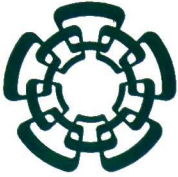


UT-T00068-001

Don.: 2.014



CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS
DEL INSTITUTO POLITÉCNICO NACIONAL

Laboratorio de Tecnologías de Información,
CINVESTAV-Tamaulipas

**Un Sistema Seguro para el
Almacenamiento y Distribución
de Documentos Digitales con
Servicio de Rastreo de Usuarios
Deshonestos**

Tesis que presenta:

Mario Diego Muñoz Hernández

Para obtener el grado de:

**Maestro en Ciencias
en Computación**

Dr. Miguel Morales Sandoval
Dr. José Juan García Hernández

Cd. Victoria, Tamaulipas, México.

Diciembre, 2013

**CINVESTAV
IPN
ADQUISICION
LIBROS**

CLASIF..	VT 00068
ADQUIS..	VT-700068-SSI
FECHA:	22-10-2014
PROCED..	Don. - 2014
\$	

10:216346-1001



RESEARCH CENTER FOR ADVANCED STUDY
FROM THE NATIONAL POLYTECHNIC INSTITUTE

Information Technology Laboratory,
CINVESTAV-Tamaulipas

**A Secure System for Storage and
Distribution of Digital Documents
with Dishonest User Tracing
Service**

Thesis by:

Mario Diego Muñoz Hernández

as the fulfillment of the
requirement for the degree of:

**Master of Science
in Computer Science**

Thesis Directors:

Dr. Miguel Morales Sandoval
Dr. José Juan García Hernández

Cd. Victoria, Tamaulipas, México.

December, 2013

© Copyright by
Mario Diego Muñoz Hernández
2013

The thesis of Mario Diego Muñoz Hernández is approved by:

Dr. Víctor Jesús Sosa Sosa

Dr. Arturo Díaz Pérez

Dr. Miguel Morales Sandoval, Co-Director

Dr. José Juan García Hernández, Director

Cd. Victoria, Tamaulipas, México., December 9 2013

To my mother and sister

Acknowledgements

- Let me begin by thanking my family, for its support and sacrifices, for the efforts of my mother Bertha Guadalupe for giving to me education and formation, and the comprehension of my sister Marisol during the time of hard work. This is our achievement.
- To my thesis director Dr. José Juan García and my co-director Dr. Miguel Morales, for their patience and guidance. I learned science and human quality from them.
- To my friends, for sharing with me this travel of two years, for believing with me that this was possible.
- To all professors I had in CINVESTAV. For their love to share knowledge. It was impossible to talk to any of them and do not learn something new.
- To my generation partners, who provided to me their feedback and comments.
- To SVAM International and its directives Anil Kapoor, Bob Hart and Othon Rodriguez, who have supported my professional growth. Also, I thank Yessica Heredia from Human resources and Sergio Ponce and Calixto Conde my project leaders during this two years for providing to me the time to achieve this goal.
- To CINVESTAV and CONACyT for supporting this work under grant CB-2010-1-50910.
- Finally, my greatest gratitude to God, without Him nothing of this would be possible. I put my trust in You, and I have not been defrauded.

Contents

Contents	i
List of Figures	v
List of Tables	ix
Publications	xi
Resumen	xiii
Abstract	xv
Nomenclature	xvii
1 Introduction	1
1.1 Context and Motivation	1
1.2 Problem Description	4
1.2.1 Illegal Access to Digital Documents from Unauthorized Users	5
1.2.2 Illegal Distribution of Digital Documents by Authorized but Dishonest Users	6
1.3 Investigation Question	6
1.4 Objectives	7
1.4.1 General Objective	7
1.4.2 Specific Objectives	7
1.5 Thesis Outline	8
2 Background of Security Information Services for Digital Documents	9
2.1 Confidentiality	9
2.1.1 Symmetric Key Cryptography	10
2.1.2 Asymmetric Key Cryptography	13
2.1.3 Digital Envelope	14
2.1.4 Key Exchange	15
2.1.5 Digital Image Encryption	15
2.2 Integrity	16
2.2.1 One Way Hash Functions	16
2.2.2 Message Authentication Codes	17
2.3 Authentication and Non-Repudiation	17
2.3.1 Digital Signatures	18
2.3.2 Public Key Infrastructure	19
2.3.3 Mutual Authentication	19
2.4 Access Control	20

2.4.1	Discretionary Access Control	21
2.4.2	Role Based Access Control	21
2.4.3	Mandatory Access Control	21
2.5	Digital Document Management Security Architectures	21
2.5.1	Encrypted Server-based Document Management System	22
2.5.2	Terminal-Based Digital Document Management System	22
2.5.3	Hardware-Based Encryption Document Management System	23
2.5.4	Cloud Computing Environment-Based Document Management System	24
2.6	Summary	24
3	Background of Fingerprinting Techniques for Digital Documents	27
3.1	Digital Watermarking	27
3.1.1	Watermarking Classification	28
3.1.2	Watermarking Applications	29
3.1.3	Watermarking Attacks	30
3.1.4	Watermarking Techniques	31
3.1.4.1	Least Significant Bit Substitution	31
3.1.4.2	Patchwork	32
3.1.4.3	Quantization Index Modulation	33
3.1.4.4	Spread Spectrum	34
3.1.5	Analysis of Watermarking Techniques for Digital Images	35
3.1.6	Digital Text Watermarking	36
3.1.6.1	Syntactic Approach	36
3.1.6.2	Semantic Approach	37
3.1.6.3	Image-Based Approach	37
3.2	Traitor Tracing	39
3.2.1	Digital Fingerprinting	40
3.3	Image Quality Assessment	41
3.3.1	Peak Signal-to-Noise Ratio	41
3.3.2	Structural Similarity Index	41
3.4	Discrete Cosine Transform	42
3.5	Summary	44
4	Related Work on Security and User Tracing for Digital Documents	45
4.1	Secure Digital Document Management Systems	45
4.2	User Tracing for Digital Documents	53
4.3	Analysis	57
4.4	Summary	58
5	Proposed Secure Document Management System	59
5.1	Information Security Services Module	59
5.1.1	Proposed Strategy	61
5.1.1.1	Proposed Strategy for Secure Transmission of Digital Documents	61

5.1.1.2	Proposed Strategy for Secure Storage of Digital Documents	62
5.1.2	Access Control	62
5.1.3	Public Key Infrastructure	63
5.1.3.1	Digital Certificate Generation	64
5.1.3.2	Certificates Revocation	64
5.1.3.3	User Authentication	64
5.1.4	Secure Digital Documents Approval	64
5.1.5	Secure Digital Documents Storage	66
5.1.6	Secure Digital Documents Distribution	66
5.1.7	Selection of Cryptographic Algorithms	70
5.1.8	Implementation	71
5.1.8.1	Cryptographic Object Storage	71
5.1.8.2	Secure Sockets Layer	71
5.1.8.3	Software Components	72
5.2	Fingerprinting Module	73
5.2.1	Fingerprinting Technique Selection	74
5.2.2	Fingerprint Generation	74
5.2.3	Fingerprint Insertion	75
5.2.4	Fingerprint Detection	77
5.2.5	Software Implementation	78
5.3	Integration	79
5.3.1	General System	79
5.3.2	User Identity	81
5.3.3	Secure Fingerprint Insertion	82
5.3.4	Secure Fingerprint Detection	82
5.4	Summary	84
6	Results of the Proposed Secure Document Management System	85
6.1	Information Security Services Module	85
6.1.1	Validation	86
6.1.2	Performance Evaluation	87
6.1.2.1	Digital Documents Server Operations	87
6.1.2.2	Digital Documents Transmission	88
6.1.3	Comparison	88
6.2	Fingerprinting Module	90
6.2.1	Perceptual Transparency	90
6.2.1.1	Sets of Parameters Evaluation	91
6.2.1.2	Inquest Evaluation	94
6.2.2	Collusion Resistance	96
6.2.2.1	Robustness Factors	97
6.2.2.2	Fingerprint Length	98
6.2.2.3	Attack on Digital Documents in Lossy Format	102
6.2.3	Performance Evaluation	102

6.2.4	Comparison	104
6.3	Integrated System	106
6.3.1	Operations Validation	106
6.3.2	Performance Evaluation	107
6.3.2.1	Download Digital Documents	107
6.3.2.2	Fingerprint Detection	108
6.4	Summary	109
7	Conclusions and Future Work	111
7.1	Main Contributions	113
7.2	Future Work	114

List of Figures

1.1	Lifecycle for Document Imaging.	2
1.2	Type of users in a Document Management System.	3
1.3	Use cases for Reviewer, Consumer, Auditor and Administrator users in a Document Management System.	4
2.1	Private key cryptography scheme.	11
2.2	Public key cryptography scheme.	13
2.3	Digital envelope scheme.	15
2.4	Hashing process.	16
2.5	Generation and use of a Message Authentication Code for a digital document.	18
2.6	Digital signature scheme.	19
2.7	General scheme of mutual authentication.	20
2.8	Encrypted server-based document management system architecture.	22
2.9	Terminal-based digital document management system scheme.	23
2.10	Hardware-based encryption digital document management system scheme.	24
2.11	Cloud computing environment-based document scheme.	25
3.1	General scheme of digital watermarking.	28
3.2	Diagram of watermarking classification.	28
3.3	Least Significant Bit substitution.	31
3.4	General scheme of syntactic text watermarking.	37
4.1	Scheme of the Secure Automated Document Delivery System [1].	46
4.2	Document management system scheme proposed by Liouy <i>et al.</i> [2]	47
4.3	eSign Architecture proposed by Shi and Ouyang [3].	49
4.4	Model for a Secure Document Management System with cloud storage proposed by Kamara and Lauter [4]	51
4.5	Architecture of an enterprise content management system for secure contracts proposed by Chieu <i>et al.</i> [5]	52
4.6	Scheme of the encoder proposed by Brassil <i>et al.</i> in [6].	54
4.7	Scheme of the decoder proposed by Brassil <i>et al.</i> in [6].	54
4.8	Scheme of the encoder and decoder proposed by Darwish in [7].	56
4.9	Accumulation of user IDs that allow tracing back a traitor according to the work of Schick and Ruland [8].	57
5.1	Selected techniques to provide the required information security services in the proposed Secure Document Management System	60
5.2	Topology for the proposed Secure Digital Document Management System.	63
5.3	Digital Certificate generation process.	65

5.4	Sequence diagram of the Approval stage in the proposed Secure Document Management System.	67
5.5	Sequence diagram of the signature verification process of digital document in the proposed Secure Document Management System.	68
5.6	Sequence diagram of the Distribution stage in the proposed Secure Document Management System.	69
5.7	Upload and download scheme of the SDMS implementation.	72
5.8	Model of the software implementation in the SDMS server with the submodules of the Information Security Service Module.	73
5.9	Diagram of fingerprint insertion method for the proposed fingerprinting module.	76
5.10	Diagram of fingerprint detection method for the proposed Fingerprinting Module.	79
5.11	High level view of the SDMS server integrating Information Security Services and User Tracing.	80
5.12	Certificate - User ID mapping process.	81
5.13	Secure Fingerprint Insertion process.	82
5.14	Generation of the detection results file after the fingerprint detection.	83
5.15	Mapping user IDs to digital certificates.	83
6.1	Distortion (PSNR) generated in digital documents due to fingerprint insertion varying the insertion position (P_w), with fixed values of robustness factor assigned to users (β_u), groups(β_g) and fingerprint length (L). Low PSNR values implies high distortion.	92
6.2	Similarity (SSIM Index) between an original document and its fingerprinted copy varying the insertion position (P_w) with fixed values of the robustness factor assigned to users (β_u), groups(β_g) and fingerprint length (L). High SSIM Index values implies a high similarity.	92
6.3	Distortion (PSNR) generated in digital documents due to fingerprint insertion varying the robustness factor for users (β_u) and groups(β_g), with fingerprint length (L) and insertion position (P_w) fixed.	93
6.4	Similarity (SSIM Index) between an original document and its fingerprinted copy varying the robustness factor for users (β_u) and groups(β_g), with fingerprint length (L) and insertion position (P_w) fixed.	94
6.5	Distortion of digital documents due to fingerprint insertion.	95
6.6	Perception of 100 respondents about fingerprinted digital documents with different distortion levels (PSNR).	97
6.7	Detection of 2 colluded users with IDs 300 and 600, with $B_u = 200,000$, $B_g = 50,000$, $P_w = 1/6$ and $L = 350,000$. The horizontal line defines the threshold value.	99
6.8	Number of detected colluders after a collusion attack for fingerprints generated with configurations 1, 2, 3, 4 and 5 from Table 6.7 and $P_w = 1/6$	99
6.9	Ratio of detected colluders after a collusion attack for fingerprints generated with configurations 1, 2, 3, 4 and 5 from Table 6.7 and $P_w = 1/6$	100
6.10	Number of detected colluders after a collusion attack for fingerprints generated with $B_g = 50,000$, $B_u = 200,000$ and $P_w = 1/6$	101

6.11 Probability of detected colluders after a collusion attack for fingerprints generated with $B_g = 50,000$, $B_u = 200,000$ and $P_w = 1/6$ 101

6.12 Amount of detected colluders, after a collusion attack with lossy digital documents for fingerprints generated with $B_g = 50,000$, $B_u = 200,000$, $L = 350,000$ and $P_w = 1/6$ 103

6.13 Average time of fingerprint insertion, decryption and secure trasmission, required to download a digital document. 108

List of Tables

3.1	Advantages and disadvantages of watermarking techniques.	35
5.1	Relation of user types and functions in the proposed Secure Document Management System.	60
6.1	Tests performed for validation of the Encryption and Signature modules.	86
6.2	Characteristics of digital documents used to evaluate the Information Security Service Module.	87
6.3	Comparison of the proposed Information Security Services Module against representative works in the literature. Information Security Services: Confidentiality (C), Integrity (I), Authenticity (A), Access control (X), and Non-Repudiation (N).	89
6.4	Characteristics of digital documents used to evaluate the Fingerprinting Module.	90
6.5	Configuration parameters for fingerprint insertion.	91
6.6	Inquest results of perceptual transparency of fingerprinted digital documents at different levels of distortion (PSNR) and similarity (SSIM Index).	96
6.7	Configurations of β_u , β_g and L that satisfy perceptual transparency for a fixed P_w	98
6.8	Performance evaluation of insertion and detection of fingerprints.	104
6.9	Comparison of the Fingerprinting Module in this thesis against other insertion techniques in the literature applicable to fingerprinting.	105
6.10	Test performed to validate the correct functioning of the system integrating the ISS and fingerprinting modules.	106

Publications

Munoz-Hernandez M. D., Garcia-Hernandez J. J., Morales-Sandoval M. and Larranaga-Cepeda A.(2013). Study on the Impact of Fingerprints on the Perceptual Transparency in Digital Documents. To appear in the Proceedings of 4th European Conference of Computer Science (ECC'13), October 29-31, Paris, France. 2013

Munoz-Hernandez M. D., Garcia-Hernandez J. J., and Morales-Sandoval M. (2013). A Collusion-Resistant Fingerprinting System for Restricted Distribution of Digital Documents. To appear in the *PLoS ONE* Journal, october 2013.

Un Sistema Seguro para el Almacenamiento y Distribución de Documentos Digitales con Servicio de Rastreo de Usuarios Dishonestos

por

Mario Diego Muñoz Hernández

Laboratorio de Tecnologías de Información, CINVESTAV-Tamaulipas

Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, 2013

Dr. José Juan García Hernández, Director

Dr. Miguel Morales Sandoval, Co-Director

Con el incremento en la capacidad de los medios de almacenamiento a finales de los 80, gran cantidad de organizaciones y empresas iniciaron una migración de sus documentos físicos a formatos digitales. Hoy en día, uno de los principales activos de las organizaciones son sus documentos digitales, los cuales suelen contener información sensible que solo debe estar disponible para usuarios autorizados. Estos documentos deben ser protegidos para evitar el acceso ilícito por parte de usuarios no autorizados.

Se han desarrollado técnicas criptográficas que brindan servicios de confidencialidad y control de acceso para restringir el acceso a los documentos digitales solo a usuarios autorizados. Sin embargo, cuando un usuario legítimo pero deshonesto obtiene un documento digital en claro, está en la capacidad de distribuirlo ilícitamente a usuarios no autorizados. Es necesario determinar la identidad los usuarios deshonestos para detener la distribución ilegal de documentos digitales lo cual se puede lograr empleando técnicas para el rastreo de usuarios en documentos digitales. El rastreo de usuarios en documentos digitales está siendo incipientemente explorado y actualmente no se ha reportado un sistema para la administración de documentos digitales que emplee esta técnica.

En este trabajo de tesis se diseña e implementa un sistema seguro para la administración de documentos digitales para enfrentar el problema del acceso a documentos digitales por parte de usuarios no autorizados y la distribución ilícita de documentos digitales por part de usuarios

autorizados pero deshonestos. La solución para los problemas abordados en esta tesis hacen uso de técnicas criptográficas y técnicas de control de acceso para proveer seguridad a documentos digitales a través de los servicios de confidencialidad, integridad, no repudio, autenticación y control de acceso. Además, se emplean técnicas de fingerprinting resistentes a colusión reportadas en el estado del arte para proveer un servicio de rastreo de usuarios.

Se reportan los resultados del sistema propuesto, concluyendo que es capaz de proteger documentos digitales de forma efectiva y además es capaz de rastrear usuarios deshonestos, brindando una solución integral para el acceso ilícito y la distribución ilegal de documentos digitales.

A Secure System for Storage and Distribution of Digital Documents with Dishonest User Tracing Service

by

Mario Diego Muñoz Hernández

Information Technology Laboratory, CINVESTAV-Tamaulipas
Research Center for Advanced Study from the National Polytechnic Institute, 2013
Dr. José Juan García Hernández, Advisor
Dr. Miguel Morales Sandoval, Co-advisor

With the increasing in the capacity of storage media in the late 80s, many organizations have migrated their physical documents to digital formats. Nowadays, digital documents contain most of the information of organizations and they are used in a wide variety of processes in which it is necessary to access their content. Due to the private nature of some digital documents, they must be protected from illicit access of unauthorized users. In order to protect digital documents, information security services are required. These services can be provided through document management systems that protect them from threats such as access or alteration by unauthorized users. However, when authorized users obtain a digital document they could act dishonestly, by distributing digital documents to unauthorized users. It is necessary to determine the identity of dishonest users to stop the illegal distribution of digital documents. To achieve this, user tracing techniques for digital documents can be used to determine the identity of dishonest users. User tracing techniques for digital documents are incipiently explored and currently it has not been reported that document management systems provide this service.

In this thesis work, a secure document management system is designed and developed to face the illegal access to digital documents from unauthorized users and their illegal distribution by authorized but dishonest users. The solution provided in this thesis to the tackled problem makes use of cryptographic techniques and access control mechanisms to provide security to digital documents, through the provision of information security services of confidentiality, integrity non-repudiation,

authentication and access control. Also, state-of-the-art collusion-resistant fingerprinting techniques are used to provide the user tracing service.

Results of the system are provided, concluding that the system is able to protect digital documents from unauthorized users in an effective way, and also it is able to trace back dishonest users, providing an integral solution for unauthorized access and illegal distribution of digital documents.

Nomenclature

DMS	Document Management System
SDMS	Secure Document Management System
OCR	Optical Character Recognition
DI	Document Imaging
NIST	National Institute of Standards and Technology
ISS Module	Information Security Services Module
SKC	Symmetric Key Cryptography
AKC	Asymmetric Key Cryptography
PKI	Public Key Infrastructure
CA	Certification Authority
RA	Registration Authority
MAC	Mandatory Access Control
DAC	Discretionary Access Control
RBAC	Role-Based Access Control
SSL	Secure Sockets Layer
VPN	Virtual Private Network
LSB	Least Significant Bit
DCT	Discrete Cosine Transform
IDCT	Inverse Discrete Cosine Transform
QIM	Quantization Index Modulation
SS	Spread Spectrum
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity
DDR	Digital Document Repository

1

Introduction

1.1 Context and Motivation

The adoption of information systems is changing the way organizations work, allowing the automation of their processes and thus making them more efficient. One important process for organizations is the management of large volume of documents. Since most of the organization's documents are in physical format, many organizations have digitalized these documents to reduce the use of paper and the space required for their storage. Digital documents (also referred simply as *documents* for the rest of this thesis) are managed by Document Management Systems (DMS) that allow instant access to information, faster search of documents and their content, easily back up and centralized security [9, 10]. The main task of a DMS is to control the document lifecycle which is the workflow that defines how users interact with the documents, from their creation until their usage in other organization processes. The document lifecycle of an organization depends on the nature of the documents (contracts, business letters, invoices, budgets, reports, etc). Also, the document format

is important when the document lifecycle is defined.

