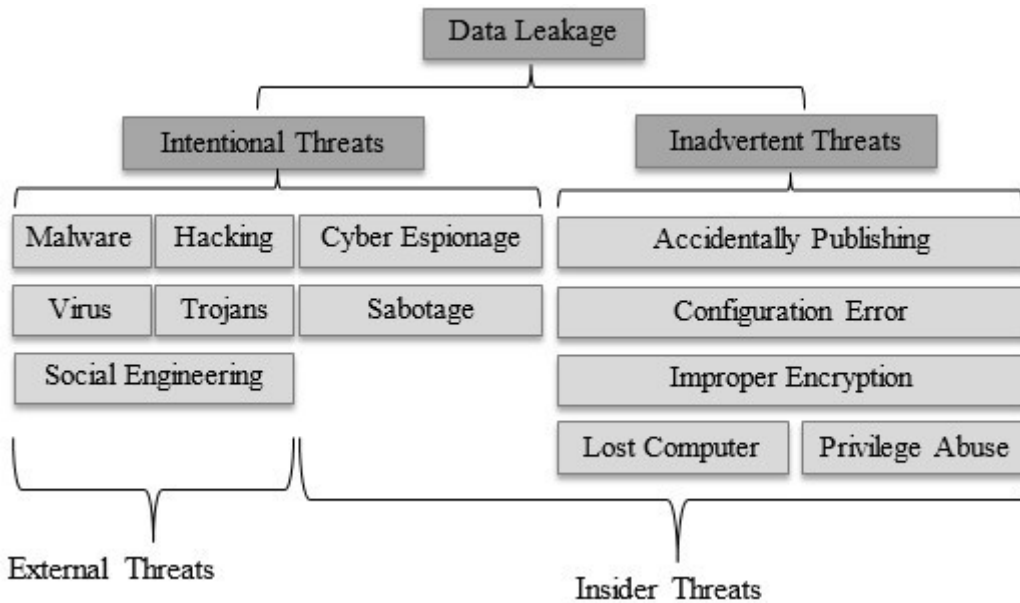


Figure 2-1: Classification of Data Leak Threats

It is evident from Figure 2-1 that intentional breaches can be caused by either insider or external threats whereas inadvertent threats or breaches are due to insider threats

only.

2.2 Data Leakage Detection and Prevention

Companies are doing businesses through different agents and organizations, which exchange massive amount of Data among multiple parties. During data exchange, a great probability of data leakage exists which could be because of malicious intent or unintentional act by any agent. However, consequences of data leakage can be seriously massive and could cause huge financial and strategic losses. Keeping in view the importance of data security and confidentiality, various researchers have contributed to protect flow of sensitive information towards to unauthorized hands.

One of those researches has been done by Rajat Verma [14] in which he analyzed and compared many existing Data Leakage Detection (DLD) models in detail. The DLD models compared in this research work includes “Minimum Overlap Model”, “Fake Object Model” and “Guilt Assessment Model” Minimum overlap model states that the probability of leaked file is dependent on distributor and number of agents involved in data sharing or information processing. In case there is one agent the Probability of guilty agent $Pr = \frac{1}{n+1}$ (where n is a number of agents). In the fake object model, the fake objects are added in original data in an attempt to identify source of data breach. Although, fake objects are not real data objects but like watermarking, they can be linked to a set of users so that source of leakage can be traced. In the guilt assessment model the guilt probability is calculated to already available data sets, which includes user/agents, original data objects, fake objects, and leaked data objects to find the guilty user who is responsible for data leakage.

2.3 Novel User Level Data Leakage Detection

Novel user level data leakage detection algorithm (NULDLDA) [27] has served one the primary motivation for the proposed framework in this research paper. In this paper, they have implemented NULDLDA and compared it to existing Data leakage detection (DLD) algorithms. NULDLDA mainly focused to find the exact location of

an agent whose data is leakage among several other agents of the system. “Data Allocation Module” of NULDLDA is responsible for the distribution of data that how data can be shared intelligently to all agents (recipients) so that it produces a chance of detecting a guilty agent. The distributor adds fake articles to original data by using “Fake Object Module”. Fake articles are objects that act as secret code inside shared data. The concept to use fake objects is similar to watermarking as fake articles are used as a means to establish original ownership of distributed objects. “Optimization Module” is used which ensures that secret code is different for all agents. “Data Distribution Module” is used to specify an objective to distribute its data among approved agents. The objective is to detect the agent who causes data leakage. Finally, the goal of the “Agent Guilt Module” is to find out or distinguish the agent who has the most probability of leaking his data (under investigation).

2.4 Guilt Model

Guilt model [14] focused on detecting the agents using allocation strategies, instead of performing modifying in the original data objects represented by set $T = \{t_1, t_2, \dots, t_m\}$ that is by insertion of fake objects F . The researcher has used a variety of data distribution strategies to improve the distributor’s chances of identifying the guilt agent U_i . A guilt agent is one who leaked a portion (sample data) or complete (explicit) data shared with him. The researcher has implemented an algorithm using fake objects to improve the chances to detect the guilt agent. Fake objects are used as a link to establish the ownership of distributed objects i.e. with respect to receiving party. The concept of using fake objects is also similar to watermarking techniques as it provides the target agents with some kind of identification information, which is linked to the identity of the receiver (agent). The researcher has also developed a framework “Agent Guilt Model” to generate fake objects. The researcher has divided agent requests to access data objects into two classes that are sample data requests and explicit data requests. In sample data requests, the agent has request the distributor to share a portion of a data object whereas in explicit data request the

distributor is requested to share a complete data object (based on some conditions). Based on agent request four fake objects are created with names $EF, E\bar{F}, SF$ and $S\bar{F}$. Here E is used for explicit requests, S for sample requests, F for fake objects, and \bar{F} is used for fake objects (as fake objects may impact the accuracy of what agents do, so they may not be allowable). The researcher also introduced the concept of guilt probability to maximize the chances of detecting a guilty agent (among a set of agents " V_t ").

In order to find the probability that an agent p is guilty, a set S (leaked data) is given to guess the probability p . He first computes the probability that the guilty agent U_i leaked an object t to S which equals to $P = (1 - p)$. Then he computed another probability by assumed that all agents can leak t to S with equal probability and calculated accurate guilt probability that is $Pr = (1 - \frac{p}{vt})$.

2.5 Watermarking Technique

While keeping the importance of data privacy and its integrity, various researches have been done so that data in electronic form can be exchanged safely. One of those researches has been done by Milad Taleby Ahvanooy [3] in which he analyzed existing watermarking techniques in detail specifically when it come to their effectiveness. The research specifically outlines two watermarking techniques namely linguistic (or natural language) and structural (format based). The shortcomings in both the techniques are elaborated in detail along with a comprehensive comparison between the two and conclude that structural techniques (format based) are more suitable for ensuring privacy and integrity of the sensitive documents.

The researcher explained a general architecture of "Text Watermarking" Techniques as well which consist of two main phases. The first phase is "Watermark Embedding" in which a watermark code is created and inserted into the original document whereas the second phase "Watermark Extraction" the watermark code is extracted and then authenticated (to verify integrity of the watermark code). The researcher also elaborated various types of attacks and their effectiveness of different watermarking techniques i.e. advantages and weaknesses of current watermarking techniques. The detail explanation of structural text watermarking techniques are

discussed in Section 3.5.2.

2.6 Summary

This chapter presented studies related to the proposed framework, such as techniques for Data security, data authentication, data privacy and data leakage threats. In addition, some of the practical implementation various data leakage detection and prevention algorithms considered which are based on watermarking code, fake objects and guilt probability including Text Watermarking, NULDLD and Guilt Model.

Finally, all data leak detection techniques and algorithms were analyzed to propose and implemented an improved data leakage detection framework which is based on existing algorithms and history research work.

Chapter 3

Research Methodology

Research is an inventive and abstruse approach to assay a tangled issue, study past work and analysis on collected information to ascertain new facts regarding a particular issue or concern. This chapter includes the entire strategy encompassing different techniques of the research methodology of this thesis to identify the research question and abstract the solution. The steps followed during this study consists of the following steps:

- Identification of General Research Question
- Preliminary Research on Problem Statement
- Define the Exact Research Question
- Detailed Research on Exact Research Question
- Present Solution to the Research Question

The pictorial representation of 5 steps [28] that were used in this research project are shown in Figure 3-1

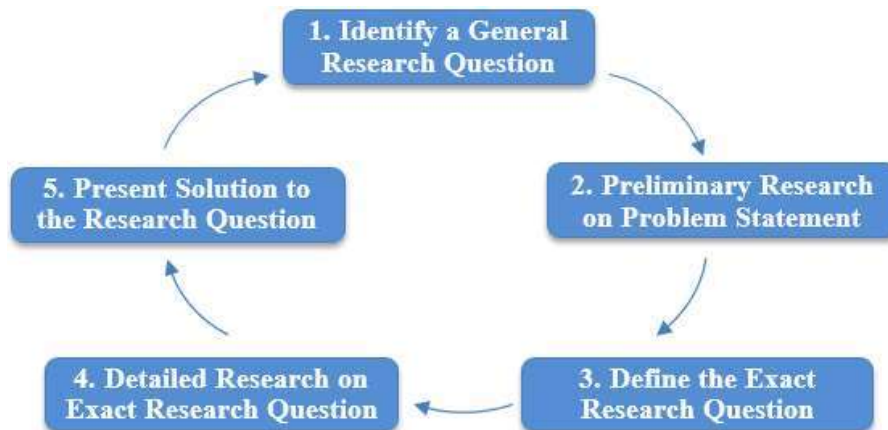


Figure 3-1: Research Methodology

In this subsequent chapter, we discussed the five steps of research cycle for this study. For better understanding, this section also includes the sub-steps of these main phases using appropriate diagrams and tables for a comprehensive understanding

3.1 Identify a General Research Question

As a researcher, your goal should always be to find gaps that your work could fill. One of the best approaches to find a problem that has particular relevance for your job is reading reports and articles from well-known websites and blogs of your field [29]. British Airways was fined \$26m in 2018 for a data breach of more than 0.4 million customers [30]. The same year also witnessed a breach of 4 million passenger's data of Cathay Airlines [31]. In 2018, the United Kingdom Act of Parliament also replaced Data Protection Act 1998 with European Union's General Data Protection Regulation (GDPR) to strengthen individual rights regarding the processing and free movement of their personal data [32].