When documents are generated or received in physical format, organizations digitalize them, using for example a scanner. Despite of the existence of tools for Optical Character Recognition (OCR) to extract the text of scanned documents, physical documents are digitalized as images in order to preserve hand-written signatures, seals of approval or any other symbol that validates the document's content. The process of scanning physical documents and store them as digital images is known as Document Imaging (DI) [11]. DI defines a specific document lifecycle for a DMS that is illustrated in Figure 1.1 [12, 13, 14].

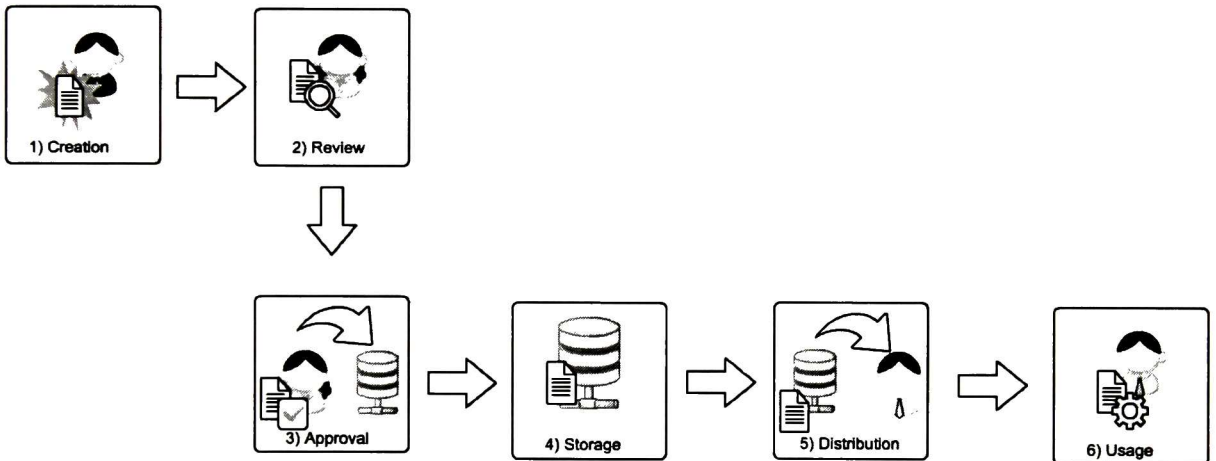


Figure 1.1: Lifecycle for Document Imaging.

The document lifecycle begins with the Creation stage, when a user creates a physical document. Then, that document is moved to the Review stage where it is reviewed and validated e.g. with handwritten signatures. After that, the document reaches the Approval stage and it is digitalized as an image. From this stage, the digital document must not be modified anymore, since the signatures on it express the will of the signors for the content that they received. Once digitalized, the document is uploaded and enters to the Store stage in a DMS which can use OCR tools to extract key words from the digital document and indexing it. Finally, the digital document is ready for being used, in the Distribution and Usage stages.

Usually, the access to digital documents is restricted only to authorized users, that are those who can be identified by the DMS. However, organizations are often divided to departments, and when a user uploads a digital document to the DMS, final users of that document belong to the same department. Hence, two types of DMS users per department are identified: Reviewer users that approve and upload digital documents, and Consumer users that retrieve digital documents from the system and use them. An special type of Consumer user is the Auditor. This type of user validates digital document's content in the DMS, and it is able to access any digital document to trace back Reviewers that uploaded a certain digital document. An Administrator user can delete digital documents and manage users. The specialization of users is shown in Figure 1.2 and the actions that they can perform are illustrated in the use case diagram in Figure 1.3.

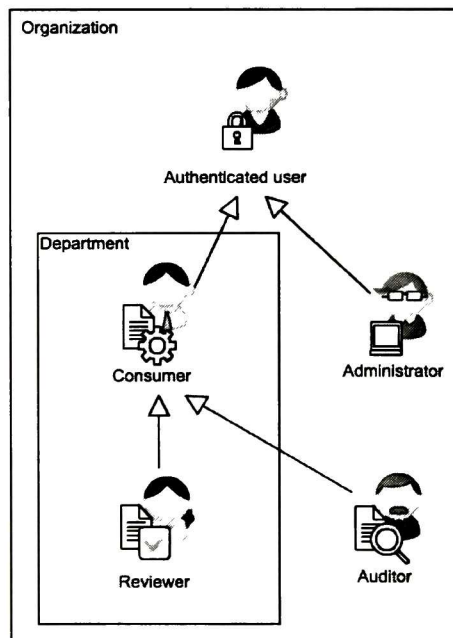


Figure 1.2: Type of users in a Document Management System.

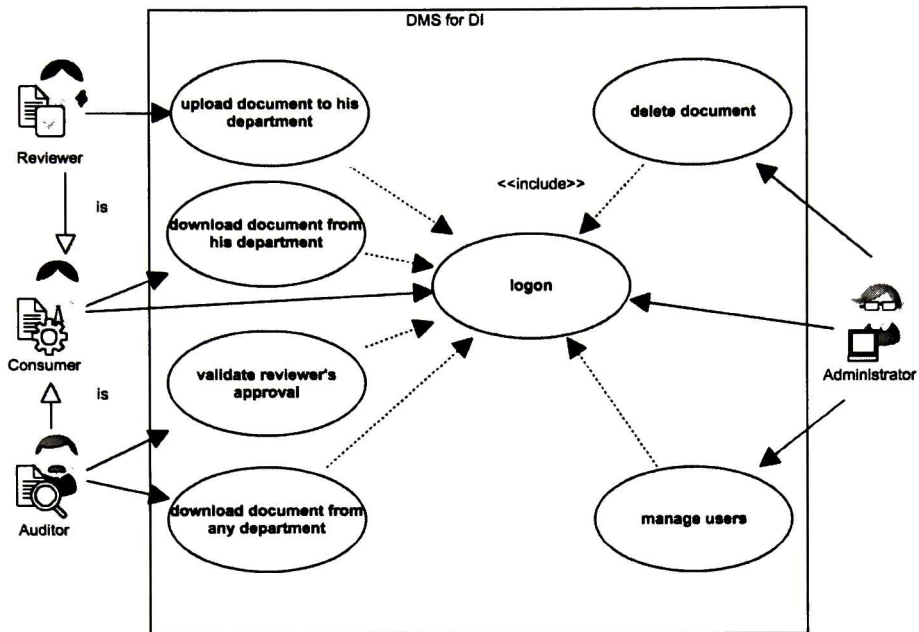


Figure 1.3: Use cases for Reviewer, Consumer, Auditor and Administrator users in a Document Management System.

1.2 Problem Description

Digital documents contain most of the information of organizations and they are used in a wide variety of processes in which it is necessary to access their content. Due to the importance of digital documents for organizations, they must be kept secure. In 2010, the Computer Emergency Response Team (CERT) reported an inquest applied to more than 500 respondents (including business, government executives, professionals and consultants) which revealed that 36% of information leakage is due to unauthorized access of malicious users and dishonest users that belong to the organizations and distribute illegally these information. Also, it was reported that 64% of the respondents are not able to take corrective actions due to the lack of evidence and because it is not possible to identify the individuals that distribute illegally the information [15]. In order to keep secure the digital documents in a DMS, there are two problems that must be solved: illegal access

to digital documents from unauthorized users, and the illegal distribution of digital documents by authorized but dishonest users. These problems are described in the next sections.

1.2.1 Illegal Access to Digital Documents from Unauthorized Users

One of the security requirements to protect digital documents is to prevent unauthorized access. Some consequences when malicious users access the content of confidential digital documents are: illegal distribution of documents, exposure of internal business strategies, privacy violation of medical records of patients, decrease of customer's confidence towards organizations, and in general, damages to the prestige of organizations. An unauthorized user can request a digital document directly to the DMS, therefore, when users download digital documents from a DMS, it is necessary to identify them as part of the authorized users. A similar case occurs in the upload of a digital document in which only authorized user can perform the upload, and the DMS must be able to identify unequivocally the user that approved, uploaded and stored a digital document at any time. Also, despite of validating the identity of authorized users, not all the digital documents are able to every authorized user e.g. a Consumer user of department A could not be able to download a digital document of department B. Unauthorized users could try to modify the content of digital documents during their transmission and storage, so document security is also required under this scenario.

In order to protect digital documents in a DMS, information security services are required. These services are [16, 17]:

- Confidentiality: ensures that digital documents are only accessed by authorized users.
- Access control: defines actions that users can perform on a digital document.
- Integrity: ensures digital documents have not been altered.
- Authentication: ensures the origin of the digital document.
- Non-repudiation: prevents that a user denies his/her signature on a digital document.

With these services, a DMS would be able to protect the digital documents from threats such as alteration or access by unauthorized users.

1.2.2 Illegal Distribution of Digital Documents by Authorized but Dishonest Users

Once an authorized Consumer user obtains a digital document from the DMS, he/she obtains it as plaintext for the Usage stage, but this leaves the digital document without protection. Due to vulnerability of digital documents during the Usage stage, Consumer users could act dishonestly, by distributing digital documents to unauthorized users. Digital documents distributed illegally are known as *pirate documents* and the users that distribute them as *traitors*. It is necessary to identify the traitors when pirate documents are detected to stop the distribution of pirate documents. Non-computational techniques can be used to find the identity of the traitor, however, it could take an undetermined time to do it. During this time the traitor can continue distributing new pirate documents, hence, not only the traitor user must be identified, but his/her identification must be done as fast as possible when a pirate copy is detected. The problem of pirate documents can not be solved if it is not possible to identify the traitors in organizations [18, 19, 20, 21].

1.3 Investigation Question

Given the two problems of illegal access to digital documents from unauthorized users and their illegal distribution by traitor users, the following investigation question is raised:

Is it possible to build a secure system for storage and distribution of digital documents that can be able to trace dishonest users?

To build a system with these characteristics could ensure the secure distribution of digital documents and their storage in a secure way. Furthermore, providing the required information security services ensures that only authorized users can distribute pirate documents, and in that case, the system would be able to detect the traitor users, enabling organizations to act accordingly in a quick and effective way.

1.4 Objectives

To respond the investigation question, a set of objectives are established in the next two sections.

1.4.1 General Objective

To develop a system that allows the storage and distribution of digital documents securely, and it is able to effectively identify dishonest users.

1.4.2 Specific Objectives

- To provide a mechanism that brings information security services for the protection of digital documents in the Approval, Storage and Distribution stages of the document lifecycle.
- To provide a robust mechanism for secure distribution of digital documents which allows to trace traitor users that distribute pirate documents during the Usage stage of the document lifecycle.
- To integrate the mechanism that brings information security services with the tracing traitor mechanism in such way that both mechanisms can work in an effective way to protect digital documents in the Approval, Store, Distribution and Usage stages of the document lifecycle.

1.5 Thesis Outline

This thesis is organized in 6 chapters. The first chapter is intended to provide the context of this thesis work. Chapter 2 is dedicated to the theory that supports this work. Chapter 3 reviews the state-of-the-art of secure document management systems and user tracing for digital documents. Chapter 4 describes the proposed system for secure storage and distribution of digital documents with dishonest user tracing. Chapter 5 presents and discusses the results of main modules of the proposed system as well as the complete integrated system. Finally, Chapter 6 concludes this work and points out future work.

2

Background of Security Information Services for Digital Documents

This Chapter presents a review of concepts and techniques that provide information security services to digital documents. These techniques are grouped according to the security service provided. Also, common architectures for secure document management systems are reviewed.

2.1 Confidentiality

Confidentiality is a security information service which guarantees that information can be accessed only by authorized users. A common way to provide confidentiality is by means of cryptographic techniques. Cryptography allows to hide a message by making it illegible, allowing only to those entities that are authorized to recover the original message. A secure communication of messages between two parties involves the next elements.

- **Sender:** it is the entity who sends the message.
- **Receiver:** it is the entity that receives the encrypted message.
- **Clear message:** it is the legible message.
- **Keys:** they are a sequence of bits that are used to encrypt and decrypt the message.
- **Encryption algorithm:** it is the method to transform the clear message into a encrypted message using a key.
- **Encrypted message:** it is the message resulting from the encryption algorithm applied to the clear message.
- **Decryption algorithm:** it is the method to transform the encrypted message into a clear message using a key.

A system that uses encryption and decryption algorithms to send messages in a secure way is known as a *cryptosystem* [22]. In the context of a DMS, digital documents are considered as the message that must be stored and sent in a secure way. The basic attack that a cryptosystem must resist is the brute force attack which consists on computing all the possible keys and try to decrypt the message with all the computed keys [23].

2.1.1 Symmetric Key Cryptography

Symmetric Key Cryptography (SKC) also named private key cryptography, consists in using a unique key to encrypt and decrypt a digital document, that means the sender and receiver have the same key. This scheme is shown in Figure 2.1.

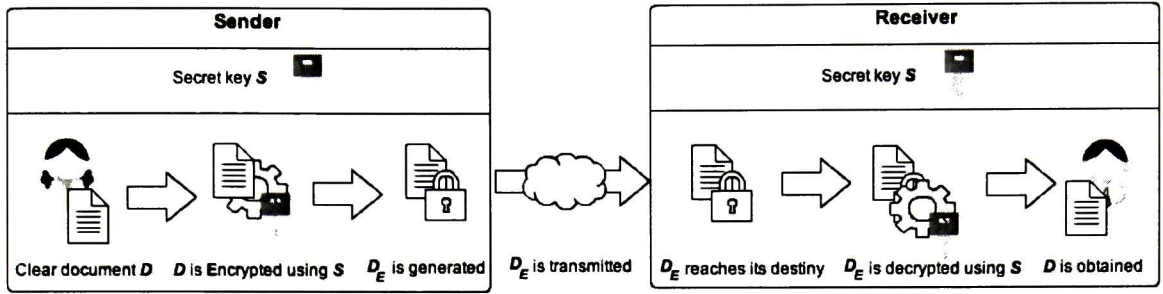


Figure 2.1: Private key cryptography scheme.

In SKC, using a secret key S , a digital document D is encrypted by a function $Enc(S, D)$ generating an encrypted digital document D_E . Using a decryption function $Dec(S, D_E)$ it is possible to obtain the original digital document D . The algorithms based on this technique are secure if computing S from $Enc(S, D)$ is a computationally intractable problem. SKC allows the transmission of digital documents between two systems securely over a network in a simple and efficient way. However, it is necessary that the sender and receiver to share the same key before the establishment of a secure channel, thus the secure sharing of the private key is a problem. Another problematic situation is that, in order to establish a secure communication between U users are required $U * (U - 1)/2$ keys. Therefore, the number of keys increases exponentially [24]. Representative SKC algorithms are:

- **Data Encryption Standard (DES):** it is a block cipher that uses a 56-bit key length (allowing a key space of 2^{56}). It was designed in 1973 and it was the first encryption standard. In 1976 it was approved by the NIST in the United States, and adopted later by most of the countries. With the increasing of the processing power, a brute force attack was performed successfully over DES on 1997 [23, 25].
- **3DES:** it was created to replace DES. This algorithm has a 168-bit key length since it uses 3 DES keys k_1, k_2, k_3 to perform a encrypt-decrypt-encrypt sequence defined in Equation 5.1 in order to encrypt the message, and the inverse operations defined on Equation 5.2 to obtain

the original message.

$$c = E(k_1, D(k_2, E(k_3, m))) \quad (2.1)$$

$$m = D(k_1, E(k_2, D(k_3, c))) \quad (2.2)$$

A key space of 2^{168} is very hard to compute, so 3DES is considered a secure symmetric algorithm. However, 3DES algorithm has presented performance issues [23, 25].

- **Blowfish**: it was presented in 1993 and designed to be implemented in hardware. Blowfish uses a key in the range from 32 to 448-bit length. Currently there is not a way known to break its security [25].
- **Advanced Encryption Standard (AES)**: it is the current encryption algorithm standardized recommended by the NIST. AES is a block cypher that can use a 128, 192 or 256-bit key length. The data to be encrypted is considered as a byte matrix and several processing rounds are performed on it depending on the key length: for a 128-bit key length 9 rounds are performed, for a 192-bit key length 11 rounds are performed and for a 256-bit key length 13 rounds are performed. The processing rounds of AES are defined by the next steps:
 - Bits substitution: change the block bits using a substitution table.
 - Row shifting: permute the rows of the block.
 - Column mixing: a substitution method is used for each block column.
 - Key addition: the logical XOR operation is performed using the key and the block.

The advantages of the AES algorithm are its speed and flexibility. Today, it can be found on different platforms including mobile devices [25].

2.1.2 Asymmetric Key Cryptography

Asymmetric Key Cryptography (AKC) also named public key cryptography, was proposed to solve the problems of SKC previously mentioned. AKC was proposed by Whitfield Diffie and Martin Hellman [26]. In AKC, each user has a public key P_u that is known by the other users, and a private key P_r that each user keeps in secret. A sender can send a message D to a receiver in a secure way by encrypting D with the public key P_u of the receiver using a function $Enc(P_u, D)$ generating the encrypted message c . Then, only the receiver is able to decrypt c with his correspondent private key P_r using a function $Dec(P_r, D)$, allowing to recover D . The AKC scheme is shown in Figure 2.2.

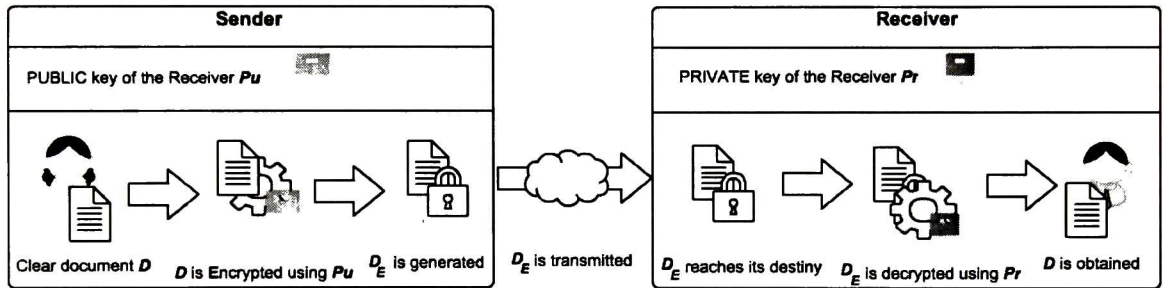


Figure 2.2: Public key cryptography scheme.

AKC provides the confidentiality and integrity services, in addition, users can associate their identity to digital documents by generating a digital signature using their private keys, providing the non-repudiation service [16, 27]. Despite of this benefits, AKC algorithms are more computationally expensive than SKC algorithms. AKC algorithms are divided into two families: those based on the discrete logarithm problem in which RSA is the most representative algorithm, and those based on Elliptic Curve Cryptography (ECC):

- **RSA:** it is the most widely used public key cryptosystem, especially in e-commerce systems. RSA defines three basic operations that are keys generation, encryption, and decryption. For the generation of the keys, it is required a value n that is obtained from the multiplication of two prime numbers p and q , then a function $\varphi(n) = (p - 1)(q - 1)$ allows the generation

of the public key e , that must be coprime of $\varphi(n)$ and satisfies the condition $1 < e < \varphi(n)$. For generating the private key d , $e^{-1} \bmod \varphi(n)$ is calculated. A message m is encrypted as $c = m^e \bmod n$. The decryption is defined as $m = c^d \bmod n$.

- ECC: it was proposed by Koblitz [28] and Miller [29] as an alternative mechanism for implementing AKC. It has the advantage of using shorter keys than RSA, bringing the same security. ECC is based on the discrete logarithm problem on elliptic curves that is likely to be harder than the classical discrete logarithm problem. To generate the private and public key it is necessary to select a finite field q in which a elliptic curve $EC(q)$ is defined, also G is defined as an elliptic curve point generator of order n . Then a d value in the range $[1, n - 1]$ is chosen, this value is used as the private key. The public key Q is defined by $Q = k * G$ where $k * G$ is a scalar multiplication. The security level of the keys depends on $EC(q)$, there are predefined elliptic curves that met required security levels for standardization [30]. In order to encrypt a document m the sender selects a value k in the range $[1, n - 1]$ and generate the encrypted messages $c1 = (k * G)$ and $c2 = M + (k * Q)$, $c1$ and $c2$ are sent to the receiver that using his private key d can retrieve m by performing $m = c2 - (d * c1)$.

2.1.3 Digital Envelope

The digital envelope is a concept which consists in encrypting a message D with a symmetric key k obtaining D_k and then, encrypting k with AKC obtaining k_{AKC} . D_k and k_{AKC} are transmitted to the receiver that decrypts k_{AKC} , and uses k to decrypt D_k obtaining the original message D . Digital envelopes are fast in encryption and decryption since these functions are performed using symmetric algorithms and only k is encrypted with asymmetric algorithms. The general scheme of the digital envelope is shown in Figure 2.3.

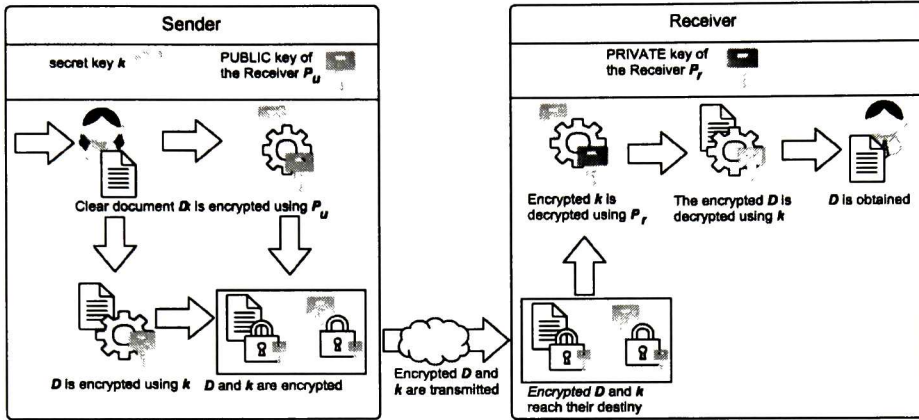


Figure 2.3: Digital envelope scheme.

2.1.4 Key Exchange

Key exchange algorithms are used to establish a secret key between two entities. That key can be used to encrypt a communication session between the entities. A representative key exchange algorithm is Diffie-Hellman, in which the sender and the receiver need to agree on a finite field \mathbb{F}_q and a base element g of order $N|q-1$. Then, the sender chooses secretly a random positive integer $k_s < N$ and computes $g^{k_s} \in \mathbb{F}_q^*$, and sends it to the receiver. The receiver acts in a similar way, choosing $k_r < N$ and performing and sending $g^{k_r} \in \mathbb{F}_q^*$ to the sender. The sender has the value k_s and receives g^{k_r} . In the same way, the receiver has the value k_r and receives g^{k_s} . Finally, both are able to generate the key $g^{k_s k_r} \in \mathbb{F}_q^*$. Even if an attacker knows both exchanged values, he/she must face the discrete logarithm problem (DLP) to obtain the private keys k_s and k_r . However, for F_q with order enough large, the DLP is considered a hard problem in computer science [31].

2.1.5 Digital Image Encryption

Due to digital documents in a DMS are stored as images, encryption techniques for images can be considered as an option to provide confidentiality. Two levels of security are considered in digital image encryption: high level security and low level security. In the low level, the image quality is

degraded by noise but it is possible to interpret the image content whereas in the high level security the image just look like noise. Digital image encryption can be carried out in both the spatial and the frequency domain [32, 33, 34]. Furthermore, this technique can be based in SKC or AKC, and it is more efficient than non-specialized cryptographic techniques for images. It is considered that certain amount of distortion can be generated when the image is decrypted [35].

2.2 Integrity

Integrity is a security information service that guarantees the accuracy and consistency of digital documents over their entire lifecycle. In a DMS, integrity services ensures that during the transmission and storage of digital documents these do not suffer any change in their content. Integrity can be provided by one way hash functions, message authentication codes and digital signatures.

2.2.1 One Way Hash Functions

A one way hash function, $H(m)$ receives a message m of arbitrary length as input and generates a fixed length hash value h . The generated value h is considered as a fingerprint of the message m [36]. The process of hashing is shown in Figure 2.4. Hash functions are non-reversible, this means that given m it is easy to generate h , however, it is very hard to compute m from h . Also, hash functions are collision resistant, that is, it is practically impossible to have two different messages m and \hat{m} such that $H(m) = H(\hat{m})$.

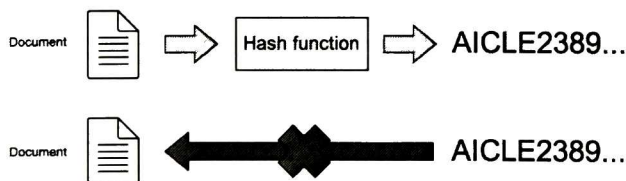


Figure 2.4: Hashing process.

Nowadays, 2 of the most widespread hash algorithms are:

- Message Digest 5 (MD5): it is based on the previous hash function MD4 proposed by Rivest [31]. These functions were designed for high-speed software implementations, trying to be as simple as possible and optimized for general purpose microprocessor architectures. MD5 as MD4 produces a 128-bit hash as output composed by a set of four 32-bit blocks. Some weaknesses that can compromise security in these functions have been reported.
- Secure Hash Algorithm (SHA): it is a family of hash functions proposed by the National Security Agency (NSA) and published by the National Institute of Standards and Technology (NIST). The SHA functions are based on the principles proposed by Rivest for MD4, and they have four versions that are: SHA-0, SHA-1, SHA-2, and SHA-3. No effective attacks for SHA functions for the third version are known [31].

2.2.2 Message Authentication Codes

Message authentication codes (MAC) are small fixed-size blocks of data that are generated based on a message of variable length using a private pre-shared key. In Figure 2.5 it is shown the general scheme for message authentication codes, in which a hash function receives a digital document as input and a private key in order to generate a hash value. When the digital document is transmitted, both, the document and its MAC are sent. Users that know the private key are able to validate the digital document integrity by calculating the MAC of the document received and comparing it against the MAC. If both MACs are same, the digital document integrity is validated.

2.3 Authentication and Non-Repudiation

Authentication is a security information service well suited to ensure the origin of a message. Digital signatures are used to prove message authenticity by verifying a signer's digital identity. Other ways

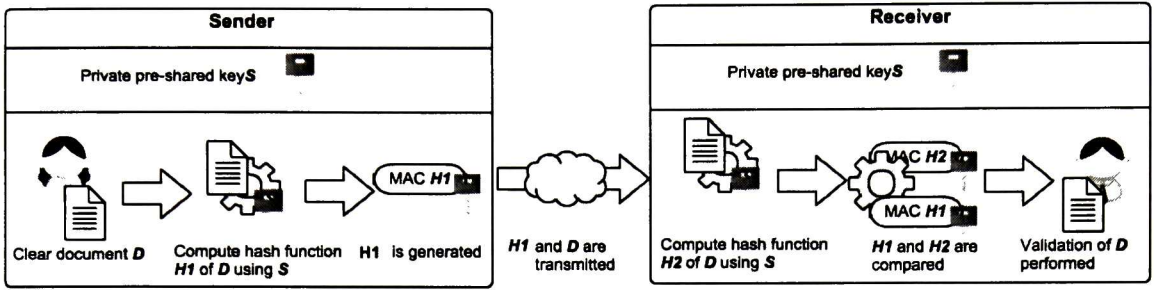


Figure 2.5: Generation and use of a Message Authentication Code for a digital document.

to provide this service are: registration of unique identifiers in a repository, meta-data with evidence of authenticity or watermarking. The non-repudiation security information service prevents that a user denies its participation in a communication. The non-repudiation service can be provided by digital signatures.

2.3.1 Digital Signatures

A digital signature is an electronic signature that can be used to authenticate the identity of the sender of a message, and that the message was not modified during its transmission. Basically, digital signatures are generated by obtaining a hash h_s of the message D that later is encrypted with the private key of the sender. It is known as *signed message* to the message in plain text with its hash encrypted using the private key of the sender. When the signed D reaches a receiver, the hash h_r of the D is computed. Then, the receiver decrypts the signature of D with the public key of the sender, obtaining h_s . If h_r and h_s are equal, the identity of the sender is validated, along with the integrity of D . This process is known as *digital signature validation*. The digital signature validation is expressed graphically in Figure 2.6 considering as message a digital document.

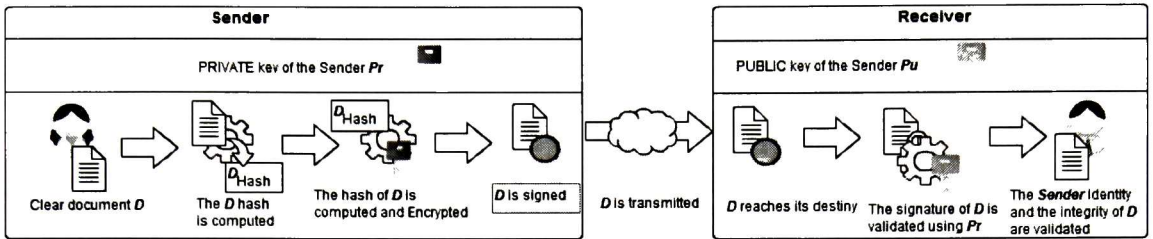


Figure 2.6: Digital signature scheme.

2.3.2 Public Key Infrastructure

Public Key Infrastructures (PKI) are commonly used for the implementation of AKC algorithms and digital signatures. A PKI distributes the public keys of users and ensures that a key belongs to a user by using digital certificates. The current standard of digital certificates is *X.509 standard ITU-T* (International Telecommunication Union) [31]. The PKI is composed by the following components:

- **Certification authority (CA):** it is responsible to assign and revoke certificates.
- **Digital certificate:** it is a document digitally signed by the CA, and establishes a link between a subject and its public key.
- **Registration authority (RA):** it is responsible for handling requests of certificates generation.
- **Repository of certificates and distribution system:** it provides a mechanism for storing keys, certificates and certificate revocation lists.

Today, the PKI is widely spread and there are standards that depend on it as *IPSec*, *SSLv3.0*, standards for sending emails, shopping online or bank transactions.

2.3.3 Mutual Authentication

Mutual authentication refers to two parties authenticating each other. The mutual authentication process starts when a client requests data to a server and the server responds with its certificate.

The client sends the obtained server certificate to a CA for validation, if the server certificate is validated, the client sends its own certificate to the server for a similar validation, in other case the communication is finished. After user and server authenticate each other, the client is able to download the data. Mutual authentication is described in Figure 2.7.

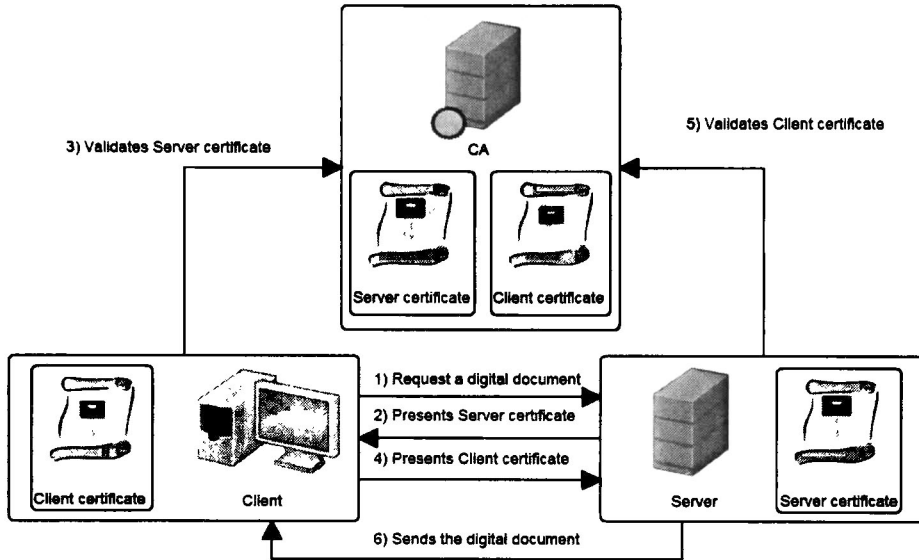


Figure 2.7: General scheme of mutual authentication.

2.4 Access Control

Access control is a security information service that allows or denies access and/or manipulation of a resource. Access control requires the definition of policies to determine the conditions that must be satisfied for the execution of certain actions. Conceptually, the essential component of access control models is the *reference monitor* that has the responsibility of enforcing the usage of the policies to access digital documents [37, 38]. The reference monitor can enforce policies in different ways that are described in the following sections.

2.4.1 Discretionary Access Control

In Discretionary Access Control (DAC), the owner of a digital document determines the actions that users can perform on digital documents based on their identity. Commonly, an access control list determines users permissions.

2.4.2 Role Based Access Control

In environments where users are created and deleted frequently, Role-Based Access Control (RBAC) grants the resource access to user groups in a flexible way and reduces administration effort. If a user does not belong to a group anymore, the system administrator just needs to change the role of that user.

2.4.3 Mandatory Access Control

Mandatory Access Control (MAC) determines the access privileges of users to resources by policies that are set by the system administrator. It is a more restrictive model because common users can not grant privileges to other users.

2.5 Digital Document Management Security

Architectures

The DMSs that integrate techniques to provide information security services often are based on common architectures that describe the document management environments, and coarse-grained strategies to cover the security requirements [39, 40]. In this section these architectures are reviewed.

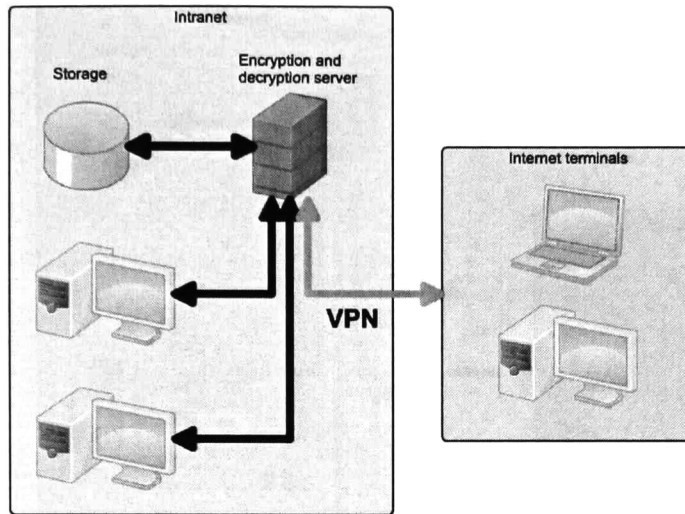


Figure 2.8: Encrypted server-based document management system architecture.

2.5.1 Encrypted Server-based Document Management System

Encrypted server-based secure DMS architecture is based on a client-server model and is common for small and medium organizations. Digital documents are generated by clients and sent unencrypted to an encryption server. The encryption server encrypts the digital documents, and after this point, the digital document can be sent to users or another servers in a secure way. In this architecture, information security services are provided by infrastructure services such as Secure Sockets Layer (SSL) or a Virtual Private Network (VPN), in order to protect the digital documents when they are transmitted from the client to the server and viceversa. This architecture is shown in Figure 2.8.

2.5.2 Terminal-Based Digital Document Management System

This architecture is commonly used by institutions and commercial companies, all the digital documents are encrypted in the terminal that is protected by using security software. The access to documents is authorized via MAC which sets the policies that determine the digital documents that certain terminal can access. This architecture is shown in Figure 2.9.

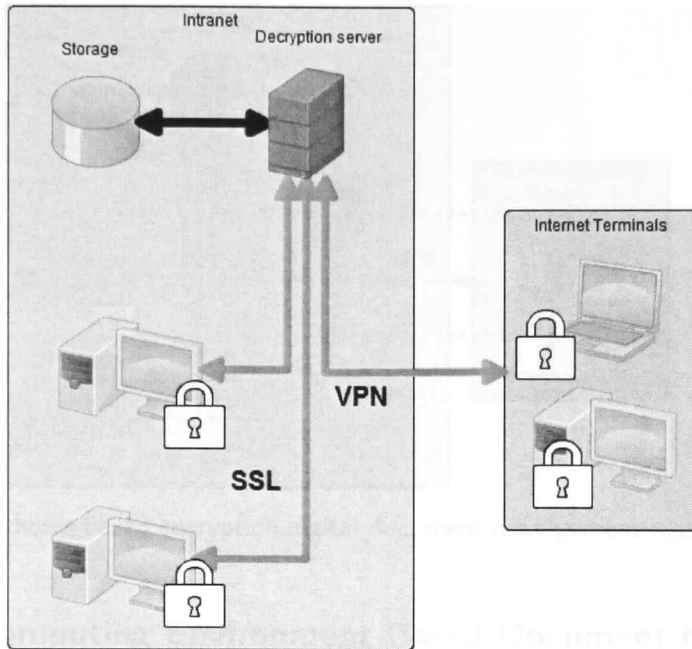


Figure 2.9: Terminal-based digital document management system scheme.

2.5.3 Hardware-Based Encryption Document Management System

In this architecture, the digital document encryption is performed by dedicated hardware in the terminal. Access control and decryption occur on the server side. This architecture is considered highly secure, with a better performance than encryption using software since it does not use the microprocessor of terminals. The drawbacks of this model are that a large amount of hardware is required to be deployed making implementation expensive and it could be difficult to change the hardware provider whereas software implementations use high level abstractions avoiding that problem. This architecture is shown in Figure 2.10.

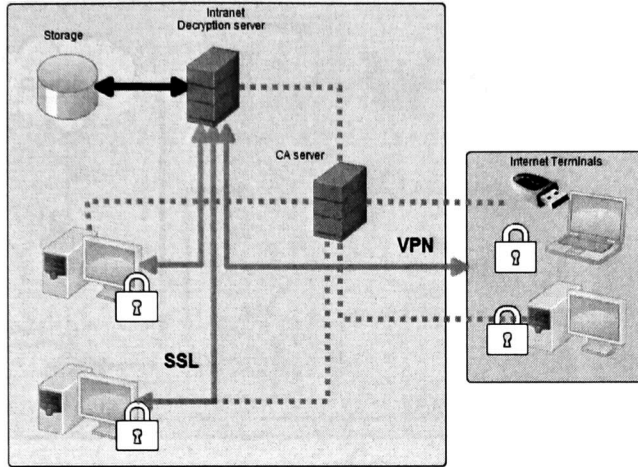


Figure 2.10: Hardware-based encryption digital document management system scheme.

2.5.4 Cloud Computing Environment-Based Document Management System

In this architecture, digital documents are easier to centralize, providing unified management. An organization is able to contract an external provider or implement its own private cloud. In order to secure the transmission of digital document all the users access the DMS by a VPN. This model is elastic since resources are centralized and are consumed as they are required, also, it is scalable due to the clients are connected to virtual machines, so new hardware components can be added without expensive changes. The centralized resources allows to apply information security services not only to a DMS but all the systems in an organization. This architecture is shown in Figure 2.11.