United Nations Security Council passed a resolution in 2017 to reiterate the importance of sharing Advance Passenger Information (API) and Passenger Name Record (PNR) among States and different Security organizations to improve border control management to track terrorists [33]. The States were focusing on legislation to Secure Personal Data whereas Security Organizations were probing the importance of sharing Passenger data (API & PNR) to strengthen border controls in the aviation industry. This created an opportunity to explore techniques that could be used to secure data, which is supposed to be shared among different Parties. Also Forensic

CHAPTER 3. RESEARCH METHODOLOGY

Investigation of data breaches that is being shared among different states and organizations is a complex task [34] as data is shared and processed by different stakeholders at different geographical locations using separate IT systems. Even contrary to the above statement, Forensics analysis can produce results that could identify the possible causes and source of a data breach (data which was shared with multiple parties), but still it is very difficult to fix the responsibility on a single one as same data was delivered to everyone in the system

Up to this point, we were able to draft a general problem statement that, a framework, which could address below mentioned answers:

- How to ensure Data Privacy and Protection
- How to ensure Data Authentication (Integrity and Sender's Identity)
- How to identify source of data breach (without complex forensic investigation)

3.2 Preliminary Research on Problem Statement

Information cannot be shared until sufficient data protection measures have not been taken. For example, an individual needs assurance that the received bank statement is a true copy issued by the bank (authenticity) and has not been altered (integrity) at any stage during the delivery process or it has not been viewed/copied by anyone except the addressee (confidentiality). At least the individual would like to have the guarantee that the bank statement is original (authenticated) [35].

The goals for the preliminary research were to have a thorough review of relevant literature on data privacy, data security, data forensics and non-repudiation and to identify gaps & techniques that could address the research questions.

3.3 Steps of Research Process

In this study, the following steps will be followed for literature review:

- Search for Relevant Research Literature e.g. research paper, article, books etc.
- Shortlist the more Relevant Papers

CHAPTER 3. RESEARCH METHODOLOGY

- Review the Abstract and Conclusion of each link
- Read all shortlisted literature and write important points in bullets
- Analyze all the bullets of shortlisted literature collectively
- Select the most relevant Papers
- Explore the final selected Papers comprehensively

3.3.1 How to Ensure Data Privacy and Protection?

The terms data protection and data privacy are mostly used interchangeably, but in actual there is a key difference between the two. Data privacy defines who has access to data whereas data protection deals with policies and tools to actually restrict access to the data [36]. Data is classified into two categories based on the type of protection it needed [15].

- **Data in Transit** - Data that is being transferred from one system to another or one location to another
- **Data at Rest** - Data stored in a system or a location

Considering the research question under reference the author has focused security of data in transit.

3.3.1.1 Encryption

Encryption is the process of transforming information (plaintext) to an unreadable form (ciphertext) whereas the reverse process that transforms ciphertext to plaintext is called Decryption. Encryption systems use keys that combine with an encryption algorithm to generate ciphertext from plaintext and recover plaintext from ciphertext [35].

There are two main types of encryption keys that is asymmetric and symmetric.

- **Symmetric Encryption**
The encryption technique, which uses the same key for both encryption and decryption, is called symmetric encryption. Symmetric key encryption is not only fastest but also difficult to break for keys of large sizes.

CHAPTER 3. RESEARCH METHODOLOGY

- Asymmetric Encryption

In this encryption technique, separate keys are used for encryption and decryption. Diffie and Hellman, (1976), also know asymmetric encryption as public key encryption, which was first purposed. One key is private that is only known to the owner whereas the public key is distributed publicly to the other parties of the system. The core idea is that it is impossible to compute a private key from provided public key. RSA[37] and ElGamal [38] are prominent asymmetric algorithms of the system. There are two methods of using public key encryption to secure data as shown in Figure3-2.

1. Method-1: Use private key for encryption and public Key for decryption
2. Method-2: Use public key for encryption and private key for decryption

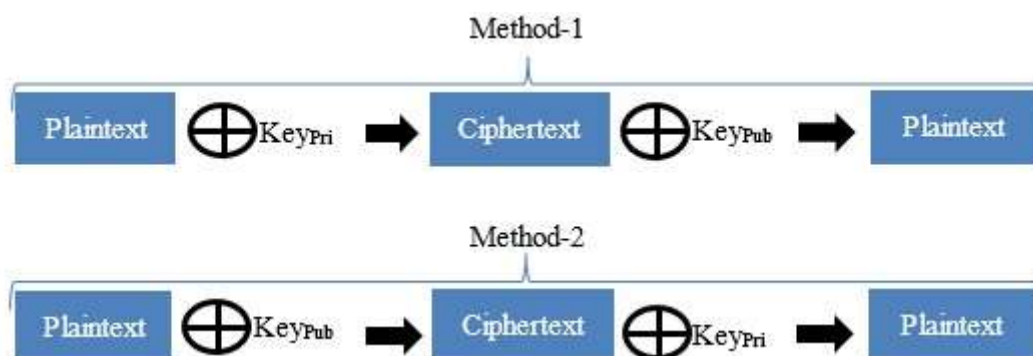


Figure 3- 2: Public Key Encryption Schemes

Method-1 of public key encryption is used for non-repudiation tasks for proof of origin (for detail please refer to Section 3.3.3.6) whereas method-2 is be used for privacy assurance or confidentiality of data. Therefore, the problem for the research question#1 is very simple that is to use public key encryption method-2.

3.3.2 How to Ensure Data Authentication?

Digital data is considered as authentic if it is prove able that it has not been corrupted after its creation [39]. To ensure data authentication, one of the most widely used cryptographic functions is the digital signature which is analog of regular physical

CHAPTER 3. RESEARCH METHODOLOGY

signature [40]. A digital signature serves the same purpose as of regular signatures:

- Digital signatures offer data authentication.
- Digital signatures provide non-repudiation of origin as well.
- Digital signature also ensures the integrity of signed data.

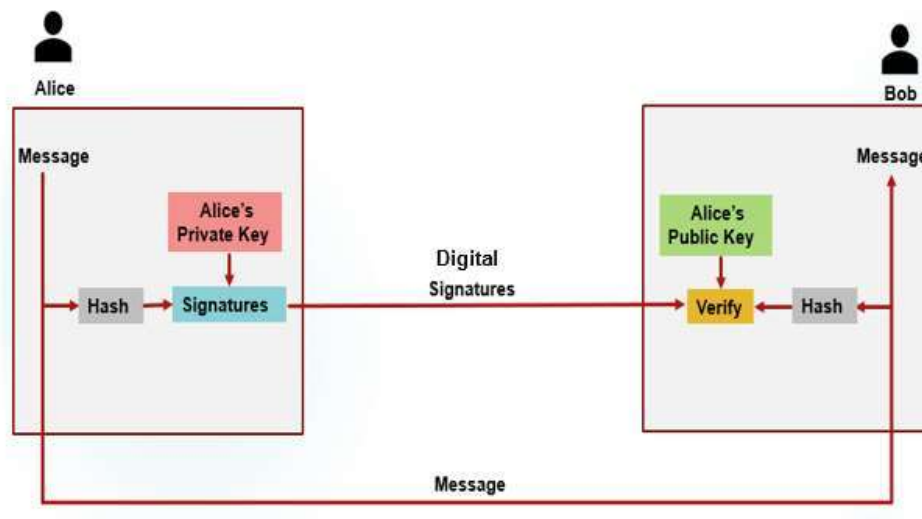


Figure 3-3: Digital Signature

Figure 3-3 shows the working principle of digital signatures. Here, Alice wants to send a message to Bob with a digital signature. To develop a digital signature, Alice takes the hash of the message in the first step. In the second step, Alice sign the message by encrypting the hash with her private key. Now Alice can send the message to Bob along with the signature. To verify the signatures, Bob also takes a hash of the received message. Bobs also decrypt the signature sent by Alice using Alice's public key. If the hash value generated by Bob and decryption value matches with each other, the Signature verification is successful else signature verification is considered as failed.