2.6 Summary

In this Chapter, concepts related to information security services and common ways to provide them were reviewed. These concepts and fundamentals are used in the next Chapters dedicated to the related work of security and user tracing for digital documents, the proposal of the secure

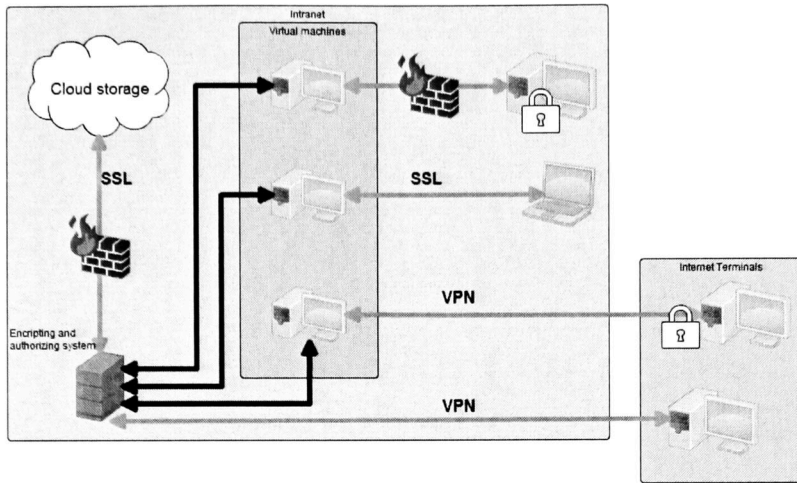


Figure 2.11: Cloud computing environment-based document scheme.

management system and its implementation.

3

Background of Fingerprinting Techniques for Digital Documents

In this Chapter, watermarking concepts are explored being digital fingerprinting the most important application of watermarking in this thesis work. Also, the Discrete Cosine Transform and the image quality assessment metrics are reviewed.

3.1 Digital Watermarking

Digital watermarking is the process of inserting information that cannot be removed easily into digital multimedia content such as image, audio or video. The inserted information (which is called *watermark*) in a watermarked document can be extracted or detected later. Since in this thesis work digital documents are represented as images, all the watermarking sections are focused on images. The general scheme of watermarking is shown in Figure 3.1. A watermark w is inserted in an image m

using a secret key s to allow only the holders of s to detect and retrieve the inserted watermark. The watermarked image m_w is transmitted through a channel in which m_w can be altered intentionally or unintentionally. Then, when the altered image \hat{m}_w is received a detector must be able to retrieve w from \hat{m}_w [41].

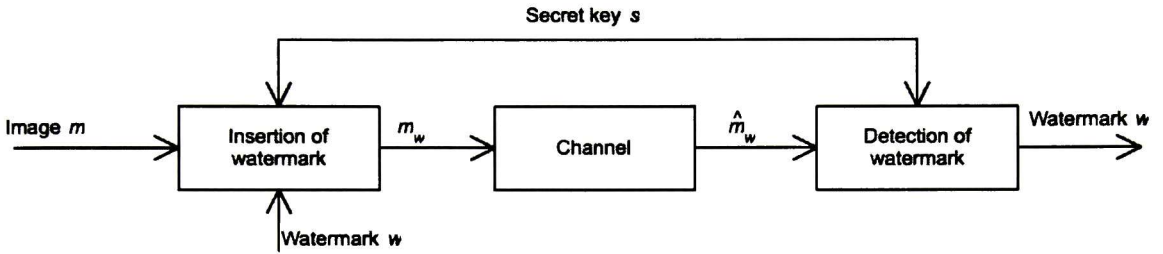


Figure 3.1: General scheme of digital watermarking.

3.1.1 Watermarking Classification

Watermarks can be classified for different characteristics, such as their perceptual transparency, their resistance to attacks, better known as robustness, or whether the original content is required when the detection of the watermark is performed [42]. Classification of watermarks are graphically shown in Figure 3.2. Next, these classifications, are described:

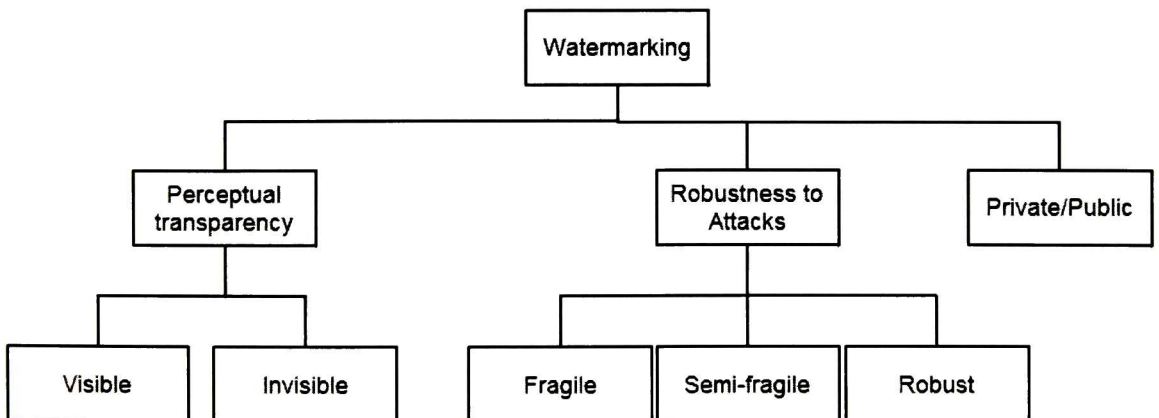


Figure 3.2: Diagram of watermarking classification.

- **Perceptual transparency:** Perceptual transparency refers to whether the watermark can be perceived by the senses of a user. Under this classification a watermark can be visible or invisible. Visible watermarks can be a text or a logo used to identify the author of the image. Invisible watermarks are inserted into an image in such a way that it cannot be perceived by human eye. It is used to protect the image authentication and prevent it from being copied.
- **Robustness to attacks:** When a watermarked image is sent through a channel, it is susceptible to intentional and unintentional attacks (these attacks are described in Section 3.1.3 of this Chapter). Watermarks can be classified depending on their capacity to resist attacks such as:
 - **Fragile watermarks:** these type of watermarks fail to be detectable if the watermarked image is modified intentionally or unintentionally.
 - **Semi-fragile watermarks:** these watermarks resist common operations such as lossy compression, rotation or size changes. However intentional attacks are able to destroy the watermark.
 - **Robust watermarks:** these watermarks are able to survive intentional and unintentional attacks.
- **Private and public watermarks:** in private or non-blind watermarks, the original content is required to retrieve the inserted watermark in a content copy whereas public or blind watermarks can be detected without the original content.

3.1.2 Watermarking Applications

The watermark's characteristics allow a wide variety of applications, such as:

- **Content protection:** inserting a visible watermark in an image, it can be publicly and freely distributed.

- **Content labeling:** watermarks can be used to bring information about the watermarked image.
- **Tamper detection:** detecting if a fragile watermark inserted in an image is degraded, it is possible to state that the image has been modified.
- **Proof of ownership:** if someone claims to be the author of a watermarked image, this can be proved by detecting the watermark.
- **Digital fingerprinting:** this is a process used to detect the owner of the image. In digital fingerprinting, watermarks will be unique to the owner and are known as *fingerprints*. Digital fingerprinting is an application of high relevance for this thesis work because it can be used to trace a traitor user from a pirate document. A more detailed explanation of digital fingerprinting is provided in Section 3.2.1 of this Chapter.

3.1.3 Watermarking Attacks

A watermarked image can suffer unintentional or intentional attacks. Unintentional attacks are those alterations in the watermark that occur during the normal use of watermarked images. Intentional attacks occur when malicious users perform actions to remove or degrade the watermark from a watermarked image [43, 44]. Attacks that can remove or destroy watermarks are:

- **Signal processing attacks:** these attacks occur by the effect of recompression, resampling, requantization, analog-to-digital and digital-to-analog conversion, and changes in color or contrast.
- **Geometric attacks:** these attacks occur for transformations in the image such as rotation, translation, cropping and scaling.
- **Synchronization attacks:** the basic idea of these attacks is to detect synchronization patterns, remove them, and then apply desynchronization techniques, such as global affine

transformations. After an attack of this nature, the watermark is still in the image, but it is not possible to detect it.

- **Collusion attacks:** these attacks occur when traitor users combine their content copies in order to generate a new pirate content destroying their fingerprints in the process [45]. A common way to perform collusion attacks is through an average of the pirate copies.

3.1.4 Watermarking Techniques

There are different watermarking techniques that define how the watermarks are inserted and detected. In the next sections these techniques are detailed.

3.1.4.1 Least Significant Bit Substitution

Least Significant Bit (LSB) substitution is a technique in which the least significant bit of the pixels of an image is overwritten with one bit of the watermark. For example, as shown in Figure 3.3 for the insertion of a bit in a gray scale pixel with the value 128, the LSB is changed resulting in the new value 129 [46].

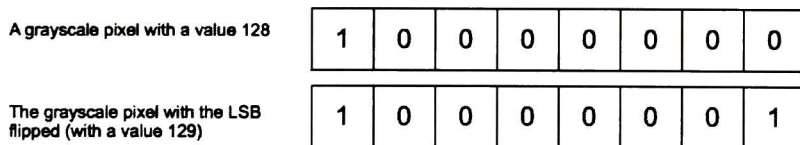


Figure 3.3: Least Significant Bit substitution.

The inserted watermark is invisible since the human visual system is not able to detect the difference in very similar colors. The LBS substitution technique presents disadvantages when the watermark must be robust. Despite of this watermarking technique can survive simple transformation such as cropping [46], it is not able to resist compression or noise addition. Also it can be detected statistically by an attacker [47]. Furthermore, a malicious user can overwrite the LBS of all the image pixels by a random bit, destroying the watermark.

3.1.4.2 Patchwork

The patchwork technique is based on a pseudo-random statistical process. It was proposed in 1996 by Bender [48] for marking images. Yeo and Kim [49] generalized the patchwork algorithm for images based on Blender's work, the Arnold's contributions [50] and the MPA algorithm [51]. This algorithm is known as *Generalized Patchwork Algorithm (GPA)*. To insert a watermark in GPA, the image is transform using a Discrete Cosine Transform (DCT, this concept is reviewed in section 3.4 of this Chapter) block $N_R \times N_C$, indexes sets $I = \{I_1, \dots, I_{2n}\}$ are generated and selected based on a random function that receives the watermark. The indexes are integrated by integers $[Z_1, Z_2]$ where this pair is in the selected positions of zig zag pattern in the DCT block. Through an index set $I^0 = \{I^{0+}, I^{0-}\}$ of pixels, it is described the position of a encoded binary value 0 and with $I^1 = \{I^{1+}, I^{1-}\}$ the position of an encoded binary value 1. It is divided each subset in 2 subsets, expressed in Equation 3.1 and Equation 3.2.

$$I_{0+} = \{I_1^0, \dots, I_n^0\}, I_{0-} = \{I_{n+1}^0, \dots, I_{2n}^0\} \quad (3.1)$$

$$I_{1+} = \{I_1^1, \dots, I_n^1\}, I_{1-} = \{I_{n+1}^1, \dots, I_{2n}^1\} \quad (3.2)$$

Having the coefficients DCT as $A^j = \{A_1^j, \dots, A_n^j\}$ corresponding to I^{j+} and as $B^j = \{B_1^j, \dots, B_n^j\}$ corresponding to I^{j-} , $j = \{1, 0\}$, For the case of the encoded binary value 0 it is computed the average as is expressed in Equation 3.3 and Equation 3.4.

$$A_{average}^0 = n^{-1} \sum_{t=0}^n A_n^0 \quad (3.3)$$

$$B_{average}^0 = n^{-1} \sum_{t=0}^n B_n^0 \quad (3.4)$$

Then, the variance is computed as shown in Equation 3.5 and Equation 3.6.

$$S^2 A_0 = \frac{1}{(n-1)} \sum_{t=0}^n (A_n^0 - A_{average}^0)^2 \quad (3.5)$$

$$S^2 B_0 = \frac{1}{(n-1)} \sum_{t=0}^n (B_n^0 - B_{average}^0)^2 \quad (3.6)$$

Finally, the DCT coefficients are replaced by A^0 and B^0 as is expressed in Equation 3.7 and Equation 3.8.

$$A_t^{0*} = (1 + \text{sign}(S_{A_0}^2 - S_{B_0}^2)P_1)A_t^0 + \text{sign}(A_{average} - B^0)\sqrt{P^2}S_{E0}/2 \quad (3.7)$$

$$B_t^{0*} = (1 + \text{sign}(S_{A_0}^2 - S_{B_0}^2)P_1)B_t^0 - \text{sign}(A_{average} - B^0)\sqrt{P^2}S_{E0}/2 \quad (3.8)$$

Where P_1 y P_2 have the control of the robustness and perceptibility of that watermark, and sign is a function to extract the sign of a real number. This algorithm decreases the possibilities of a false positive detection, it is resistant to signal process and to the compression techniques and it is highly inappreciable.

3.1.4.3 Quantization Index Modulation

Quantization Index Modulation (QIM) refers to inserting information, modulating an index or sequence of indices with the watermark information and then quantizing the host signal with the associated quantizer or sequence of quantizers. To insert a watermark bit w in an image x it is necessary to choose a quantization value that depends on the binary value of w , representing a quantization step size Δ , if even it is denoted by $Q\Delta_{even}(\cdot)$ and if odd it is denoted by $Q\Delta_{odd}(\cdot)$ [52].

The insertion techniques based on QIM present a good balance between the performance and

robustness of the watermark, the degradation of the image that is inserted and the amount of information on the watermark. However, QIM has some disadvantages in terms of safety, because with a single copy of a watermarked image it is possible to determine the location of the watermarks through the generation of a high-resolution histogram of the image which allows to observe increases Δ and hence remove them [53]. Another aspect to consider is the sensitivity to an amplitude scale attack which affects considerably the performance of the watermark detector [54].

3.1.4.4 Spread Spectrum

Spread spectrum (SS) is widely used in telecommunications and involves transmitting a narrow band signal into a wide band signal such that the energy at each frequency is undetectable. Using SS as a watermarking technique, the image is considered as a communication channel and the watermark as a message to be transmitted, so distortions caused by an attack or by image processing are considered noise. The carrier signal, in this case an image, is also considered a noise source in the model of the communication channel [55].

By using SS, the watermark is inserted in the most significant regions of the image, because the less significant regions are the most affected or removed by applying filters or other image processing techniques. Inserting the watermark in the most significant regions has the advantage that if a user tries to remove the watermark, the image would be severely damage [56].

Cox [57] proposed the first watermarking scheme based on the SS technique for images. In this scheme, the watermark is represented by a sequence $X = \{x_1, \dots, x_n\}$ of real numbers in which each value in the sequence is calculated based on a gaussian distribution. To insert the watermark X in an image D , a sequence $V = \{v_1, \dots, v_n\}$ is extracted from D , in which X is inserted generating a sequence $\hat{V} = \{\hat{v}_1, \dots, \hat{v}_n\}$. The sequence \hat{V} is inserted in D by replacing the sequence V , given as a result a watermarked image \hat{D} . In order to extract the watermark of \hat{D} , it is extracted a sequence $\hat{V} = \{\hat{v}_1, \dots, \hat{v}_n\}$ that is compared with the sequence V given as a result a sequence $\hat{X} = \{\hat{x}_1, \dots, \hat{x}_n\}$ that is the inserted watermark \hat{X} . Due to \hat{X} can differ from X due to alterations

suffered by intentional or unintentional attacks, a threshold is computed and the similarity of \hat{X} and X is evaluated, if their similarity is over the threshold, the watermark is detected. SS is robust against geometric transformations, signal processing and collusion attacks. However, SS requires a lot of computational resources because it establishes a correlation between the extracted watermark and the watermarks stored in the system. Also, the possibility of false-positive detection increases if the watermarks are represented by short sequences [58].

3.1.5 Analysis of Watermarking Techniques for Digital Images

The advantages and disadvantages of the presented approaches for image watermarking are summarized in Table 3.1.

	LSB	Patchwork	SS	QIM
Advantages	Highly imperceptible	Highly imperceptible	Robustness against common attacks, and collusion attacks	Efficiency, good balance between fingerprints size and image degradation
Disadvantages	Statistically detectable, not robust against image transformations	Not robust against image transformations	Small fingerprint, establish correlations is not efficient	Detectable by analyzing the image histogram
Disadvantages	Statistically detectable, not robust against image transformations	Not robust against image transformations	Small fingerprint, establish correlations is not efficient	Detectable by analyzing the image histogram

Table 3.1: Advantages and disadvantages of watermarking techniques.

From these techniques, SS presents resistance to collusion attacks, which makes it suitable for fingerprinting applications. SS has been applied for fingerprinting in the context of the copyright protection in an effective way, because it is robust to a wide variety of attacks. Since the first SS work in watermarking proposed by Cox [57], SS schemes for fingerprinting have evolved taking into account the behavior of traitor users and for improving their performance. Wang *et al.* [59] considered

that traitors are more likely to collude with users who share common characteristics such as social circumstances or geographic location. Under this assumption, a hierarchical fingerprinting technique based on SS was proposed. Using that technique, users are assigned to groups, and the fingerprint is generated from the user ID and its group ID. At the detection stage, the group of the colluders is identified and then, the users who belong to that group are identified. This strategy reduces the computational cost along with the probability of false positive detection. Kuribayashi proposed in [60] a hierarchical fingerprinting scheme based on Code Division Multiple Access (CDMA). In that scheme, users are organized in groups, and the user fingerprint is represented by two sequences: a SS sequence for the user ID and other for the group ID. These sequences are orthogonal because they are DCT basis vectors modulated by a pseudo-noise sequence that is a pseudorandom sequence of 1 and -1 values allowing retention of orthogonality. The scheme presented in [60] has a better performance detecting colluders than Cox and Wang schemes due to a fast DCT implementation, which reduces the computing time from linear to logarithmic scale.

3.1.6 Digital Text Watermarking

Digital text watermarking is the process of inserting a digital watermark on a digital text document in order to avoid illicit re-distribution and copyright violations. Limited progress for digital text watermarking has been reported. There are three main approaches for digital text watermarking that are: syntactic approach, semantic approach and Image-based approach [61, 62]. These approaches are explained in the next sections.

3.1.6.1 Syntactic Approach

Hassan *et al.* proposed the natural language watermarking by performing morphosyntactic alterations to the text [63]. First, the text in the digital document is transformed into a syntactic tree diagram where the hierarchies and the functional dependencies are made explicit and watermark is inserted. The watermarking process is shown in Figure 3.4. In order to generate watermarked sentences, the

watermarking insertion mechanism can use dictionaries, a set of preprocessed sentences, syntactic tools or web tools as Wordnet. The problem with this approach is that depending on the content of

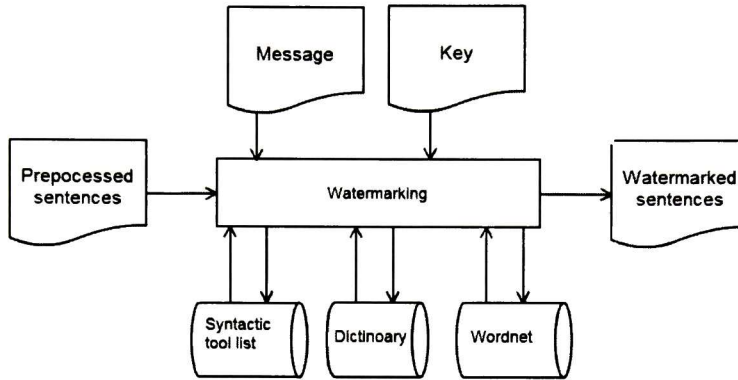


Figure 3.4: General scheme of syntactic text watermarking.

the text, it could be not possible to change it, e.g. digital documents of poetry, contracts or music sheets.

3.1.6.2 Semantic Approach

The semantic watermarking approach uses the text semantic structure to insert a watermark. Text contents like verbs, nouns, prepositions, words spelling, acronyms, sentence structure, grammar rules, etc, are exploited to insert the watermark in the text. Atallah *et al.* [64] proposed the semantic watermarking schemes. Later, the synonym substitution method was proposed in which watermark is embedded by replacing certain words with their synonyms without changing the context of text [65]. Xingming, *et al.* [66] proposed noun-verb based technique for text watermarking which exploits nouns and verbs in a sentence parsed with a grammar parser using semantic networks.

3.1.6.3 Image-Based Approach

In this approach of digital text watermarking, the text in a image-based document is used to insert the watermark. The main techniques of this approach are:

- **Line shifting encoding:** this technique inserts a watermark in the digital document by vertically shifting the locations of text-lines. One bit can be encoded per line. During the extraction of the watermark, text-lines are located using horizontal projection profile. The distance between adjacent text-lines is measured, this can be done by either measuring the distance between the baselines of adjacent lines or the difference between centroids of adjacent lines [67, 68].
- **Word shifting encoding:** this technique inserts a watermark in the digital document by horizontally shifting the locations of words within text lines. No watermarked lines are included to detect and compensate nonlinearities that occur in printing and copying. However, word shifting encoding is only applicable to digital documents with variable spacing between adjacent words, and because of the variable spacing requirement, decoding requires the knowledge of the space between words in the unaltered document [69, 70].
- **Inter-character space encoding:** inter-character spacing is varied to encode a binary watermark. This watermarking text strategy is designed for languages that have few or no large inter-word spaces. [71, 72].

Watermarking techniques on image-based digital documents were proposed by Brassil *et al.* [67] and Low *et al.* [68]. They consist of inserting watermarks by shifting the text-lines in a vertical way allowing a bit encode per line. These techniques only work for documents with a well defined style format and the inserted watermark is easy to remove by an average collusion attack. In [69, 70] techniques based on horizontally shifting of words in the document were proposed. However, these techniques can only be applied to documents that have variable spacing between adjacent words. The watermark detection is only possible having the knowledge of the space between words in the original document. Other techniques based on the distance between lines, words or characters have been proposed in [71, 72, 73, 74, 75, 76, 77] and most of them proved to be robust against indirect attacks such as copying, printing and scanning. Also, specific schemes have been proposed to face

these problems [78, 79]. However, the robustness of these techniques against collusion attacks is not reported.

3.2 Traitor Tracing

When digital documents are distributed to users and get the document in plain text, users can make use of them, being this the Usage stage of the digital document's lifecycle. At this point, digital documents are not under any security mechanism, so a dishonest user is able to unlawfully distribute the documents to unauthorized entities. As mention in the description problem in Chapter 1, the illicit copy of a digital document is known as *pirate document* and the authorized user which extracts originally the digital document is known as *traitor user*. Initially, piracy cases appeared in the distribution of CD ROMs, event subscription payments and access to online databases. In these cases, providers provided a key (a decoder, password, software, etc.) to a user to decrypt the content, which then distribute it illegally, allowing other users to access these contents.

In this situation, schemes for traitor tracing were necessary, that is, the identification of the traitor user [80]. The first traitor tracing schemes consisted of three elements:

- **User initialization scheme:** it is used to register a user and generates his/her key.
- **Scheme decryption:** it is used by the user to access content provider.
- **Traitor tracing algorithm:** it is used to obtain the traitor identity through the decoder mechanism using the key of the traitor user, when the mechanism is confiscated. The decoding mechanism can be hardware or software.

With the advent of online distribution of multimedia content such as copyrighted music, audio and video, the traitor tracing schemes evolved to adapt them to this new environment, in which users who have paid for some content are in the capacity of distributing it to users who have not

paid for it. A common way to provide traitor tracing is *digital fingerprinting* [81]. The next Section will delve into this concept.

3.2.1 Digital Fingerprinting

Digital fingerprinting is an application of watermarking in which the ID of a user (named fingerprint) is inserted in each digital content that he/she requests [82]. If, at a later point in time, a pirate copy of the content is found, the identity of the traitor user that distributed the pirate copy can be determined by retrieving the fingerprint. For the effectiveness of digital fingerprinting techniques, it is necessary to meet two main properties:

- **Robustness:** it is the ability of fingerprints to survive intentional and unintentional attacks after being inserted into the content. Due to fingerprints are unique per user, collusion attacks are possible. A robust fingerprint must be able to survive to collusion attacks.
- **Perceptual transparency** (namely unobtrusiveness, invisibility or imperceptibility): the original content and its fingerprinted copy must be identical to the user's perception. In the context of fingerprinting for digital documents, the perceptual transparency is measured using the natural language understanding [61, 62], so the characteristic of perceptual transparency is achieved if the text contained in the document is entirely legible after the fingerprint insertion.

For most of the existing fingerprinting techniques, robustness and perceptual transparency have an adverse effect on each other. Therefore, it is necessary to find a balance between these properties, for maximum robustness while maintaining perceptual transparency [83]. Robust fingerprinting techniques are usually not reversible, that means a user can not obtain the original content from a fingerprinted one, so the user must be able to manipulate it for legitimate purposes with the inserted fingerprint.

3.3 Image Quality Assessment

Image quality assessment metrics can be used to determine the amount of distortion generated by a watermark in an image. These metrics can be *full-referenced* if the watermarked image needs to be compared with the original image, or *non-referenced* if the metric tries to assess the quality of a watermarked image without any reference to the original one. There are several metrics that can be measured objectively and automatically evaluated. In this thesis the Peak Signal-to-Noise Ratio and Structural Similarity Index are used as metrics. These metrics are detailed in the next sections.

3.3.1 Peak Signal-to-Noise Ratio

Peak Signal-to-Noise Ratio (PSNR) measures the similarity of two images. It defines the relation between the maximum energy of a signal and the noise that affects expressing this difference in decibels [84, 85]. Given an *8-bit* grayscale image f and a copy of the altered image g , both of size $M \times N$, the PSNR between f and g is defined in Equation 3.9 and Equation 3.10.

$$PSNR(f, g) = 10 \cdot \log_{10} \left(\frac{255^2}{MSE(f, g)} \right) \quad (3.9)$$

$$MSE(f, g) = \frac{1}{(M \cdot N)} \sum_{i=1}^M \sum_{j=1}^N (f_{ij} - g_{ij})^2 \quad (3.10)$$

For MSE (Mean Square Error), the difference between pixels f_{ij} and g_{ij} is consider as an error that generates image quality loss. As MSE tends to zero, the value of PSNR approaches to infinity. Thus, higher values for the $PSNR(f, g)$ indicate higher image quality.

3.3.2 Structural Similarity Index

Structural Similarity Index (SSIM) is a particular implementation of the structural similarity philosophy [84], and it is considered correlated with the human visual system [86]. For two image

signals x and y , the comparison of three components luminance, contrast and structure is necessary. These components are relatively independent because object structures in images neither depend on illumination nor contrast. The luminance is defined by the function in Equation 3.11 where μ_x is the standard deviation of x .

$$l(x, y) = \frac{(2\mu_x\mu_y + C_1)}{(\mu_x^2 + \mu_y^2 + C_1)} \quad (3.11)$$

Then, the closeness of the contrast of the images is measured as shown in Equation 3.12 where σ_x is the variance of x .

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (3.12)$$

The structure comparison is defined by Equation 3.13 where σ_{xy} is the covariance between x and y .

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (3.13)$$

Finally, the three components are combined to get the overall similarity measure expressed Equation 3.14, where the exponents α , β and γ are positive integers that define the importance of each component.

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma \quad (3.14)$$

3.4 Discrete Cosine Transform

The Discrete Cosine Transform (DCT) expresses a sequence of data points in terms of a sum of cosine functions oscillating a different frequencies [87]. The DCT is an orthogonal transform that takes a discrete signal in the space domain and transforms that signal into its discrete frequency domain representation. It has a strong "energy compaction" that is the ability to pack the energy of the spatial sequence into as few frequency coefficients as possible. Due to this characteristic, the DCT is used for lossy compression of multimedia content, and also for inserting watermarks in the frequency domain of images and audio [60, 88]. There are four types of common DCT types

in which N are the number of real basis vectors whose components are cosines and j is the j – th component of the k – th basis vector. These DCT types are:

- **DCT-I:** it was introduced by Wang and Hunt [89]. It is expressed in Equation 3.15 and it is its own inverse.

$$\cos(jk\frac{\pi}{N-1}) \quad \text{divided by } \sqrt{2} \text{ when } k \text{ or } j \text{ is } 0 \text{ or } N-1 \quad (3.15)$$

- **DCT-II:** it was introduced by Ahmed *et al.* [90]. It is considered the most common DCT type and used for image and video compression. This DCT type is expressed in Equation 3.16.

$$\cos((j + \frac{1}{2})k\frac{\pi}{N}) \quad \text{divided by } \sqrt{2} \text{ when } k = 0 \quad (3.16)$$

- **DCT-III:** it is the inverse of the DCT-II. This DCT type is expressed in Equation 3.17.

$$\cos(j(k + \frac{1}{2})\frac{\pi}{N}) \quad \text{divided by } \sqrt{2} \text{ when } j = 0 \quad (3.17)$$

- **DCT-IV:** it was introduced by Kekre and Solanki [91] and it is used in audio coding algorithms. This DCT type is expressed in Equation 3.18.

$$\cos(j(j + \frac{1}{2})(k + \frac{1}{2})\frac{\pi}{N}) \quad (3.18)$$

The DCT is used in the solution for tracing users in the proposed system. Currently, software implementations of Fast DCT algorithms are available [92]. The selected implementation of the Fast DCT for this thesis work is based in the computation of the Fast Fourier Transform (FFT) algorithms, obtaining the DCT as a special case of the FFT in which a real sequence with even symmetry is received as input, canceling the sine terms in the FFT leading to a unnormalized non-orthogonal

DCT. Since the DCT obtained from the FFT is non-orthogonal, it is necessary to normalize the result of the DCT to retain orthogonality (the result of the scalar product of two vectors are equal to zero). In this thesis, orthogonality is a property that is exploited for tracing dishonest users. The normalized definition of DCT-II for arrays of two dimensions is expressed in Equation 3.19.

$$B_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2n+1)q}{2N}, \quad \begin{array}{l} 0 \leq p \leq M-1 \\ 0 \leq q \leq N-1 \end{array} \quad (3.19)$$

where:

$$\alpha_p = \begin{cases} \frac{1}{\sqrt{4M}} & \text{if } p = 0 \\ \sqrt{\frac{1}{2M}} & \text{otherwise} \end{cases} \quad (3.20)$$

$$\alpha_q = \begin{cases} \frac{1}{\sqrt{4N}} & \text{if } q = 0 \\ \sqrt{\frac{1}{2N}} & \text{otherwise} \end{cases} \quad (3.21)$$

A is a matrix of size $M \times N$ and B is the output matrix with the same dimensions of A . In this thesis, the DCT is computed for a matrix because it is applied to image-based digital documents.

3.5 Summary

In this Chapter, watermarking techniques and their application for digital fingerprinting were introduced along with image quality assessment metrics and the Discrete Cosine Transform. These concepts and techniques are used in the next Chapters dedicated to the related work of security and user tracing for digital documents, the proposal of the secure management system and its implementation.

4

Related Work on Security and User Tracing for Digital Documents

This Chapter is dedicated to review proposed solutions in the literature to protect digital documents and it is divided into two parts. The first one is dedicated to the review of Secure Document Management Systems (SDMS) and the second is dedicated to review recent papers on user tracing for digital documents. To the best of the author's knowledge, there is not any publications in the literature that has previously considered the integration of security information and user tracing in a SDMS.

4.1 Secure Digital Document Management Systems

In the 1980s the growth in the capacity of storage media, advances in networking and Internet access allowed organizations to centralize their information which was accessible to its members through a

network of computers [10]. Several approaches have been proposed to create SDMSs.

Casey *et al.* [1] proposed one of the first approaches of SDMS called Secure Automated Document Delivery System (SADDS). SADDS was developed as a British Library research project, allowing to search in bibliographic databases, and request digital documents using e-mail protocols. The general scheme of SADDS is shown in Figure 4.1. Information security services for SADDS are authentication of users, integrity of digital documents, non-repudiation of client actions, confidentiality of digital documents and access control. The selected technique to provide information security services

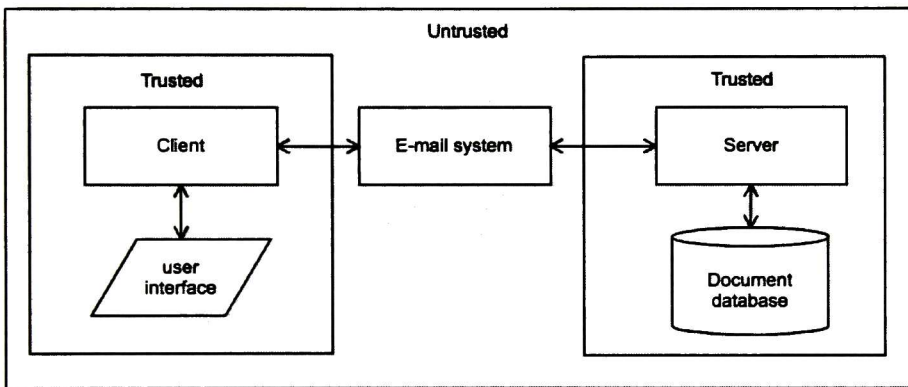


Figure 4.1: Scheme of the Secure Automated Document Delivery System [1].

in SADDS implements the RSA algorithm, using a Public Key Infrastructure (PKI). The scheme of SADDS was considered straightforward, cost effective, and provides a high degree of security. However, a SADDS uses the SMTP protocol to transmit the digital documents via email. This is a disadvantage in scenarios when new client applications need to be added.

Liroy *et al.* [2] presented a reusable scheme that provides information security services of integrity, authentication of sender and receiver, privacy and secure digital document management. This scheme based on a PKI is shown in Figure 4.2. When an Author uploads a digital document to the server, he/she is able to define the authorized users that are allowed to decrypt his/her digital documents using an extended access control list (XACL). When a Reader user r requests a digital document D , the Document server validates with the XACL if that user is able to get D . If r is able to get D ,

the Document server signs and encrypts D with a symmetric key DK . Then, DK is sent, packaged in a digital envelope to the Reader user, allowing the Reader user to decrypt the digital document. The PKI uses the RSA algorithm and the document is encrypted using the 3DES algorithm [93].

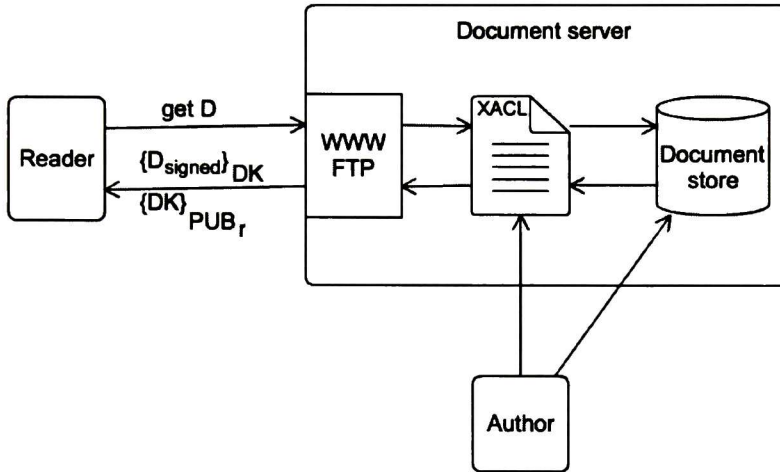


Figure 4.2: Document management system scheme proposed by Liroy *et al.* [2]

One of the main problems reported in the scheme proposed for Liroy *et al.*, is that there are no mechanisms against attacks of Distribute Denial of Service (DDoS) in which the Document server can be saturated by requests of an attacker denying to respond legitimate requests. Also, it is reported that the management of large XACLs becomes a problem since user privileges are set per each user. One of the main problems reported in the scheme proposed for Liroy *et al.*, is that there are no mechanisms against attacks of Distribute Denial of Service (DDoS) in which the Document server can be saturated by requests of an attacker denying to respond legitimate requests. Also, it is reported that the management of large XACLs becomes a problem since user privileges are set per each user. One of the main problems reported in the scheme proposed for Liroy *et al.*, is that there are no mechanisms against attacks of Distribute Denial of Service (DDoS) in which the Document server can be saturated by requests of an attacker denying to respond legitimate requests. Also, it is reported that the management of large XACLs becomes a problem since user privileges are set per

each user. One of the main problems reported in the scheme proposed for Liyo *et al.*, is that there are no mechanisms against attacks of Distribute Denial of Service (DDoS) in which the Document server can be saturated by requests of an attacker denying to respond legitimate requests. Also, it is reported that the management of large XACLs becomes a problem since user privileges are set per each user. One of the main problems reported in the scheme proposed for Liyo *et al.*, is that there are no mechanisms against attacks of Distribute Denial of Service (DDoS) in which the Document server can be saturated by requests of an attacker denying to respond legitimate requests. Also, it is reported that the management of large XACLs becomes a problem since user privileges are set per each user.

Gerasimov [94] suggested that the creation of a SDMS for a state organization can be achieved with technologies such as firewalls, secure Virtual Private Networks (VPN) and a PKI. Firewalls can be used to face DDoS attacks by an appropriated configuration. Secure VPNs allow a secure end to end transmission by encrypting all the traffic transmitted through the network. An advantage of secure VPNs is that they allow transmitting sensitive information over Internet in a secure way. Despite of this advantage, Gerasimov does not provide a mechanism for the secure digital document storage. Shi and Ouyang [3] also took advantage of PKIs in order to support commercial transactions by developing a system called *eSign* to manage digital signatures. The general scheme of *eSign* is shown in Figure 4.3. *eSign* is based on a PKI and allows registering users in an account that requires their first and last name, their certificate and their private key. Then, at the signing process when a user u uploads a digital document, *eSign* signs the digital document with the private key u . Administrator users can access and validate any signed digital document while other users can only access and validate their own digital documents. All the signing process and signature validation is performed by *eSign*. The signing strategy can lead to a security problem, because all the user private keys are centralized in *eSign*.

Kwok and Nguyen [95], researchers in the Crypto Group at Microsoft Research, proposed an SDMS for contracts by considering that different file formats can be used in different stages of the

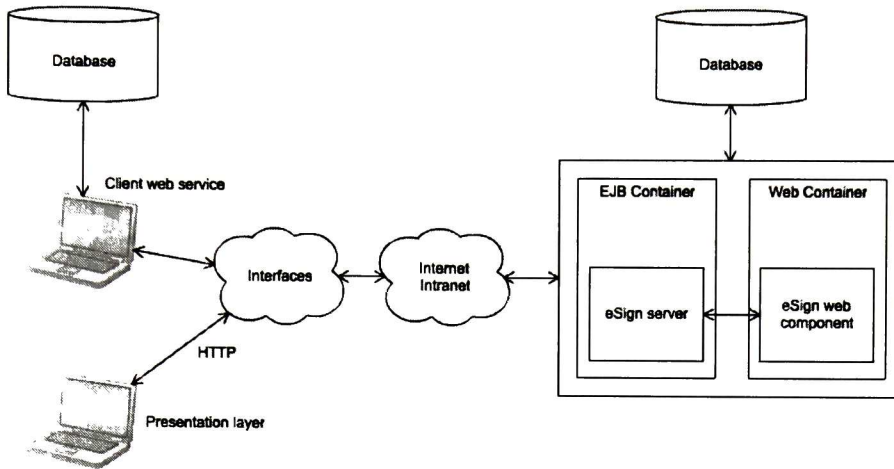


Figure 4.3: eSign Architecture proposed by Shi and Ouyang [3].

contract lifecycle. The file formats used in that system are PDF and XML. If the digital documents are received in any image format, the SDMS is able to convert the digital document to a useful format for the system with a XML and PDF converter that use an OCR engine. Then the XML and PDF generator will generate their corresponding XML and secure PDF contracts using contract templates, XML schemas, Cascading Style Sheets (CSS), and PDF styles. The XML contract is used during negotiations. When all the involved parties approved the final XML contract, the XML and PDF contract generators generates the final secure PDF contract which is digitally signed for all the participants on its creation. While this is a flexible way to handle contracts, information security services are provided until the last stage on the contract lifecycle, so in most of the contract lifecycle, contracts are not being protected.

Kamara and Lauter [4] proposed a model for the storage of digital documents in public clouds to avoid the costs of developing a private cloud for organizations. The scenario of this model considers that employees of an organization upload digital documents that are available for the partner's organization. The sequence of steps to upload and download digital documents are:

1. Each Organization employee and Partner employee receives a credential

2. Organization employees send their digital document to a specific repository
3. The data processor processes and encrypts the documents before sending it to the cloud
4. The Partner employee sends a keyword (such as the name of the document that request) to a specific computer in the organization containing the repository of digital documents
5. The computer storing the digital document returns a token
6. The Partner employee sends the token to the cloud
7. The Cloud Provider uses the token to find the appropriate encrypted document and returns it to the Partner employee
8. At any point in time, Organization's Data Verifier can verify the integrity of Organization's data

This sequence of steps are illustrated in Figure 4.4. The information security services provided for that model are confidentiality and integrity during the storage of the documents in the cloud. Whereas that work is focused on cover the inherent risks of storage sensitive documents on a public cloud, does not cover a document lifecycle. Also, access control is not considered in that model.