3.3.3 How to Identify Source of Data Breach (without complex forensic investigation)

As the data in this research question flows between different agents of various stakeholders, so it is very difficult to pinpoint the exact source of a data breach (guilty

CHAPTER 3. RESEARCH METHODOLOGY

agent). Even if we identify a guilty agent using a fake object, guilt probability or watermarking, it is apparently looking impossible to fix the responsibility of data leakage on the guilty agent as same data is processed by the distributor agent/object as well. The assurance that agent cannot deny leaking of data object is case on non-repudiation classification or study.

3.3.3.6 Non-Repudiation

Considering message handling scenarios in consideration, there are three kinds of non-repudiation scenarios that should be ensured by originator, by the recipient and by the delivery agent [41]

- Non-Repudiation of Origin (NRO)
NRO provides proof to the receiver of a message that it is being sent from a specified originator.
- Non-Repudiation of Delivery (NRD)
NRD provides proof to the originator of a message that the message has been delivered to the specified recipient.
- Non-Repudiation of Receipt (NRR)
NRR provides proof to the originator of a message that the message has been receipt by the specified recipient.

Non-repudiation in peer-to-peer protocol is implemented by sending a signature with the message, thus providing the recipient with the proof of origin for the message. Similarly, the recipient in turn can also send a signed proof of receipt to the originator as shown in Figure 3-4

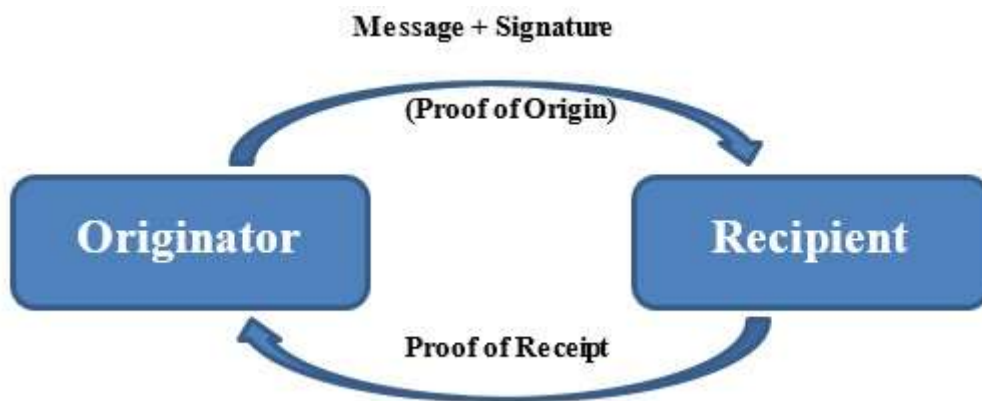


Figure 3-4: Simple Non-Repudiation Scheme

Generally enforcing mandatory evidence of receipt with non-repudiation scheme is difficult to implement and optional in nature [42]. Here it is very important to reiterate the problem statement of the research question #3 that is “How to Identify Source of Data Breach (without complex forensic investigation)”. Even though if we implement evidence of receipt/proof of receipt module in the proposed framework, the same data is distributed to all receipt states or organizations of the system. Also same data is available at the originator side [42].

3.4 Define the Exact Research Question

At this stage of research, many research papers were reviewed to identify any non-repudiation technique or framework, which could identify exact source of data breach even same data, is being distributed to other parties as well.

- Public key encryption using public key of the recipient for encryption and private key for decryption provides
- Confidentiality and data privacy assurance to address question # 1 of the proposed research. Similarly use of Digital Signature ensures data authentication to address question # 2.
- Whereas non-repudiation even with proof-of-receipt cannot provide the system with exact identification of source of data breach to address problem stated in question # 3 of this research proposal.

CHAPTER 3. RESEARCH METHODOLOGY

“At this stage, question # 3 of this research question had given an opportunity to define exact research question of this proposal”.

To find solution of question # 3, differ research literature were reviewed, discussed and exchanged the relevant information with different IT security and Data forensic experts to identify minimum requirements to address the requirement of the proposed framework, which includes:

- Bind unique identities or tags before sharing a data to any recipient.
- Use unique identities for each recipient (if a data is being shared with more than one recipient)
- Tagged data with unique identity should be confidential (only be shared with intended recipient, even originator of data should not able to see unique identities or tagged data).
- Information regarding Unique Identities, tagged data and intended recipient must be stored in a secure way that data breach of this data would not affect the integrity of the framework (to ensure non-repudiation of recipient)

The specific objective of the study is to develop a framework to address all the three research questions (mentioned at the end of section 3.1). Therefore, it is, therefore; important for the researcher to confine the scope of his work to achieve the given task only. Theoretically, the proposed framework supports all kind of data either electronic, scanned or printed documents but the developed framework only supports Microsoft .doc files in electronic form. From above points, the literature review is concluded and the exact researchable question is identified that is:

“Development of a framework for data breach detection and privacy assurance of electronic document”

3.5 Detailed Research to Answer the Exact Research Question

Data Origin Authenticity in the research questions means that when a document is

CHAPTER 3. RESEARCH METHODOLOGY

found, the origin of that resource should be traceable whereas non-repudiation states that the true recipient of the document should be detected and proven. Encrypting data with public key of recipient party ensures that data is readable for the recipient party only and provides a sense of non-repudiation of delivery (NRD) but once data is decrypted it loses the property of NRD again as same data in plaintext form is available at originator side as well [3].

After reading many research papers on non-repudiation and data security techniques it can be concluded to find a method, which enables the originator to embed any secret information before sharing document and extract the same information to detect the recipient of the document. Information hiding can be a useful technique to hide an information across different media types e.g., text, image, audio, or video[43]. The concept of cryptography and information hiding techniques is same in a sense that both are used to secure sensitive information. However, the imperceptibility is the difference between both techniques; that is, information hiding concerns how to hide information unnoticeably. The information hiding techniques can be further categorized into steganography and watermarking as shown in Figure 3-5.

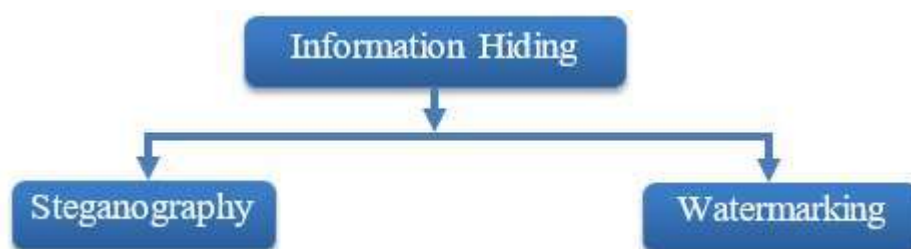


Figure 3-5: Information Hiding Techniques

3.5.1 How to Ensure Data Privacy and Protection?

3.5.1.1 Steganography

Steganography is the combination of two Greek words “Stegano” and “Graphy” [44]. Stegano means sealed and Graphy means writing and hence the combined meaning produced as secret writing.

CHAPTER 3. RESEARCH METHODOLOGY

Steganography aims to hide information before transmitting the secret messages message without the observer being able to notice the hidden information inside the secret message therefore, the main goal is how the hidden information remains unnoticeable [44].

3.5.1.2 Watermarking

During last decade, watermarking techniques have been used to protect copyright protection, media forgery, authentication of data and for prevention of tempering of digital assets (e.g., text, image, audio, or video) [15]. Watermarking is concerned with embedding a unique identity into a digital asset such that the identity remains unnoticeable or unidentifiable to everyone except the originator. Whenever there is a query regarding origin or owner of asset, the hidden watermark can be referred.

3.5.2 Types of Watermarking

Watermarking strategies are based on different factors which can be classified into different forms[45] as shown in Figure 3-6.

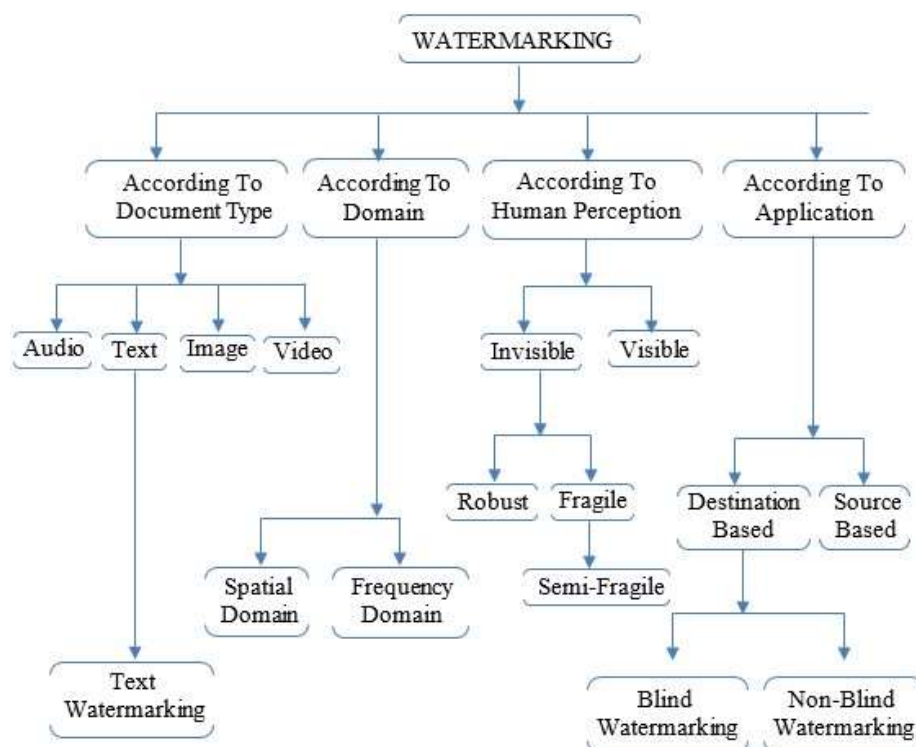


Figure 3-6: Types of Watermarking

3.5.2.1 Watermarking Techniques Based on Document Type

Watermarking techniques can be classified based on different digital media forms e.g. audio, text, image and video.