Chieu *et al.* [5] proposed a SDMS for contracts designed as an enterprise web application service. The core components of that SDMS are the access control component, the task execution engine responsible for the document lifecycle, an email notifiicator, a signature engine, an encryption engine, the document management component responsible for the organization, tracking, and storage of contract and a document search component. These components and the overall system architecture are illustrated in Figure 4.5.

The access control component is strongly related with the search component, since in the XML digital documents the access control information is inserted and then, this information is directly incorporated into the secure search-index. Each time a user requests the execution of a query,

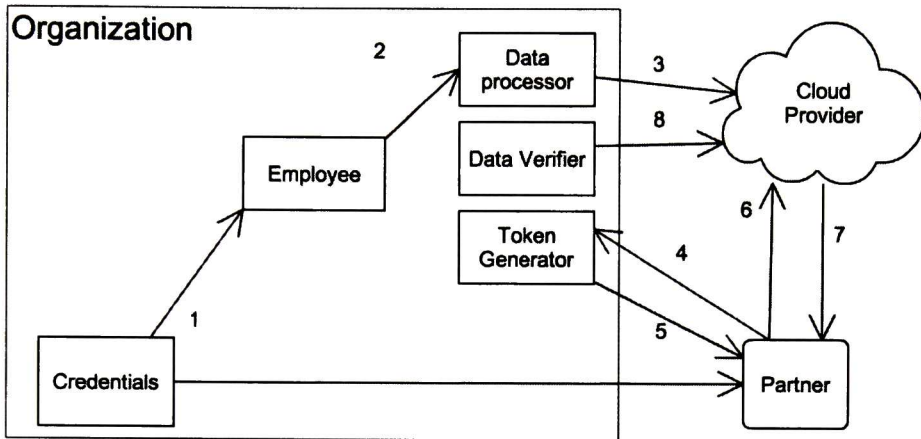


Figure 4.4: Model for a Secure Document Management System with cloud storage proposed by Kamara and Lauter [4]

unauthorized documents for that user will not be considered in the results set. To effectively utilize the secure search-index to search for authorized documents, a compound query generation mechanism is also incorporated in the search component to join the user profile information in the search query. This allows to hide the existence of highly sensitive digital documents protecting the organization's confidentiality. Hiding the existence of digital documents to unauthorized users is exposed as the main characteristic of that system, and it is stated that this mechanism makes that system more secure than those that perform a post-filtering of the result sets. However, this statement is not supported by any evidence in the paper that presents that work.

Zhao *et al.* [96] presented a scheme for a SDMS based in three-layer structure and hardware symmetric encryption. The three-layer software architecture defines the most general components of the system, which is divided into a presentation layer, a business logic layer and a data access layer. On the other hand, hardware encryption is the core of security in that scheme, cryptographic smart cards are used in client terminals, and encryption cards are used in servers to secure the communication between clients and the server. When users upload or download digital documents, client subsystems use their cryptographic smart cards to encrypt or decrypt digital documents. In

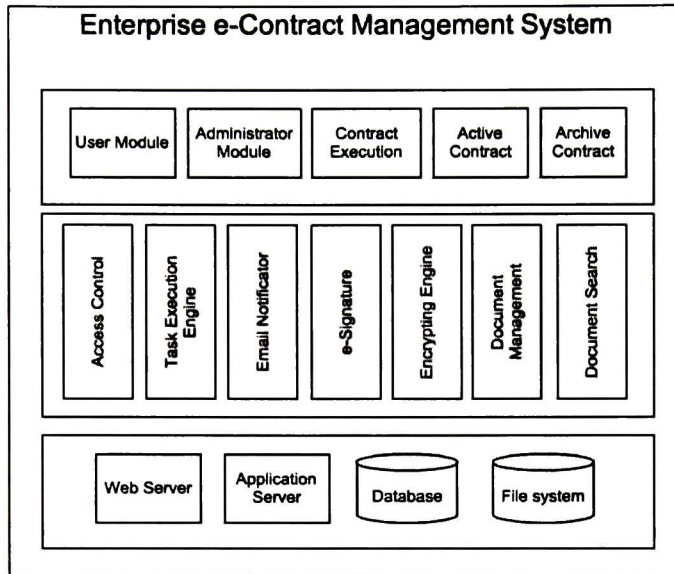


Figure 4.5: Architecture of an enterprise content management system for secure contracts proposed by Chieu *et al.* [5]

the same way, when the server distributes digital documents, it uses an encryption card to encrypt digital documents before they are transmitted.

Odagiri *et al.* [97] implemented a system based on Destination Addressing Control System (DACS) for the distribution of copyrighted documents in schools. That system implements access control through VPNs. When users require access to a digital document via HTTP using a Web browser, the system determines which digital documents are able to be accessed by users using their IP previously assigned, using Network Address Translation (NAT) to route the request to the appropriate server. This technique has the advantage of being able to revoke access to the digital document dynamically.

Bai [98] and Sevenic *et al.* [99] have proposed the use of the structure of XML documents to embed access control mechanisms by specifying a language for the definition of privacy policies that can restrict the actions per user. These solutions need a special software that interprets the XML files with the embed access control. Hence, that kind of solution must be complemented with other

type of security information techniques such as digital signatures to provide the integrity service in order to avoid a direct change or removal of the access control by malicious users.

Munier *et al.* [17] have proposed another similar approach, which encapsulates a set of security mechanisms in the digital document that is delivered to the user, so that digital documents are able to apply the policies associated with the user who manipulates them. These documents are defined as autonomous files and define their own file format. Furthermore, that scheme allows embed queryable information in digital documents. Since that work uses a non-common format, its application in a production environment can be limited when the file format used by an organization can not be changed.

Adobe [12, 16] provides a solution for a SDMS that is based on a PKI using digital documents in PDF format. Similar to the Munier work, that SDMS works for digital documents represented in a format that is defined by the creator of the SDMS with the difference that PDF is a wide stablished format. Most of the information security services provided by adobe are included in the definition of a PDF standard, allowing the use of access control, the expiration of the digital documents issued by their applications making them inaccessible after a certain time and to update the documents avoiding the generation of outdated versions. That work covers general document lifecycles, but customization is not supported and its implementation could derive in vendor locks, that in a long term they can be a risk if better solutions would be needed.

A comparative of the reviewed works on SDMSs against the proposed SDMS in this thesis is provided in Chapter 6 in Table 6.3.

4.2 User Tracing for Digital Documents

While there are several works on user tracing in multimedia content, illegal distribution of documents by traitor users is a problem that recently is becoming relevant for organizations, therefore, user tracing techniques for digital documents are being incipiently explored.

Brassil *et al.* [6] proposed a fingerprinting system based on the word shifting watermarking technique to protect distribution of documents that are delivered by subscription to a specific user. That system supports scenarios where the access to digital documents is restricted but their content is not sensible. That system receives as input a Postscript file and renders a bitmap as output in which a user fingerprint is inserted as a version of the digital document. The insertion of the user fingerprint is shown in Figure 4.6.

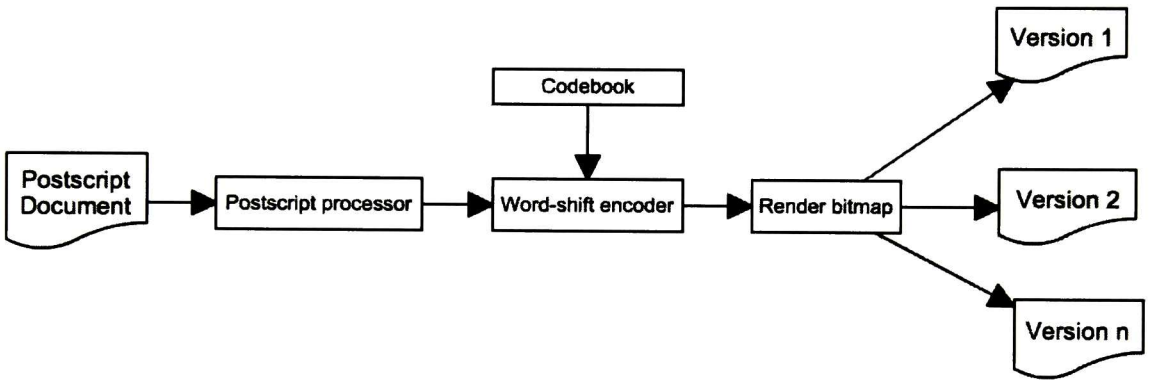


Figure 4.6: Scheme of the encoder proposed by Brassil *et al.* in [6].

When a pirate document is found in physical format, it is scanned and processed in order to reduce any noise that could be introduced. Then, the document in digital format is processed by a decoder that extracts in a non-blinded fashion the document version x that is the fingerprint. The detection of the user fingerprint is shown in Figure 4.7.

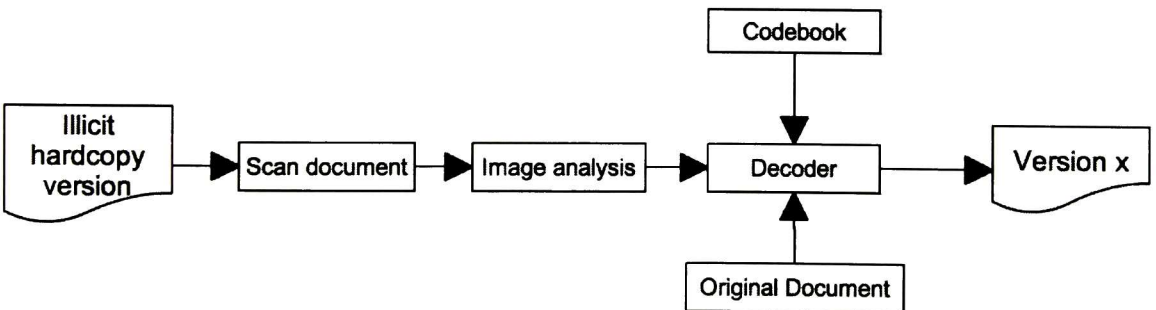


Figure 4.7: Scheme of the decoder proposed by Brassil *et al.* in [6].

Some disadvantages reported on that system are that fingerprints can be easily removed and malicious users can detect them without having the original digital document because that system is based in a watermarking scheme for digital documents with well-known vulnerabilities such as word shifting performed by malicious users[61]. Also, despite of being robust to analogic transformations, that work does not report resistance to collusion attacks.

Darwish [7] proposed the use of fingerprinting in XML documents based on the Winoing algorithm [100]. To generate a fingerprint, it is necessary to decompose the XML document in m partitions based on a secret key K_s . Then, the hash value of each tag in each partition is computed with a one-way function and the Winoing algorithm is applied generating the fingerprint by a subset of the hash codes and a user secret key. White space characters are inserted at the end of the tags of the XML document to insert the fingerprint. When a pirate document is detected, the pirate XML is split using K_s allowing to extract the fingerprint and compare it with all the known user's fingerprint. The process of insertion and detection of fingerprints is shown in Figure 4.8.

Darwish's work is resistant to collusion attacks, however, it is not able to survive to trivial attacks such as the deletion of all the inserted white spaces that represents the user's fingerprints.

Schick and Ruland [8] proposed the trace of digital documents as a new security service. The approach of that work consists on combining a digital signature σ of a digital document m with the ID named as Tracking Data (TD) of the users that manipulate a digital document. The digital document m is encrypted each time it is transmitted. Then, when m is decrypted by a user U_k , his/her TD is added to the signature as extra data with σ . Then, all extra data is signed by U_k generating a new signature σ_k that will be appended to m , σ and the extra data that represents the previous user TDs. The process of appending TD to m is shown in Figure 4.9. If a pirate copy of a digital document is detected, it is possible to traceback all the users that previously manipulates that document, considering that the last TD belongs to the traitor user. The ideas presented in that work are a recent approximation to the integration of information security services and user tracing. Schick and Ruland [8] proposed the trace of digital documents as a new security service.

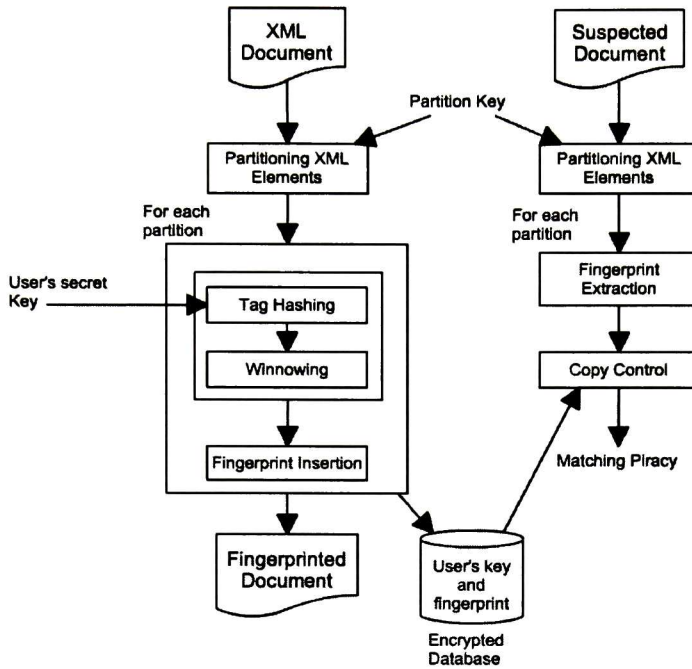


Figure 4.8: Scheme of the encoder and decoder proposed by Darwish in [7].

The approach of that work consists on combining a digital signature σ of a digital document m with the ID named as Tracking Data (TD) of the users that manipulate a digital document. The digital document m is encrypted each time it is transmitted. Then, when m is decrypted by a user U_k , his/her TD is added to the signature as extra data with σ . Then, all extra data is signed by U_k generating a new signature σ_k that will be appended to m , σ and the extra data that represents the previous user TDs. The process of appending TD to m is shown in Figure 4.9. If a pirate copy of a digital document is detected, it is possible to traceback all the users that previously manipulates that document, considering that the last TD belongs to the traitor user. The ideas presented in that work are a recent approximation to the integration of information security services and user tracing.

That work is useful when it is necessary to traceback not only the traitor user, but all the previous users that distributed the digital document to him/her. However, for that work it was not reported resistance to collusion.

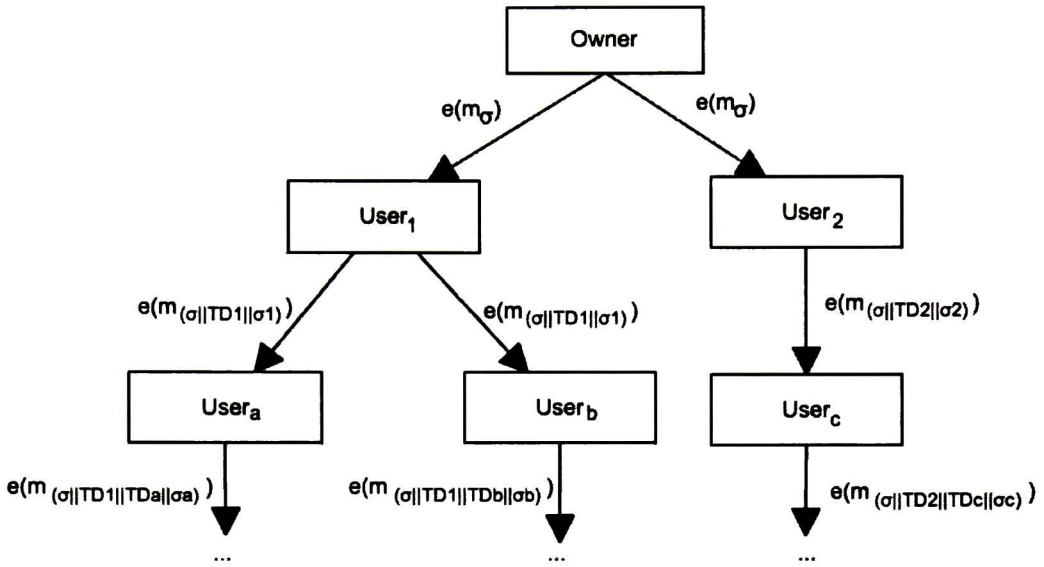


Figure 4.9: Accumulation of user IDs that allow tracing back a traitor according to the work of Schick and Ruland [8].

4.3 Analysis

Most of the reviewed papers for SDMS rely on a PKI to provide information security services [1, 2, 3, 12, 16, 94], which denotes the effectiveness of PKIs. Other common characteristic of some of the reviewed works is that a considerable part of them manage the digital documents in a XML format [5, 95, 99], taking advantage of the capacity of adding nodes without affecting the structure of the digital documents. However, that approach is not appropriated for this thesis since the problems addressed in this thesis consider image-based digital documents. The most significant difference among the reviewed works is the implementation context, since the application scenario is different in each work, the proposed solutions are also customized. A general solution for a SDMS is the one proposed by Adobe [12, 16], but even for that work there are scenarios that cannot be covered such as customized document lifecycles [101].

On the other hand, after a wide review of recent papers on user tracing in the context of digital

documents, it was found that just a few works have been reported. The works focused on user tracing in digital documents are based in digital fingerprinting techniques (these techniques were reviewed in Chapter 2), aiming to provide a solution based on digital fingerprinting. Only the work proposed for Schick and Ruland [8] reports resistance to collusion attacks. However, the application scenario considered in that work uses digital documents in XML format, whereas a solution for image-based digital documents is needed.

Reviewed works do not provide a solution to the problem addressed in this thesis, but it is possible to retake some ideas of these works for the development of a secure document management system that allows user tracing.

4.4 Summary

In this Chapter, the most relevant works focused on provide information services through secure document management systems were reviewed and the first efforts to provide a solution to trace traitor users that distribute pirate documents. It was determined that reported works have presented custom solutions according to an specific application scenario, so they are not suited for the problem addressed in this thesis work. Also, it was found that all the proposed works to trace users in the context of digital documents are based on fingerprinting, but those solutions are neither robust to collusion attacks nor appropriated for the problem scenario in this thesis.

At the moment of developing this thesis work, it was not found in the literature an integral solution for the illegal access of digital documents from unauthorized users and the illegal distribution of documents by authorized but dishonest users. In Chapter 4 an integral solution to these problems is proposed.

5

Proposed Secure Document Management System

This Chapter describes the Secure Document Management System (SDMS) proposed in this thesis. The SDMS is composed by two main modules: a module that provides information security services and a module based on the paradigm of digital fingerprinting that allows the dishonest user tracing. Both modules incorporate some ideas of related work presented in Chapters 2, 3 and 4. Also, the integration of the Information Security Services Module with the Fingerprinting Module in the SDMS is described.

5.1 Information Security Services Module

In the first Chapter of this thesis, it was defined the required information security services for the stages of Approval and Distribution for a SDMS, these are: authentication, confidentiality and integrity. Also, it was defined that during the Storage stage the required services are: confidentiality, integrity, non-repudiation and access control. The stages and the information security services to be provided are shown in Figure 5.1, and the user types and their functions are shown in Table 5.1.

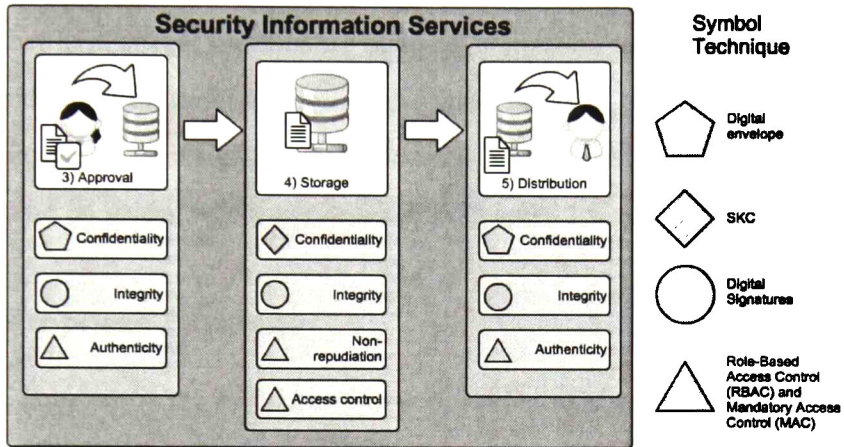


Figure 5.1: Selected techniques to provide the required information security services in the proposed Secure Document Management System

Function/User type	Consumer	Reviewer	Auditor	Administrator
SDMS Access	•	•	•	•
Access all documents			•	
Access documents by department	•	•	•	
Upload document		•		
Encrypt stored document		•		
Download document	•	•	•	
Decrypt stored document	•	•	•	
Generate valid signature		•		
Signature validation			•	
Register users				•
Delete users				•

Table 5.1: Relation of user types and functions in the proposed Secure Document Management System.