- **Audio**

A unique audio identity (pseudo noise or narrow-band signal) is embedded into an original audio signal. It is typically used to identify ownership of copyright of audios.

- **Image**

Watermark identity is embedded into the original image. The identity can be a string of bits or text or an image like a symbol or logo.

- **Video**

Watermarking technique, which refers to embedding of a unique identity in a video sequence. Video watermarking is used for copyright claims and to protect the video from illegal copying or editing.

- **Text**

In Digital text, watermarking a secret code or information is embedded into original data (plaintext) in a way that secret information is invisible to everyone except the originator.

Text Watermarking is the most difficult and important type in the digital watermarking. Text media is more challenging as any significant change in visual of text may change the meaning of the words or sentences [45].

3.5.2.2 Techniques of Text Watermarking

Text watermarking has a number of strategies and techniques, which are based on the use case. From the text processing point of view, digital text watermarking techniques can be categorized into two sub-classes namely, linguistic (natural language) and structural (format based) [3] as shown in Figure3-7.

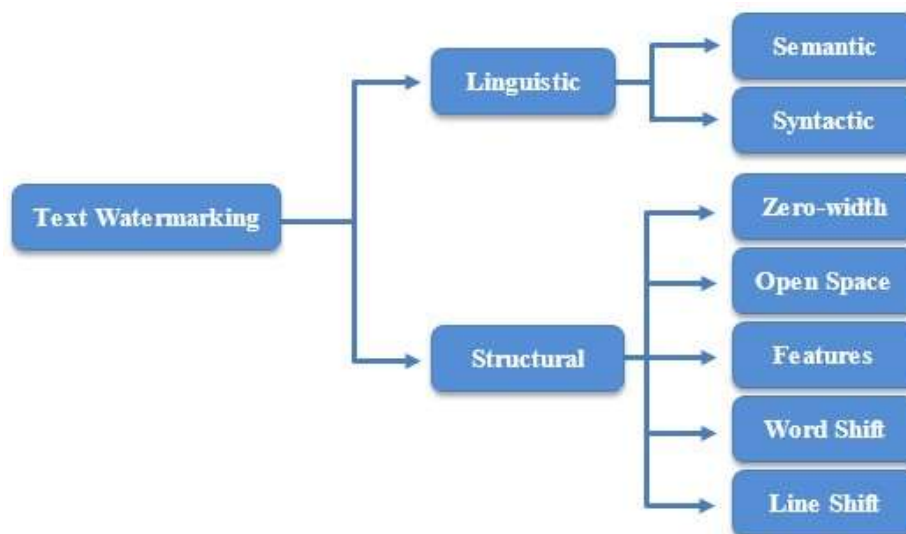


Figure 3-7: Classification of Text Watermarking

- Linguistic Watermarking.

In this technique, the content of a text document modified to hide a watermark identity or binary string. It generally deals with the structure of text using nouns, adjectives, idioms verbs etc. to embed watermark [29]. There are two methods to apply Linguistic Text Watermarking technique:

 - Semantic Watermark
A scheme in which watermark is embedded by changing the text in a way that its meaning remains unchanged.
 - Syntactic Watermark
A scheme in which structure of the sentence is changed to embed the watermark.
- Structural Watermarking.

Structural layouts or special characteristics of the text are changed in order to hide the watermark bytes or identity. Structural layouts may consist of spaces in between words or paragraphs, font size, font style etc.

 - Open Space
In this technique, he used white spaces in text documents for three different location, which includes inter-word spaces, inter-sentence spaces and end-

CHAPTER 3. RESEARCH METHODOLOGY

of-line spaces. In order to hide watermark code or identity, the algorithm inserts additional spaces. As an example two white spaces between words represent a bit value of “1” whereas one white space represents bit value of “0”. Similarly, single space between sentences represents bit value of “0” and double space between sentences represents bit value of “0”.

- Line Shift

In line shift watermarking, all even lines are shifted vertically up or down by a fractional but a predetermined value. The vertical up or down shift represents binary “0” or “1”. Odd lines are used as control lines to identify shifting of the even lines and their position [46].

- Word Shift

All even word are shifted horizontally left or right by a fractional but a predetermined value. The horizontal left or right shift represents binary “0” or “1”. Like line shift, Odd words are used as control words to identify shifting of the even words and their position [46].

- Features

In this technique, the algorithm changes some characters in the text e.g. increasing height of a specific alphabet, changing color of a specific character, changing font style of character after a specific interval. Any specific change in feature of text can be marked as bit value “1”, similarly no change in feature can be marked as bit value “0”

3.5.2.3 Watermarking Techniques Based on Domain

Considering some special characteristics of images, watermarking can be classified into two main classes i.e. spatial and frequency domains.

3.5.2.4 Watermarking Techniques Based on Human Perception

Based on Human perception, the watermarking techniques can divided into two categories visible and invisible.

- Visible Watermark

It includes any visible identity of the owner e.g. text or logo.

CHAPTER 3. RESEARCH METHODOLOGY

- Invisible Watermark

Invisible watermarks are embedded in a way that the human eye cannot notice them. The insertion rate is usually very small to be noticeable to human eye and can be retrieved through watermark extraction method.

3.5.2.5 Watermarking Techniques Based on Application

Based on use cases like data authentication, copyright protection, broadcasting monitoring, non-repudiation, forgery detection and data integrity, the watermarking techniques can be classified into two main categories i.e. destination based and source based [47].

- Source Based Watermarking

In this type of watermarking all copies of a particular data (which is supposed to be shared with multiple parties) have a unique watermark. Source based watermark identifies the owner of that data.

- Destination Based Watermark

In this type of watermarking all copies of a particular data (which is supposed to be shared with multiple parties) have different watermarks, but each watermark is unique with respect to recipient of data. Destination based watermark identifies recipient of data.

Destination Based Watermarking can be further classified into three sub-classes blind, non-blind and semi-blind [48].

- Non-blind Watermark

If the original image is required to extract the watermark, the technique is called as Non-Blind Watermark. The practical use cases of the non-blind watermark technique are limited as it requires extra storage to maintain the original image [49].

- Semi-Blind Watermark

If the watermark itself or any other extra information is required (instead of the original image) to detect the watermark code, the technique is called as Semi-Blind.

- Blind Watermark

CHAPTER 3. RESEARCH METHODOLOGY

A technique, which does not require the original or any extra information, is called as Blind Watermark

3.6 Present Solution to the Research Question

Digital text watermarking is a technique, which can be used to address the minimum requirements, which were identified in Section 3.4 regarding Question#3 of this Research Proposal. The solutions against above-mentioned requirements are stated in Table 3-1. These solutions are identified based on research work mentioned in Section 3.5.

Requirements	Solution
Bind unique identities or tags before sharing a data to any recipient.	Digital Text Watermarking
Use unique identities for each recipient (if a data is being shared with more than one recipient)	Destination based Watermarking
Tagged data with unique identity should be confidential (only be shared with intended recipient, even originator of data should not able to see unique identities or tagged data).	Encrypt Watermarked data with public Key of respective recipient party Watermarked data should not be visible to originator (sending party)
Information regarding unique identities, tagged data and intended recipient must be stored in a secure way that data breach of this data would not affect the integrity of the framework (to ensure non-repudiation of recipient)	HMAC of Watermark code will be stored in database

Table 3-1: Solution to Minimum Requirements of Research Question

3.7 Summary

This chapter is the representation of complete flow of research work followed in this project. Various research papers were studied and different news and technology forums/websites were referred to identify a general problem statement. Data leakage problems are growing many folds because hackers are growing in technology and advance techniques whereas IT experts still relying on forensic investigations of IT gadgets which is not only a complex and lengthy process but also hackers know gray areas of those forensic techniques. Existing models and techniques were explored through different research papers to shortlist techniques that could help in identification of guilty agents using leaked data file itself. Watermarking is a technique, which could be useful as it did not add noise to original data set and keeps a secret identity embedded in original data in plaintext form.

Chapter 4

Implementation

In this chapter, the design, implementation details and tools that are used in this proposed framework will be explain. This chapter will explain the individual modules of application that is developed to implement the proposed framework. Here, the complete flow and functionality of framework will be described to achieve desire goals and solutions to the research questions.

4.1 Proposed Framework

In order to solve research question that is to “Develop a Framework for Data Breach Detection and Privacy Assurance of Electronic Document”, the researcher has identified the required cryptographic and digital text watermarking modules, which are used as building blocks of this proposed framework. The proposed framework consist of two main modules, which includes:

- Generator Module
- Extraction Module

4.2 Generator Module

This module is used to generate an encrypted and watermarked file from the original (input) file, which can be forwarded to other parties of the system. This component ensures data privacy, data integrity and non-repudiation of delivery. The generator module consist of following sub-modules as shown in Figure 4-1:

- Watermark Code Generator:
It is responsible to generate unique watermark codes for all parties (agent) of the system.
- Watermark Code Embedding
This module embeds the unique watermark codes in original data file (.docx format) which is supposed to be shared with respective agents
- Data Security Module
This module encrypts the watermarked file with public key of agents
- Hash Module
Watermark Code and “encrypted + watermarked” file are hashed using HMAC to identify guilty agent in Extraction Module.

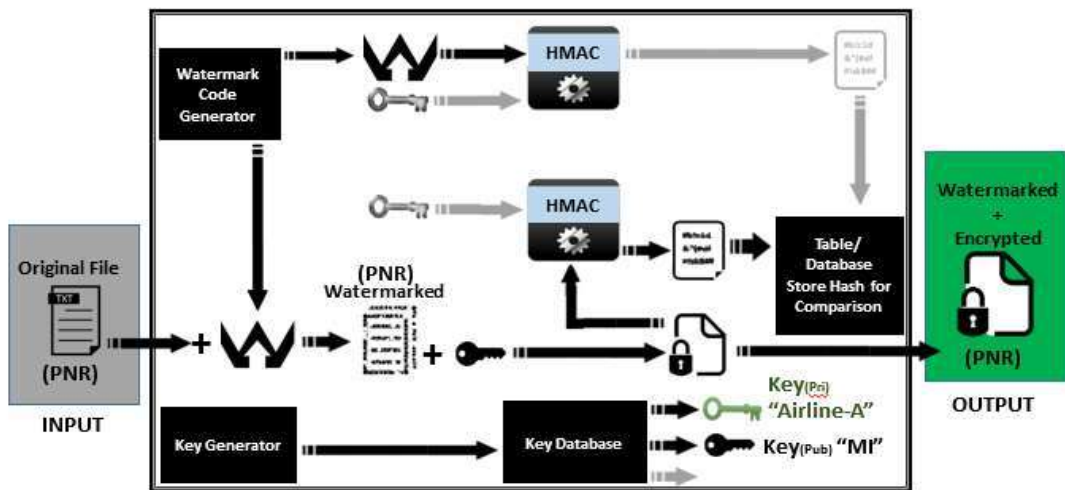


Figure 4-1: Processes involved in Generator Module

4.2.1 Watermark Code Generator

The digital asset, which is used in this research work, belongs to text class (.docx format) as shown in Figure 3-6. So according to the document type classification, text-watermarking technique is used. To develop text watermark generator, “Open Space” technique of Structural Text Watermarking is referred. Watermark code consist of 8 bits in which bit value “0” represents one white space whereas to represent bit value “1”, three kind of white spaces are used which includes two white

CHAPTER 4. IMPLEMENTATION

spaces, three white spaces and four white spaces. The reason to use three kind of white spaces is to represent a random number of size 16,777,216 (2^{24}) using only 8 bits as shown in Table 4-1. Representing such a huge number with only 8 bits helped me in reducing computation power to embed watermark code in original documents

Random Number Value	White Space Type	Bit Value
n/a	One white space	“0”
$00000 < \text{rand} \leq 000000256$	Two white space	“1”
$00256 < \text{rand} \leq 000065536$	Three white space	“1”
$65536 \leq \text{rand} \leq 16,777,216$	Four white space	“1”

Table 4-1: Random Number, White Space and Bit Value Mapping

A control signal is used to identify start of watermark code. The control signal is two white spaces immediately followed by three white spaces. Considering the length of watermark code and control signal, the minimum number of words in original document is 10, but on the safer side the implantation has restricted the minimum number of words to 15. The pseudocode of this implemented watermark code is shown in Figure 4-2

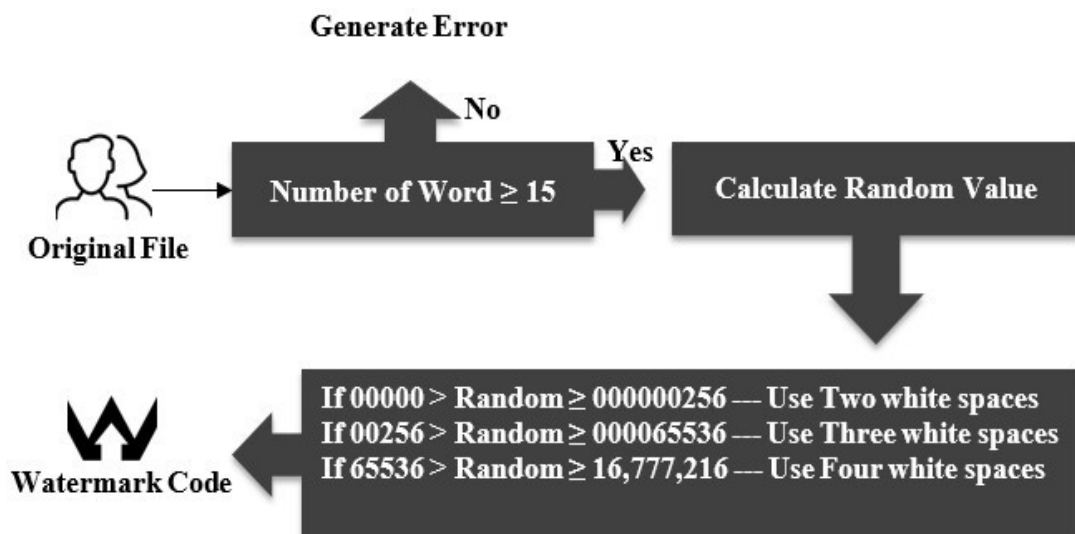


Figure 4-2: Watermark Generator

4.2.2 Watermark Code Embedding

In this module a separate watermarked file is generated for each agent using unique watermark code as shown in figure 4-3

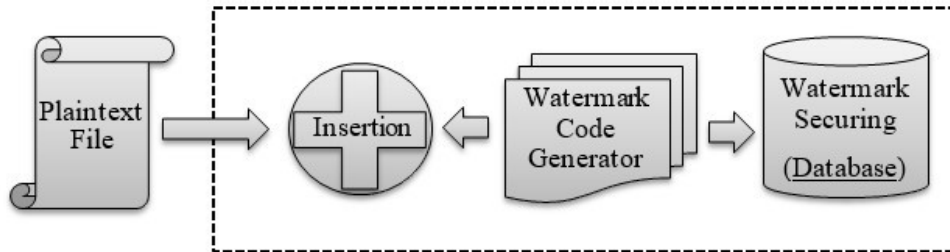


Figure 4-3: Text Watermark Embedding

4.2.3 Data Security Module

In order to ensure privacy of data and non-repudiation of delivery, it very important to encrypt or secure file in a way that watermarked file is only visible to receiver of file. To achieve confidentiality and non-repudiation goals watermarked file is encrypted using public key of receiver of confidential file/shared data using following steps.

4.2.3.1 RSA Key Pair Generation

When a user is registered on application of proposed framework RSA key pair is automatically generated in the background. The public key stores automatically in the database whereas private is exported to predefined location (which should be securely given to respective user/party of the system through any out of band channel, which is out of the scope of this research proposal)

4.2.4 Hash Module

The watermark code that was generated in previous step is hashed using HMAC and saved in a database (Microsoft access in this proposal) along with other metadata like time stamp, agent name, HMAC of “encrypted + watermarked” file etc. The saved hash of the watermark code will be used in Extraction Module to identify guilty agent. Further as shown in Figure 4-4, hash of “encrypted + watermarked” file

using HMAC is also stored in database in order to re-assure NRD claim.

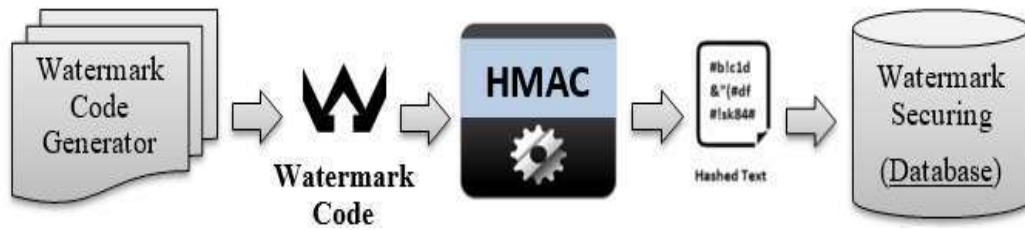


Figure 4-4: Watermark Code Securing

4.3 Extraction Module

Extraction module deals with the extraction of watermark code using a plaintext watermarked file (which is decrypted by an agent and leaked later on) to identify and proof exact name of agent who has received the encrypted file. The agent who possess the private key of respective RSA key pair, it can be easily established that the decrypted watermarked is leaked from no one except the owner of that particular public key as shown in Figure 4-5.