The works reported in the literature reviewed provide information security services using similar techniques, however, the main difference lies in the context of the implementation. Whereas there are works that have already provided the required information security services, there are not works suitable for the problem tackled in this thesis, where the main restriction is that digital documents are digital images. The design of the Information Security Services Module (ISS Module) for the SDMS is based on several techniques reviewed in Chapter 2.

5.1.1 Proposed Strategy

The SDMS is based on the Encrypted Server-Based Document Management System architecture, since it brings a simple but secure way to access to the SDMS. Despite of the Cloud Computing Environment-Based Document Management System could be more flexible, the lifetime of digital documents is considered for this thesis about two years, so older documents can be stored off-line in external storage media, and no highly scalable infrastructure would be needed [39]. Also the Terminal-Based Document Management System was discarded since a more flexible access control than MAC is needed in an implementation scenario. Finally, Hardware-Based Encryption Document Management System was discarded to avoid vendor locks.

5.1.1.1 *Proposed Strategy for Secure Transmission of Digital Documents*

The identity of sender and receiver is validated using a mutual authentication mechanism to provide the authentication service, when a digital document is uploaded and downloaded. The confidentiality service between the sender and the receiver will be provided by the implementation of a digital envelope, in order to take advantage of fast encryption in SCK used in this work for bulk encryption of digital documents. By means of the digital envelope, the symmetric key used by the SCK algorithm is securely distributed, encrypting it using an ACK algorithm. Integrity of digital documents is ensured by validating a digital signature. When Reviewer user uploads a digital document to the SDMS, the ISS Module will validate the digital signature of that user on the uploaded digital document. When

a Consumer user downloads a digital document, the ISS Module will validate the digital signature of the downloaded digital document.

5.1.1.2 Proposed Strategy for Secure Storage of Digital Documents

The ISS Module will encrypt the stored digital documents preventing unauthorized access to them. If an unauthorized user tries to obtain a digital document directly from a Digital Document Repository (DDR), he/she would not be able to access its content. The ISS Module would be responsible for encrypting and decrypting the digital documents using a pre-defined secret key. Digital documents uploaded to the SDMS must be encrypted and stored with the signature of the Reviewer user that uploaded them. Based on the nature of the documents stored in the DDR, updates are not required after they are stored. In case of a correction in the physical document, the related stored digital document must be deleted in the SDMS by an Administrator user, and the new one will be loaded as a new digital document. Since the signature of Reviewer user that uploaded a digital document is stored in the DDR, an Auditor user is able to validate that signature, avoiding that the Reviewer users deny they approved that digital document. That user is the only one with read and write privileges in the location where the encrypted digital documents are stored.

5.1.2 Access Control

Access control mechanisms based on RBAC ensure that users perform only allowed actions to them (see Table 5.1) and also that the DDR is only accessed by the SDMS through the ISS Module. The RBAC Mechanism is applied per module allowing them to be self-contained in an object oriented fashion. The access to the DDR will be protected by a Mandatory Access Control (MAC) where the SDMS is executed by a user that is defined in the execution infrastructure.

5.1.3 Public Key Infrastructure

PKI is required in this work for implementing the ACK algorithms used in the digital envelope and digital signature schemes. The proposed topology of the SDMS is shown in Figure 5.2. The SDMS will be deployed in a Local Area Network (LAN), and its main components are an encryption server in which the SDMS will be deployed, a Digital Document Repository (DDR) stores the digital documents, a Certification Authority (CA) that will generate the digital certificates and also will take the role of a Registration Authority (RA), and the clients of the system that will be users in the roles of Consumer, Reviewers, Auditors and Administrators.

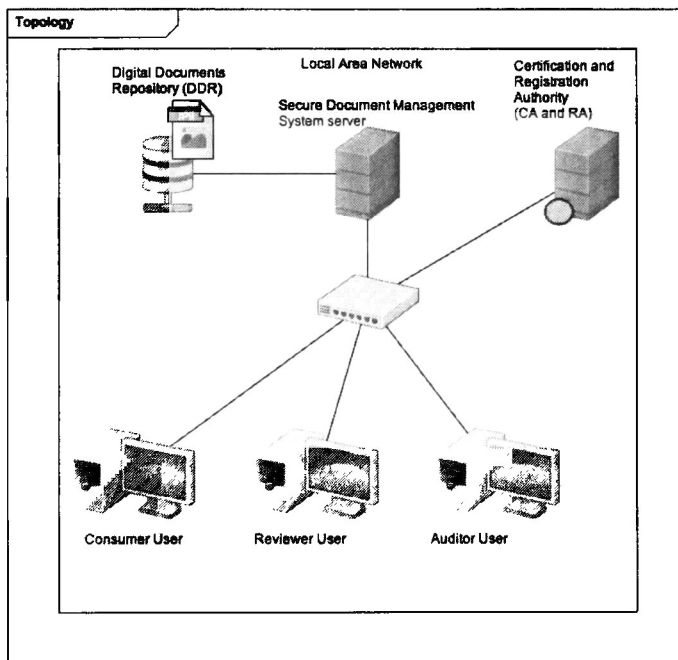


Figure 5.2: Topology for the proposed Secure Digital Document Management System.

5.1.3.1 *Digital Certificate Generation*

The digital certificate generation occurs when the Administrator user registers a new user. The new users will be responsible of the generation of their private and public keys that will be provided to an Administrator user along with contact data for the generation of the certificate for the new user. A user U generates its public key UP_u and private key UP_r that are sent to the Administrator user A , that access CA for the generation of C validating that the owner of those keys has the proper rights in the system through a RBAC mechanism. When C is generated, it is signed by the CA and stored in a certificate repository. Finally, C is sent to U which validates the CA signature in C to conclude successfully the certificate generation. The certificate generation process is graphically described in Figure 5.3.

5.1.3.2 *Certificates Revocation*

The certificate of a user could be revoked only by an Administrator user. However, the certificate revocation is not done arbitrarily, but when the user is logically deleted. The certificates will not be deleted, but they will remain stored in order to validate signatures of Reviewer users that do not belong to the organization anymore.

5.1.3.3 *User Authentication*

The method for user authentication in the system is carried out by mutual authentication. Each time a user requests an action to the SDMS, his/her identity and the server identity will be validated through the ISS Module.

5.1.4 **Secure Digital Documents Approval**

During the Approval stage of the document lifecycle, a Reviewer user signs an approved digital document. Then, the digital document and its signature are encrypted with a secret random key, that

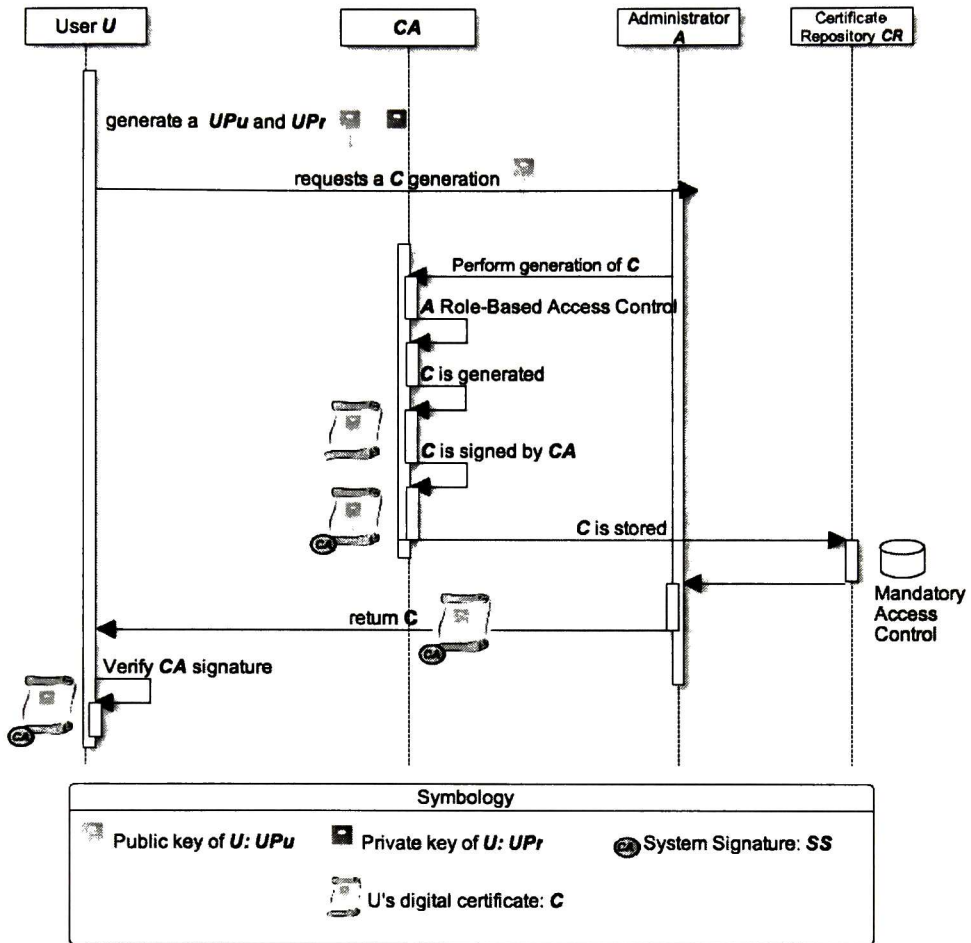


Figure 5.3: Digital Certificate generation process.

is packaged in a digital envelope for the SDMS. The digital document, the signature and the digital envelope are uploaded to the SDMS that validates the user permissions using a RBAC mechanism. After that, the digital envelope is opened and with the extracted secret key, the digital document and the signature are decrypted. The digital document in clear and the signature are encrypted using a secret key that belongs to the ISS Module, and both are stored in a digital document repository, that is accessed by the ISS Module through the MAC mechanism validating the rights of the SDMS to write in the repository. The process of the secure Approval, with all the cryptographic key involved,

are shown in the sequence diagram in Figure 5.4.

5.1.5 Secure Digital Documents Storage

When a user requests to store a digital document, that digital document is encrypted, and when the document is retrieved, it is decrypted. Both actions requested by the user are validated by a RBAC mechanism. For the case of signature validation, an Auditor is able to request the signature validation of a Reviewer over a digital document. In order to do this, the ISS Module gets the digital document along with its related signature from the digital document repository and decrypt them. Then, the signature is validated and the validation result is encrypted using a secret random key that is packaged in a digital envelope for the Auditor. The encrypted validation result is signed by the SDMS and sent to the Auditor along with the digital envelope. Finally, the Auditor validates the SDMS signature and extracts the secret key from the digital envelope allowing to obtain the result validation in clear. The process of the signature validation of digital documents in the SMDS, with all the cryptographic key involved, is graphically shown in the sequence diagram in Figure 5.5.

5.1.6 Secure Digital Documents Distribution

When a user requests a digital document, that action is validated by a RBAC mechanism. Then, the encrypted digital document is obtained from the digital document repository and it is decrypted with the secret key of the SDMS. After that, the digital document is encrypted using a secret random key, and the the digital document is signed by the SDMS. The secret random key is packaged in a digital envelope for the user, and the digital document along with the digital envelope are sent to the user which validates the SDMS signature. Then, the digital envelope is opened and with the extracted secret random key and the digital document is decrypted. The process of the secure digital document distribution, with all the cryptographic key involved, is graphically shown in the sequence diagram in Figure 5.6.

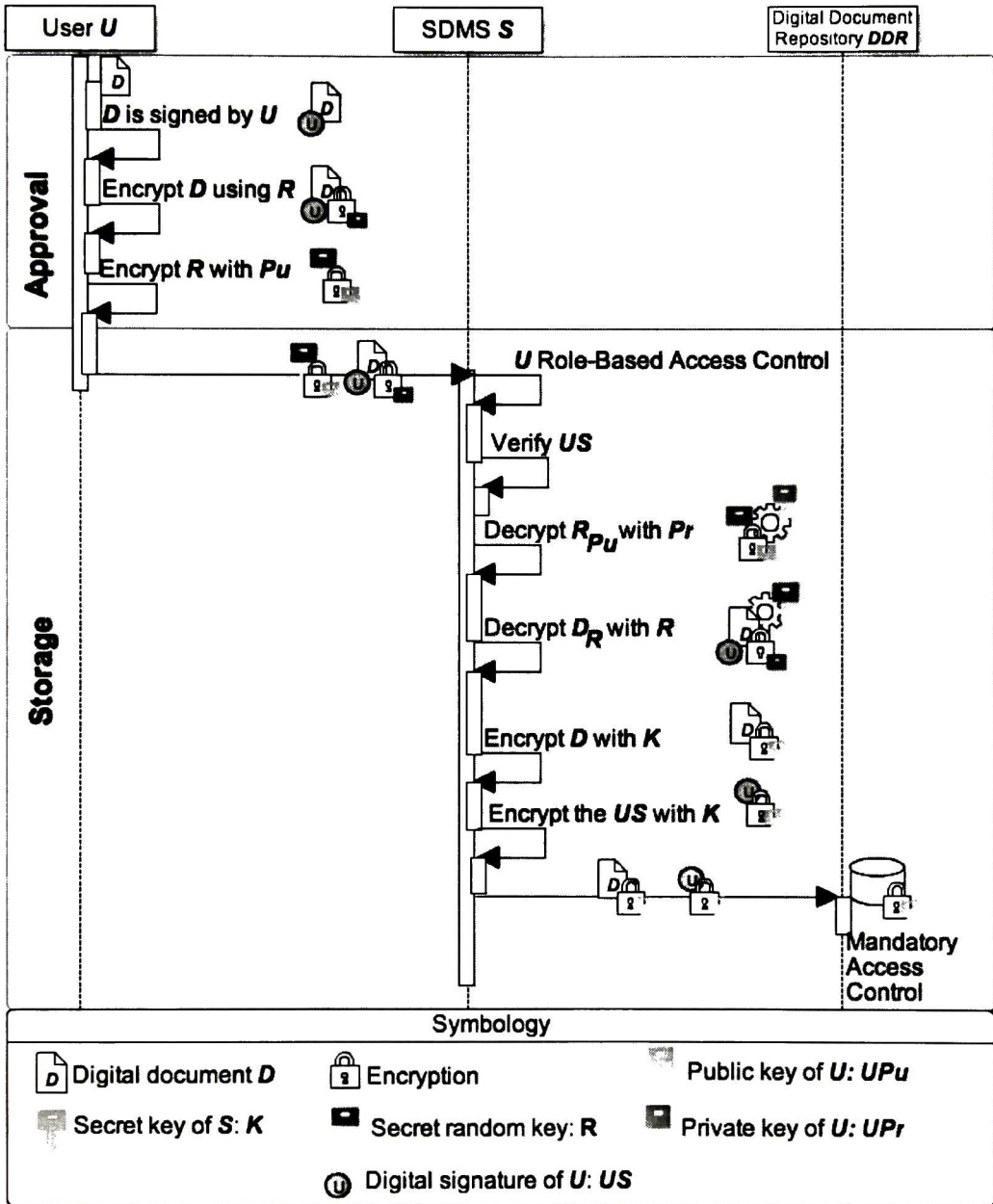


Figure 5.4: Sequence diagram of the Approval stage in the proposed Secure Document Management System.

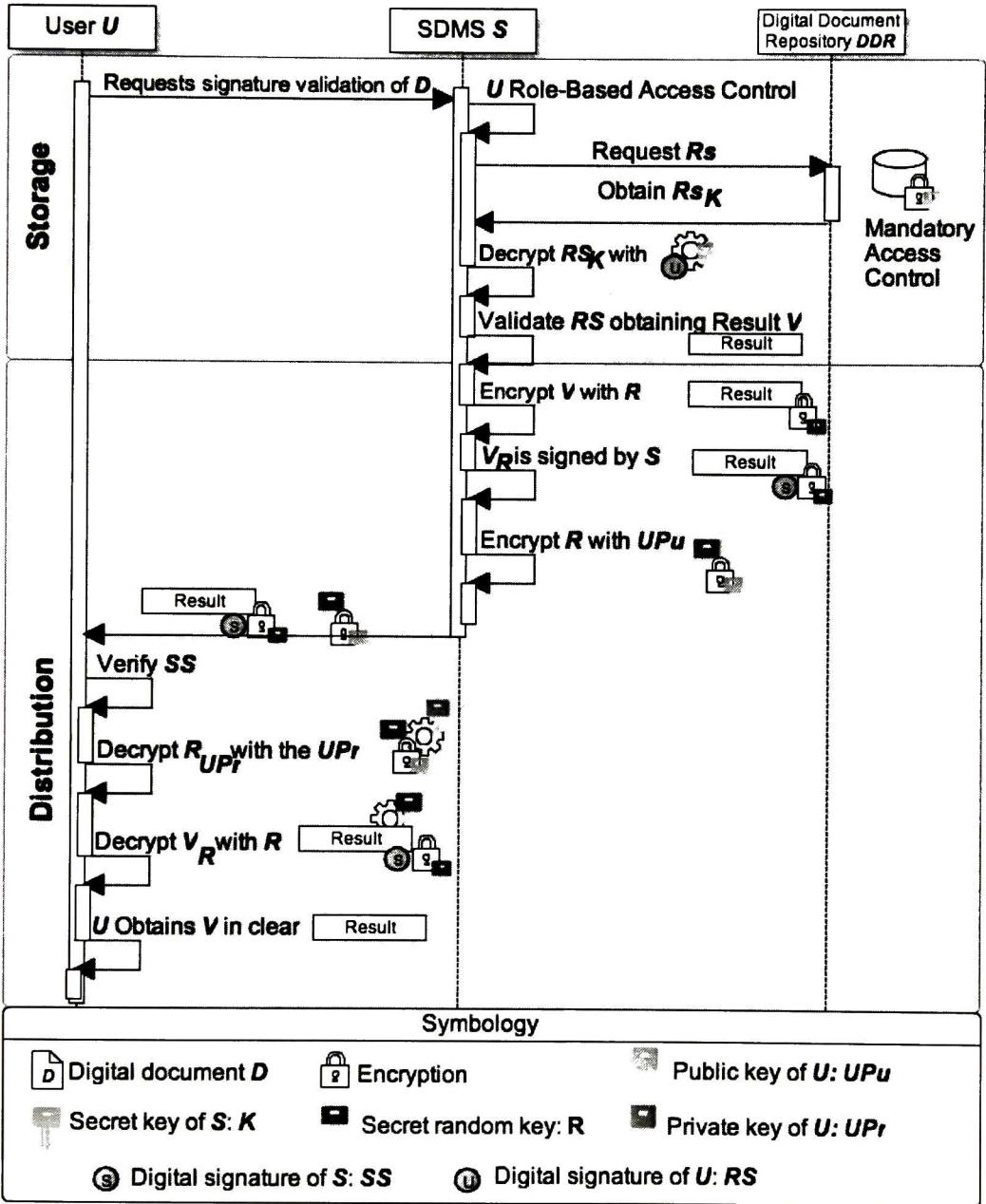


Figure 5.5: Sequence diagram of the signature verification process of digital document in the proposed Secure Document Management System.