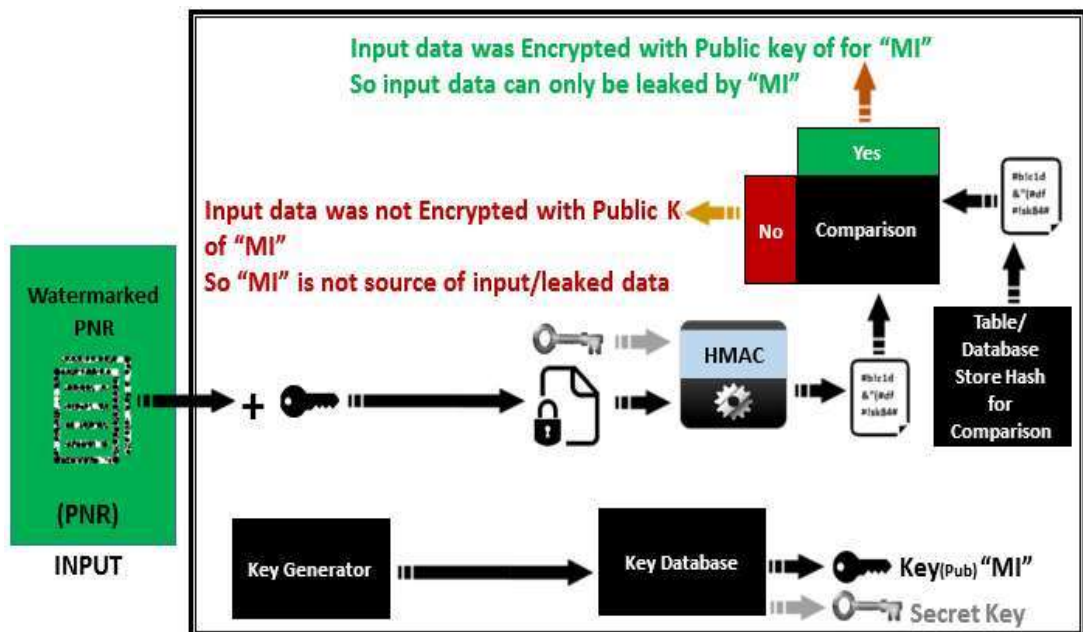


Figure 4-5: Processes involved in Watermark Extraction

The extraction module consists of following sub-modules:

- Watermark Code Extraction
In this module, a watermarked file is parsed or processed to extract embedded watermark code
- Guilt Agent Detection
In this module, hash value of newly extracted watermark code is compared within database to find an exact match and hence guilty agent.

4.3.1 Watermark Code Extraction

In this module, watermarked identity or code is extracted from an input file (watermarked file) as shown in Figure 4-6. The watermarked code is required to proof non-repudiation (non-repudiation of delivery in case of this proposal) in Guilt Agent Detection module. First control signal is detected and then 8-bit text watermark code is generated by inspecting one, two, three or four white spaces.

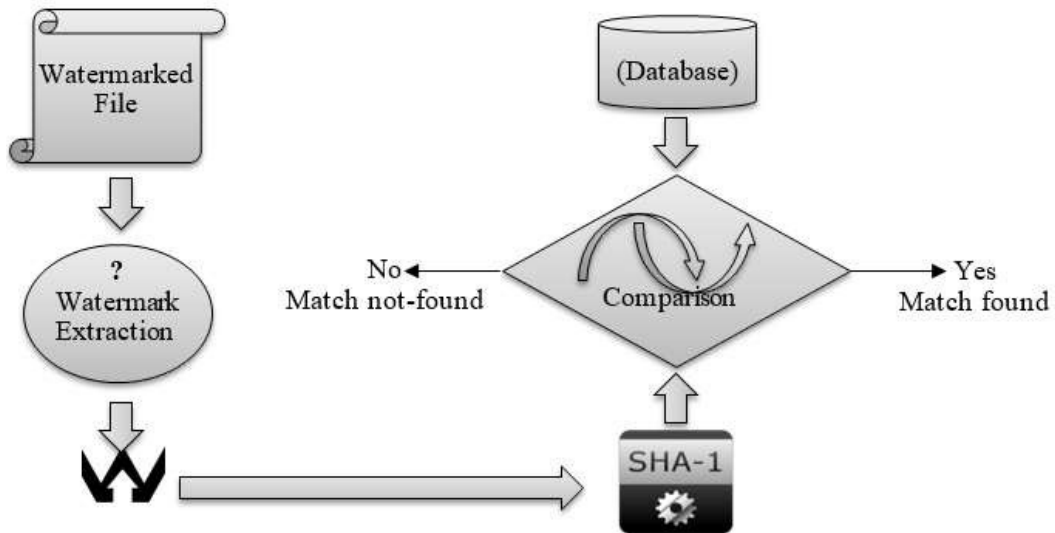


Figure 4-6: Digital Text Watermark Extraction

4.3.2 Guilt Agent Detection

In this module, the extracted watermark code and leaked watermark file are used to find guilty agent using 2 steps:

Step-1: Extracted watermark code is hashed using Hash Module (discussed in section 4.2.4). The hash value of extracted watermark code is then searched in the database to find exact match or the guilty agent.

Step-2: This step is optional but useful to re-assure the non-reputability of delivery. The leaked watermarked file is encrypted using public key of suspected guilty agent and passed to Hash Module to get a hash string. This hash string is compared in database to find an exact match again. Once hash value is matched with previously calculated hash, the non-reputability of delivery is established to fix responsibility of data leakage on guilty agent.

4.4 Proposed Application

For implementation of proposed framework, Java NetBeans platform is used to develop a desktop based application. The application is connected to Microsoft Access for database whereas different tables were used to store user information, credentials, RSA public keys, hash values (watermark code and encrypted watermarked file) etc. as shown in Figure 4-7.

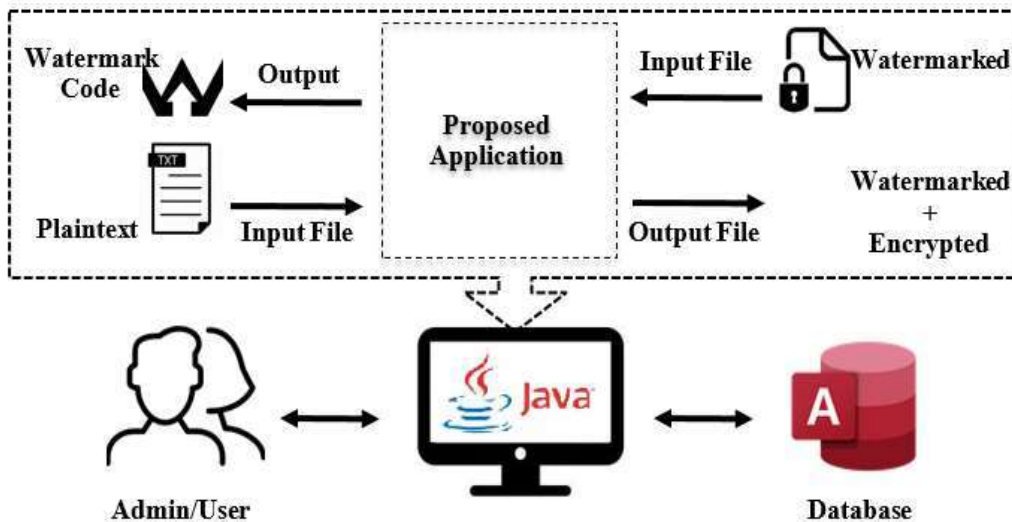


Figure 4-7: Proposed Application

4.4.1 Application Login

The first page that appears after opening the application prompt the use to tell whether he wants to proceed/login to admin or a user as shown in Figure 4-8.

4.4.2 Admin Login

After successful admin login (by entering correct username and password), the administrator shown two Tabs that is whether to proceed for watermark extraction or to manage application users as shown in Figure 4-8.

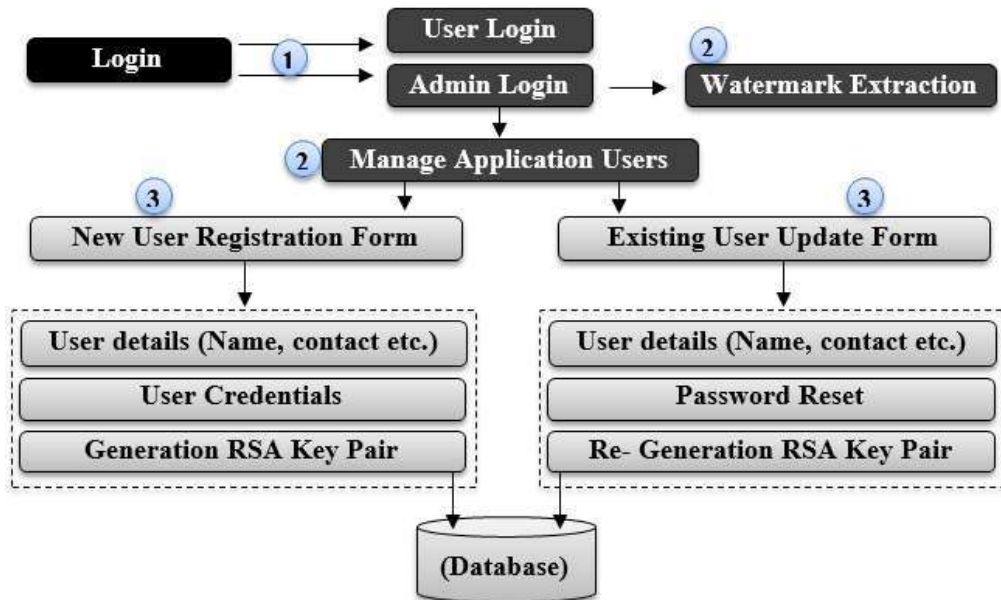


Figure 4-8: User Login/Update Form

4.4.2.1 Manage Application Users

By selecting Manage Application User Tab the administrator can create new users or update existing users as shown in Table 4-2.

	New User Registration Form	Existing User Update Form
Name	Alphabets and numeric only, maximum length = 40	ASCHI, Alphabets and numeric only, maximum length = 40

CHAPTER 4. IMPLEMENTATION

Username	Alphabets and numeric only, maximum length = 10	--
Password	Enter Initial Password Alphabets and numeric only, maximum length = 10	Reset Existing Password Alphabets and numeric only, maximum length = 10
Re-Confirm Password	Alphabets and numeric only, maximum length = 10	Alphabets and numeric only, maximum length = 10
Country	Alphabets and numeric only, maximum length = 20	Alphabets and numeric only, maximum length = 20
Organization	Alphabets and numeric only, maximum length = 40	Alphabets and numeric only, maximum length = 40
Contact	Numeric only, maximum length = 12	Numeric only, maximum length = 12

Table 4-2: Manage Application User Forms

RSA private key should be shared to respective user through a confidential and secure out of band channel (not covered in the scope of the proposed framework).

4.4.2.2 Watermark Extraction

In this page, the administrator can upload a leaked file (a watermarked word file processed via the proposed application) and identify the recipient of this file. The detailed working of Watermark Extraction Module is discussed in section (4.2.2).

4.4.2.3 Admin Login

After successful login, the user is shown options to upload a file for watermark embedding that is supposed to be shared with any other user of the system. After uploading the file the user can generate watermarked file by selecting “Watermark Embedding Tab” as shown in Figure 4-9.

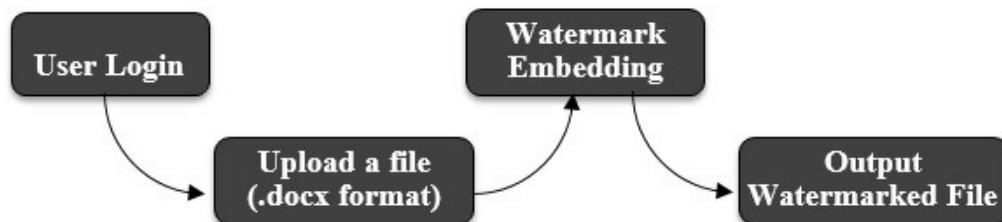


Figure 4-9: User Login

CHAPTER 4. IMPLEMENTATION

There are many hidden processes in Watermark Embedding module that is watermark code generation, hashing, and public key encryption etc., which are used to ensure confidentiality, integrity and non-repudiation of delivery (detail available in section 4.2.1). After watermark embedding, the encrypted watermarked file is saved to a specific location so that output file can be shared separately to recipient using any secure out of band channel (out of scope of this research proposal).

4.5 Summary

In this chapter, the complete flow and implementation concepts of the proposed framework are explained. The proposed framework in this research paper is an improved form of exiting Data Leakage Detection (DLD) algorithms discussed in Chapter-2. Watermark Code Generator module is similar to Fake Object Module in a sense that some attributes of original data set are changed. However, here using Open Space text watermarking, the meaning of original message does not affected at all. Watermark Code Embedding Module is similar to Data Allocation Module as both add secret code to original data set for a target agent. Similarly, Guilt Agent Detection is same in both as both are trying to recover the embedded code that watermark code and fake object respectively to identify guilty agents. The issues observed in both existing DLD algorithms are:

- The use of fake objects adds a noise in original data set, which can alter the message of original data e.g. medical reports, financial statements; passenger name records (PNRT etc.
- Non-Repudiation of Delivery (NRD) cannot be established as original data set in plane text and in embedded form (original data + fake object) is available at distributor side as well

The improvement in this work is use of public key encryption in a controlled manner ensures that watermarked data for an agent is not accessible to originator/distributor at any point of time except leaked by an agent in decrypted form.

Chapter 5

Conclusion and Future Work

This chapter concludes the presented thesis and highlights potential future research directions. It describes different research prospects of the research and identifies open research problems that could be explored further by researchers and system designers.

5.1. Conclusion

The proposed framework provides a secure and traceable solution that enables the investigators to identify a guilty agent (source of a data breach) with probability $Pr \leq 1$ by satisfying the property of the Minimum Overlap Model either. Traceable means that the ownership of any document, which is being shared with other parties of the system can be tracked easily. The proposed system holds good for both cases when the number of receiving agents of a file is ≥ 1 (1 or n).

As the probability $Pr \leq 1$, the proposed framework can provide a claim of Non-Repudiation of Delivery (NRD) which was not guaranteed by existing models of data leakage detection. To achieve NRD, the Text Watermarking technique is used. Watermarking is an information hiding technique, which is commonly used for audio, video, image and text media to provide copyright protection and detection use cases. As a watermarked file is embedded with a unique watermark code for the originator, the ownership of a file can be easily identified by extracting the unique code. The proposed framework embeds unique watermark codes for all the recipient parties of the system, it enables the system to identify the exact destination of the file (recipient party) by extracting unique code from a watermarked file. As the watermarked file

was processed at the originator party, so in order to establish a complete NRD check, the proposed framework output watermarked file after encryption with public keys of the recipient party. Encrypting Watermarked files with the public key ensures that the watermarked file in plaintext form is only visible to the recipient party. In addition to the NRD claim, encryption with a public key also ensures the confidentiality of shared files in transit.

Lastly, to further secure the system, the hash of the unique watermark codes and encrypted watermarked files are stored (using HMAC) in a database in order to reassure the proof NRD in the reverse watermark extraction process as mentioned in section 4.5. Hash of Watermark Code and encrypted file can also be used to verify the integrity of the shared file.

5.2. Future Work

Though the solution solves the core problem i.e. to identify the exact source of a leaked file, which was shared with multiple parties, the proposed application framework, is limited to work electronic word documents (Microsoft .docx format in English language). Similarly, watermark embedding module is implemented and open space technique of text watermarking is used. As mentioned in Section 1.3 of this thesis, the development of a text watermark generator is not the objective of this research work, so if any researcher wants to extend the scope of the proposed framework i.e. to support all kind of documents like printed, scanned/image, electronic etc., all kind of Languages i.e. English, Arabic, Urdu etc. or any format i.e. PDF, excel, CSV etc. the researcher needs to improve text watermark embedding and extracting modules. Natural Language Processing (NLP) technique of Artificial Intelligence can be used to support all kind of documents i.e. printed, scanned, electronic documents of all kind of formats. For example, NLP can be used to recovering and digitizing printed or image based documents to OCR formats [34]. Advanced text watermarking techniques can be used to protect embedded watermark code from different attacks i.e. reformatting, re-ordering or re-phrasing attacks [24]. Similarly, the combination of different watermarking techniques or modules can be used in the framework to support all kinds of documents simultaneously

Bibliography

- [1] K. Arlitsch, A. Edelman, C. Editor, and B. Kenning Arlitsch, “Staying Safe: Cyber Security for People and Organizations,” *J. Libr. Adm.*, vol. 54, no. 1, pp. 46–56, 2014.
- [2] S. Quinton and N. Reynolds, “Characteristics of Digital Data,” *Underst. Res. Digit. Age*, pp. 57–88, Jan. 2020.
- [3] M. T. Ahvanooy, Q. Li, H. J. Shim, and Y. Huang, “A Comparative Analysis of Information Hiding Techniques for Copyright Protection of Text Documents,” vol. 2018, 2018.
- [4] A. Alnemari, R. K. Raj, C. J. Romanowski, and S. Mishra, “Protecting Personally Identifiable Information (PII) in Critical Infrastructure Data Using Differential Privacy,” *2019 IEEE Int. Symp. Technol. Homel. Secur. HST 2019*, pp. 6–11, 2019.
- [5] M. Panjwani and M. Jantti, “Data Protection & Security Challenges in Digital & IT Services,” *2017 Int. Conf. Comput. Appl. ICCA 2017*, pp. 379–383, 2017.
- [6] M. A. Sahi *et al.*, “Privacy Preservation in E-Healthcare Environments: State of the Art and Future Directions,” *IEEE Access*, vol. 6, pp. 464–478, 2017.
- [7] JEFF PETERS, “Data Privacy Guide: Definitions, Explanations and Legislation | Varonis.” <https://www.varonis.com/blog/Data-Privacy/>.
- [8] A. Pala and J. Zhuang, “Information Sharing in Cybersecurity: A review,” *Decis. Anal.*, vol. 16, no. 3, pp. 172–196, 2019.
- [9] E. Schiavone, A. Ceccarelli, and A. Bondavalli, “Continuous Authentication and Non-Repudiation for the Security of Critical Systems,” *Proc. IEEE Symp. Reliab. Distrib. Syst.*, pp. 207–208, 2016.

- [10] B. Awojobi and J. Ding, "Data Security and Privacy," *Cybersecurity Inf. Prof.*, pp. 291–304, 2020.
- [11] O. G. Abood and S. K. Guirguis, "A Survey on Cryptography Algorithms," *Int. J. Sci. Res. Publ.*, vol. 8, no. 7, Jul. 2018.
- [12] H. R. Pial, "Digital Signature Understanding. How it Works and Importance," no. October, 2020, [Online]. Available: https://www.researchgate.net/publication/344818547_Digital_Signature_Understanding_How_it_Works_and_Importance/citation/download.
- [13] Z. Jalil and A. M. Mirza, "A review of digital watermarking techniques for text documents," in *2009 International Conference on Information and Multimedia Technology, ICIMT 2009*, 2009, pp. 230–234, doi: 10.1109/ICIMT.2009.11.
- [14] S. Dommala and M. Sreedevi, "Rajat Verma_2020 Survey on DLD," vol. 1, no. 9, pp. 42–46, 2012.
- [15] O. Ghaisas and Y. Borse, "Survey of Data Protection Mechanisms to Protect Data at Rest on Cloud," *Int. J. Eng. Trends Technol.*, vol. 59, no. 2, pp. 105–107, May 2018.
- [16] X. Zhang, D. P. van Donk, and T. van der Vaart, "Does ICT Influence Supply Chain Management and Performance? A Review of Survey-based Research," *Int. J. Oper. Prod. Manag.*, vol. 31, no. 11–12, pp. 1215–1247, 2011, [Online]. Available: <https://research.rug.nl/en/publications/does-ict-influence-supply-chain-management-and-performance-a-revi>.
- [17] M. C. J. Caniels, H. K. L. Lenaerts, and C. J. Gelderman, "Explaining the Internet Usage of SMEs The Impact of Market Orientation, Behavioural Norms, Motivation and Technology Acceptance," *Internet Res.*, vol. 25, no. 3, pp. 358–377, 2015, [Online]. Available: <https://research.ou.nl/en/publications/explaining-the-internet-usage-of-smes-the-impact-of-market-orient>.
- [18] L. Cheng, F. Liu, and D. D. Yao, "Enterprise Data Breach: Causes, Challenges, Prevention, and Future Directions," *Wiley Interdiscip. Rev. Data Min. Knowl.*

- Discov.*, vol. 7, no. 5, 2017.
- [19] L. J. Trautman and P. C. Ormerod, “CORPORATE DIRECTORS’ AND OFFICERS’ CYBERSECURITY STANDARD OF CARE: THE YAHOO DATA BREACH,” *Am. Univ. Law Rev.*, vol. 66, [Online]. Available: <https://www.nytimes.com/2016/09/23/technology/yahoo->.
- [20] Sara Heath, “Majority of 2015 Healthcare Data Breaches Due to IT Hacking.” <https://healthitsecurity.com/news/Majority-of-2015-Healthcare-Data-Breaches-Due-to-IT-Hacking>.
- [21] N. Lord, “Top 10 Biggest Healthcare Data Breaches of All Time | Digital Guardian,” *Digital Guardian*, 2018. <https://digitalguardian.com/Blog/Top-10-Biggest-Healthcare-Data-Breaches-All-Time>.
- [22] Anita Balakrishnan and Deirdre Bosa, “Uber Hack Exposes Data of 57 Million Users and Drivers, Report Says.” <https://www.cnbc.com/2017/11/21/uber-hack-exposes-data-of-57-million-users-and-drivers-report-says.html>.
- [23] Megan Leonhardt, “The 5 Biggest Data Hacks of 2019.” <https://www.cnbc.com/2019/12/17/the-5-Biggest-Data-Hacks-of-2019.html>.
- [24] EUGENE BEKKER, “2020 Data Breaches - The Most Significant Breaches of the Year.” <https://www.identityforce.com/blog/2020-Data-Breaches>.
- [25] B. Jankowski, W. Mazurczyk, and K. Szczypiorski, “Information Hiding Using Improper Frame Padding,” *Proc. 2010 14th Int. Telecommun. Netw. Strateg. Plan. Symp. Networks*, 2010.
- [26] L. Cheng, F. Liu, and D. D. Yao, “Enterprise data breach: causes, challenges, prevention, and future directions,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 7, no. 5, Sep. 2017, doi: 10.1002/WIDM.1211.
- [27] T. Lakshmi Siva Rama Krishna, P. Bandavi, K. Priyanka, and V. P. Vivek, “Novel User Level Data Leakage Detection Algorithm,” *NULDLDA_2019 Nov. User Lev. Data Leakage Detect. Algorithm*, vol. 8, no. 10, pp. 2378–2381, 2019.
- [28] JEAN-F. TOMB *et al.*, “Enhanced Reader.pdf,” *Nature*, vol. 388. pp. 539–547,

- 1997.
- [29] C. Igwenagu, “Fundamentals of Research Methodology and Data Collection,” *L. Lambert Acad. Publ.*, no. June, p. 4, 2016, [Online]. Available: https://www.researchgate.net/publication/303381524_Fundamentals_of_Research_Methodology_and_Data_Collection.
- [30] “BA GDPR Data Breach Fine Lowered to £20m Due to COVID-19 - Infosecurity Magazine.” <https://www.infosecurity-magazine.com/News/BA-GDPR-Fine-20/>.
- [31] “Cathay Pacific Breach | Information Security Buzz.” <https://informationsecuritybuzz.com/Expert-Comments/Cathay-Pacific-Breach/>.
- [32] “Data Protection Act 2018 - Wikipedia.” https://en.wikipedia.org/wiki/Data_Protection_Act_2018.
- [33] “ICAO Traveller Identification Programme Symposium 2021 (TRIP2021) and First Joint ICAO/INTERPOL Passenger Data Exchange Forum.” <https://www.icao.int/Meetings/TRIP-Symposium-2021/Pages/default.aspx>.
- [34] A. Sivaprasad and S. Jangale, “A Complete Study on Tools & Techniques for Digital Forensic Analysis,” in *2012 International Conference on Computing, Electronics and Electrical Technologies, ICCEET 2012*, 2012, pp. 881–886.
- [35] P. SDeshmukh and P. Pande, “International Journal of Computer Science and Mobile Computing A Study of Electronic Document Security,” 2014.
- [36] “Data Protection and Privacy: Definitions, Differences, and Best Practices| Cloudian.” <https://cloudian.com/Guides/Data-Protection/Data-Protection-and-Privacy-7-Ways-to-Protect-User-Data/>.
- [37] F. Meneses *et al.*, “RSA Encryption Algorithm Optimization to Improve Performance and Security Level of Network Messages,” 2016.
- [38] A. Putera Utama Siahaan, E. Elviwani, and B. Oktaviana, “Comparative Analysis of RSA and ElGamal Cryptographic Public-key Algorithms,” Jul. 2018.

- [39] J. Spacey, “What is Data Authentication? - Simplifiable.com,” Dec. 21, 2016. .
- [40] KWANGSOO SEOL, YOUNG-GAB KIM, and EUIJONG LEE, “Privacy Preserving Attribute Based Access Control Model for XML Based Electronic Health Record System,” 2018. .
- [41] W. Wu, J. Zhou, and Y. Xiang, “How to Achieve Non-Repudiation of Origin with Privacy Protection in Cloud Computing,” *Volume 79, Issue 8, December 2013*. .
- [42] D. Graft, M. Pabrai, and U. Pabrai, “Methodology for Network Security Design,” in *Conference Proceedings - Annual Phoenix Conference*, 1990, pp. 675–682.
- [43] E. J. Delp, “Multimedia Security,” pp. 116–116, 2004.
- [44] C. N. Yang, C. C. Lin, and C. C. Chang, “Steganography and Watermarking,” vol. 7, no. 6, pp. 1–412, 2013.
- [45] T. Reports, “Digital Text Watermarking Techniques Classification and Open Research Challenges : A Review,” no. June, 2020.
- [46] I. Stojanov, A. Mileva, and I. S. Stojanovi'c, *A New Property Coding in Text Steganography of Microsoft Word Documents*. .
- [47] A. K. Singh, I. Gupta, R. Verma, V. Gautam, and C. P. Yadav, “A Survey on Data Leakage Detection and Prevention,” *SSRN Electron. J.*, no. Icdam, pp. 1–7, 2020.
- [48] O. Jane, E. Elbaşı, and H. G. İlk, “Hybrid Non-Blind Watermarking based on DWT and SVD,” *J. Appl. Res. Technol.*, vol. 12, no. 4, pp. 750–761, Aug. 2014, doi: 10.1016/S1665-6423(14)70091-4.
- [49] C.-C. Chang, Y.-C. Chou, and T.-C. Lu, “A Semi-blind Watermarking Based on Discrete Wavelet Transform.”