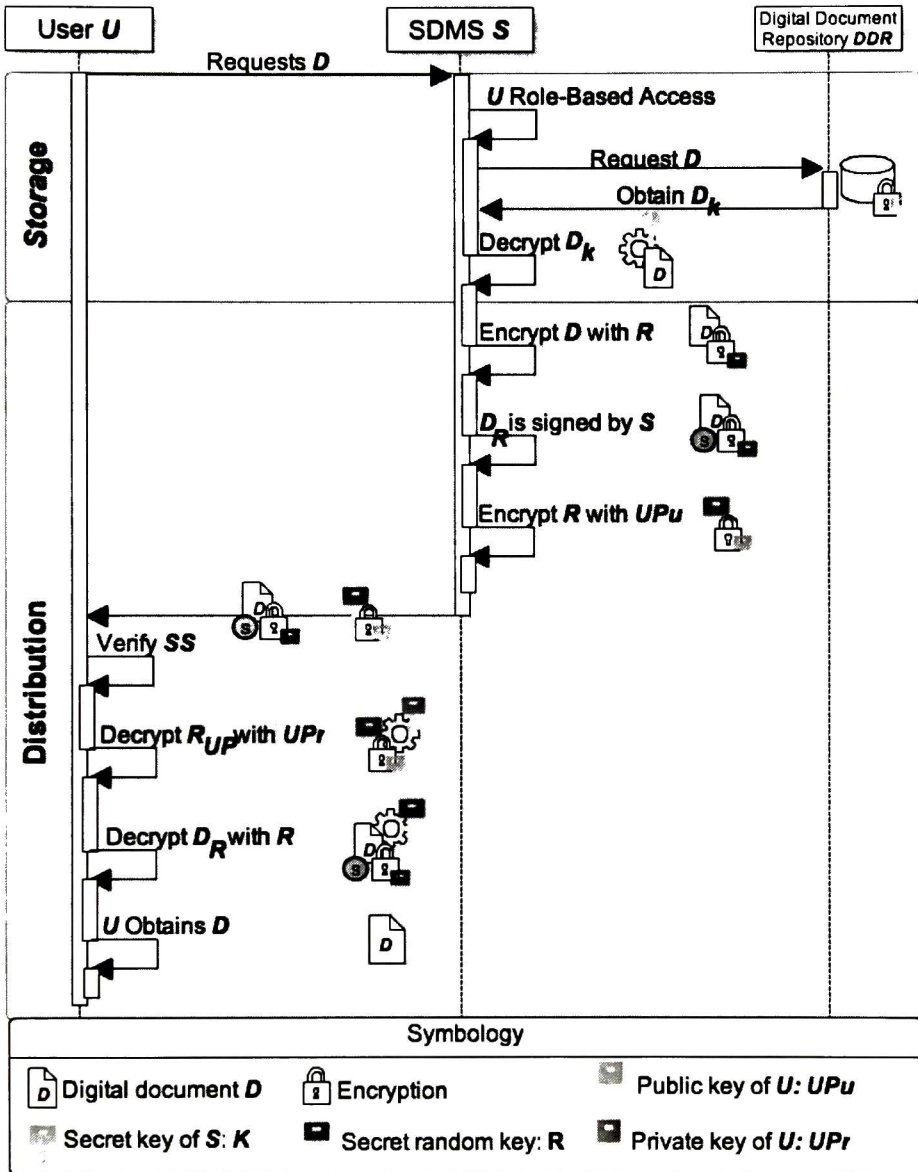


Figure 5.6: Sequence diagram of the Distribution stage in the proposed Secure Document Management System.

5.1.7 Selection of Cryptographic Algorithms

Three cryptographic algorithms were selected for implementation of the ISS Module based on their characteristics, presented in Chapter 2. These algorithms are:

- **Hash algorithm:** Since SHA algorithms have proved to be more robust and are established as a standard, these hash algorithm family is chosen for the integrity security service of the ISS Module.
- **Symmetric Key Cryptographic algorithm:** Due to AES is the current standard of the NIST and there are no successful attacks to break it, this algorithm was selected to provide the security information service of confidentiality in the ISS Module.
- **Asymmetric Key Cryptographic algorithm:** Due to AES is the current standard of the NIST and there are no successful attacks to break it, this algorithm was selected to provide the security information service of confidentiality in the ISS Module.

Having defined AES as the SKC algorithm, RSA as the AKC algorithm and SHA as hash function, the length of the keys in those algorithms must be defined. A SDMS is a system to be used for long terms, and during that time, the system must be secure. The key lengths define the resistance of cryptographic algorithms to attacks and they are established estimating the period of time the system must protect the digital documents. For the implementation of the SDMS in this work, the period of time for documents protection is 30 years. Estimating the key length according to the NIST [102], the key length for SKC algorithms must be at least 112 bits. Considering that AES as a minimum key length of 128 bits, this will be the minimum key length used for this algorithm. For RSA algorithm, NIST recommends to use a key length of 2048 bits as minimum so that size is used in the proposed ISS Module.

5.1.8 Implementation

The ISS Module was implemented as a Web application. It was coded in Java language in a desktop computer with Ubuntu 12.10 as operative system, an Intel Core i5 processor at 2.7GHz and 4GB RAM, mounted in an Apache Tomcat server version 7.0.42. The functionality of the PKI defined in Chapter 5.1.3 for the ISS Module was implemented by the key and certificate management utility of Oracle Keytool [103].

5.1.8.1 *Cryptographic Object Storage*

Users can access to the SDMS through a Web browser, in which the certificate approved by the CA must be imported along with the private key in a file with a PKCS#12 format [104], which is a standard format proposed by the RSA laboratories to store cryptographic objects. The PKCS#12 file is required to support mutual authentication configured in the Apache Tomcat server. The SDMS can not be accessed if the user certificate is not registered in the browser. In the server side, the certificate repository was created using a file with JKS format [103], instead of the PKCS#12 because JKS is the format allowed by the Apache Tomcat Server and proposed by Oracle.

5.1.8.2 *Secure Sockets Layer*

In order to implement the selected techniques in Chapter 5.1.4 and Chapter 5.1.6 for the secure transmission of the digital documents from users to the SDMS and viceversa, the Secure Sockets Layer (SSL) protocol was configured in the Apache Tomcat server. Different from digital envelope implementations, SSL uses a key exchange sharing a secret key at the beginning of the communication (in this case using RSA) instead in each message. Despite of this implementation differs with the design, security services are still provided. The SSL protocol was configured in the server to establish a secure channel using RSA, AES, and SHA-2, as defined in the design of the ISS Module. Figure 5.7 shows a high view of the interaction of the users and the SDMS and the components involved.

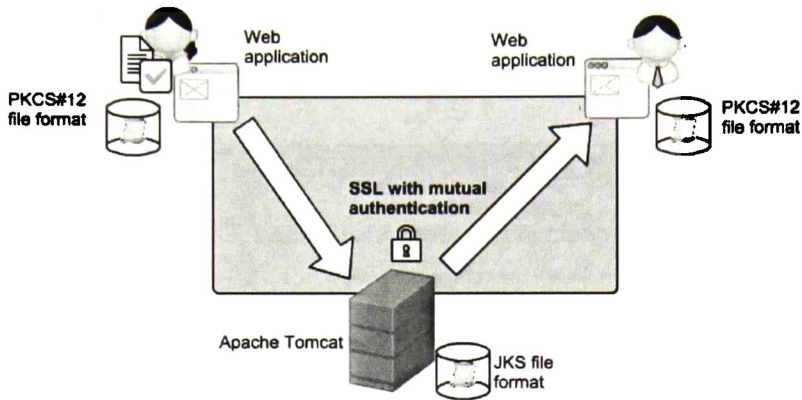


Figure 5.7: Upload and download scheme of the SDMS implementation.

5.1.8.3 Software Components

When a Reviewer uploads a digital document, a file with its signature is uploaded too. A desktop application was developed to sign digital documents. The application receives as input the PKCS#12 file of the Reviewer, the username, password and the digital document to be signed. Then the application is able to generate the digital document signature.

For the secure storage of digital documents and their signatures, the authenticity of the digital document is revalidated as part of fine-grained security mechanisms in the SDMS. Then, the User Module is loaded for the authenticated user, with the settings related to its type (Consumer, Reviewer, Auditor or Administrator). These settings are username, department, and the location of the repository that can be accessed by the user. Then, functionality of the Encryption and Signature Modules can be invoked by the authenticated user, and both modules perform RBAC validation. The Encryption and Signature Modules use AES, RSA and SHA-2 algorithms, which were provided by the `javax.crypto` package. The SDMS is executed by a user that has the proper rights to access the encrypted repository of digital documents. The high level view of the SDMS and the submodules that compose the ISS Module is shown in Figure 5.8. The way the submodules are related was inspired in the general architecture of the SDMS proposed by Chieu, *et al.* in [5].

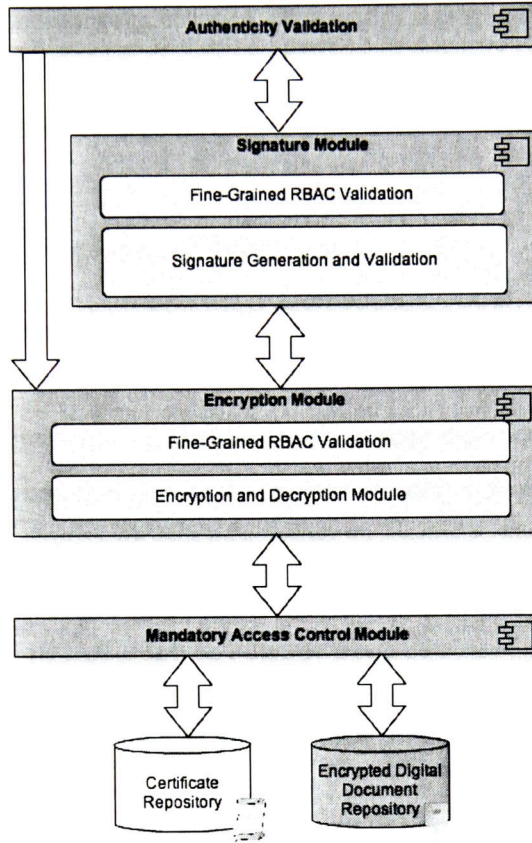


Figure 5.8: Model of the software implementation in the SDMS server with the submodules of the Information Security Service Module.

5.2 Fingerprinting Module

After the review of the related works for user tracing, it was found that all the works reported to provide this service are based on fingerprinting. However, all these works have well-known vulnerabilities, that make them inadequate to be implemented as the solution for the problem of distribution of pirate documents by traitor users. Although, other fingerprinting techniques can be used to solve this problem. An important criteria to select the fingerprinting technique for the

Fingerprinting Module is the resistance to collusion, because this characteristic allows to determine the identity of traitor users involved in the creation of the pirate document. In the next sections, a fingerprinting technique is selected for the Fingerprinting Module.

5.2.1 Fingerprinting Technique Selection

Different watermarking insertion techniques that can be used for fingerprinting of digital documents have been proposed (see Chapter 3). These techniques can be syntactic techniques, semantic techniques and image based techniques [61, 105]. In this work digital documents are represented as images, therefore, image based techniques are chosen as basis to develop the Fingerprinting Module. However, reported watermarking techniques for image-based digital documents are not robust to collusion attacks, so watermarking techniques for multimedia, in particular for digital images are explored. Due to the robustness to collusion attacks and furthermore the reduction of computational time detecting colluders, the scheme proposed by Kuribayashi was selected to be the basis of the Fingerprinting Module developed in this thesis. This scheme groups users by region, social circumstances or another characteristic can be used for this purpose. In the scenario addressed in this thesis work, users are already segmented by department, hence, this scheme not only meets the objectives established for this thesis, but the user grouping occurs naturally. The following Sections describe the generation, insertion and detection of fingerprints of the selected scheme.

5.2.2 Fingerprint Generation

The user's fingerprint is composed of a SS sequence that identifies the user and another sequence that identifies his group membership. These sequences are based on direct sequence's SS , which is generated by a pseudo-random sequence PN that modulates a carrier signal. The SS sequence for a group i is generated from a vector V of length L with all entries to 0 adding an amount of energy β_g to the entry at position i . Subsequently, the inverse DCT (IDCT) is applied to V to obtain the i -th

basis function of the DCT. Finally, V is modulated by a PN sequence generated from a secret key s , which provides security to the scheme because only the one who knows the key is able to detect groups. The SS sequence W_i that is generated for the i -th group is represented by the Equation 5.1. Each component in the spread spectrum sequence for the group ID can be assigned to a group; therefore, the total amount of groups supported is L .

$$W_i = PN(s) \otimes IDCT(i, \beta_g) \quad (5.1)$$

Generation of sequence W_j from a user j belonging to a group i is performed similar to the sequence W_i of the group, with the difference that the PN seed is given by the group ID. The sequence assigned to the j -th user is represented by the Equation 5.2.

$$W_j = PN(i) \otimes IDCT(j, \beta_u) \quad (5.2)$$

With the SS sequences of user and its group, the user fingerprint is generated as defined in Equation 5.3.

$$W_{i,j} = PN(i) \otimes IDCT(j, \beta_u) + PN(s) \otimes IDCT(i, \beta_g) \quad (5.3)$$

The resulting energy of the fingerprint is represented by $\beta^2 = \beta_g^2 + \beta_u^2$. By using the group ID as the seed of the PN, a link is established between the group and the user. As the number of groups and users per group is L , the total amount of users supported is L^2 .

5.2.3 Fingerprint Insertion

The following steps describe the fingerprint insertion method and are graphically shown in Figure 5.9:

1. Generate a user fingerprint $W_{i,j}$ as described in Equation 5.3.

2. Transform the input image into the frequency domain using the DCT function.
3. Select L coefficients of the low and middle frequencies from a position P_w . The selected coefficients are denoted as:

$$V = \{v_0, v_1, \dots, v_{L-1}\} \quad (5.4)$$

4. Insert the fingerprint additively in the extracted coefficients:

$$\hat{V} = V + W_{i,j} \quad (5.5)$$

5. Transform the image to the spacial domain using the IDCT function in order to get the fingerprinted image.

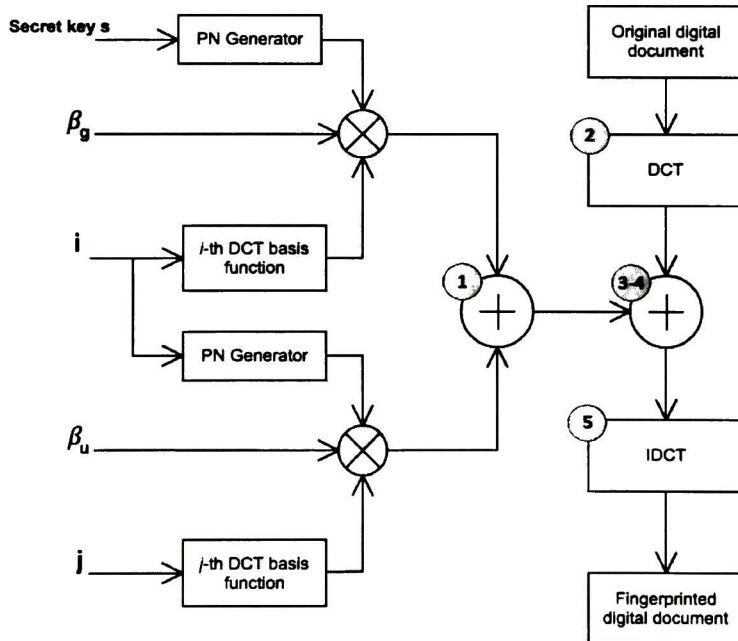


Figure 5.9: Diagram of fingerprint insertion method for the proposed fingerprinting module.

5.2.4 Fingerprint Detection

The following steps describe the detection of the user's fingerprint in an illicit copy of an image representing a digital document and are illustrated in Figure 5.10:

1. Transform the illicit copy into the frequency domain using the DCT function.
2. Select L coefficients of the low and middle frequencies from the position P_w . The selected coefficients are denoted as:

$$\hat{V} = \{\hat{v}_0, \hat{v}_1, \dots, \hat{v}_{L-1}\} \quad (5.6)$$

3. Detect the group ID :

(a) Generate PN using the secret key s .

(b) Use the DCT function to extract the detection sequence \hat{d}_g :

$$\hat{d}_g = \text{DCT}(\text{PN}(s) \otimes (\hat{V} - V)) \quad (5.7)$$

(c) Calculate the variance σ_g^2 of \hat{d}_g considering the probability distribution and determine the threshold T_g from a given false-positive denoted as Pe_g :

$$T_g = \sqrt{2\sigma_g^2 \text{erfc}^{-1}(2Pe_g)} \quad (5.8)$$

where $\text{erfc}^{-1}(\cdot)$ stands for the inverse complementary error function defined in Equation 5.9 and σ_g^2 is computed as shown in Equation 5.10.

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt \quad (5.9)$$

$$\sigma_g^2 = \frac{1}{n} \sum_{\hat{d}_{g,k} \in D_g} (\hat{d}_{g,k} - \overline{\hat{d}_g})^2 \quad (5.10)$$

(d) If \hat{d}_g in the input k exceeds the threshold T_g , it is determined whether k is the group ID.

4. Detect the user ID:

(a) Generate a PN using the ID k of the detected group.

(b) Use the DCT function to extract the detection sequence \hat{d}_u :

$$\hat{d}_u = \text{DCT}(\text{PN}(k) \otimes (\hat{V} - V)) \quad (5.11)$$

(c) Calculate the variance σ_u^2 of \hat{d}_u in a similar way as in Equation 5.10, considering the probability distribution and determine the threshold T_u from a given false positive denoted as Pe_u :

$$T_u = \sqrt{2\sigma_u^2} \text{erfc}^{-1}(2Pe_u) \quad (5.12)$$

(d) If \hat{d}_u in the input h exceeds the threshold T_u , it is determined that h is the user ID.

5.2.5 Software Implementation

The Fingerprinting Module was implemented in C++ language using the GCC compiler version 4.2.1 with Ubuntu 12.10 OS as operative system, an Intel Core i5 processor at 2.7GHz and 4 GB RAM. To perform the DCT and IDCT transforms the FFTW library version 3.3.3 was used [92]. In order to read and write images two libraries were used: libtiff version 3.6.1 [106] for TIFF images and jpeglib version 8.0 [107] for JPEG images. Finally the IQA library version 1.1.2 was used to compute

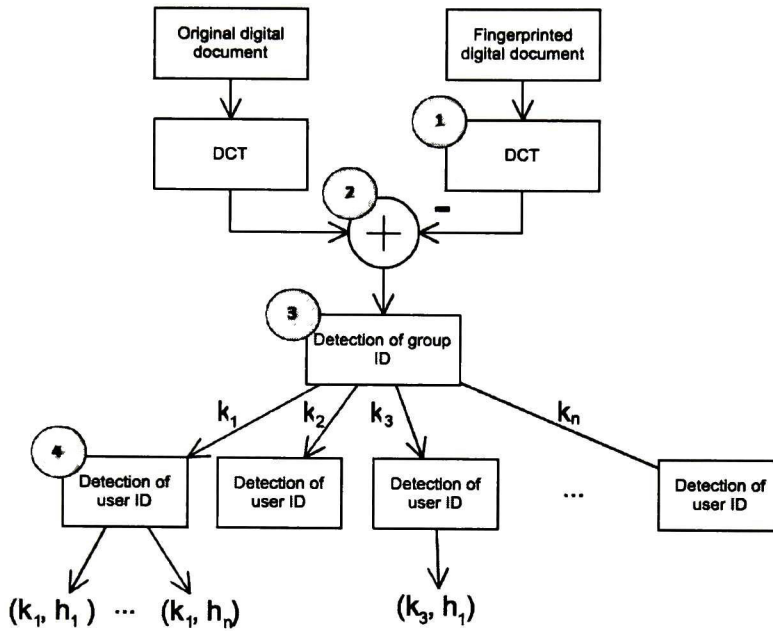


Figure 5.10: Diagram of fingerprint detection method for the proposed Fingerprinting Module.

the PSNR and SSIM Index values[108]. The Fingerprinting Module is able to read and write digital documents in JPEG and TIFF format.

5.3 Integration

This section describes the integration of the ISS and Fingerprinting Module to create the SDMS system proposed in this thesis.

5.3.1 General System

The integration of the ISS Module and the Fingerprinting Module required to adapt the model of the software implementation illustrated in Figure 5.8. Due to the secure document transmission is independent of the insertion and detection of fingerprints, no adjustments were needed in this part.

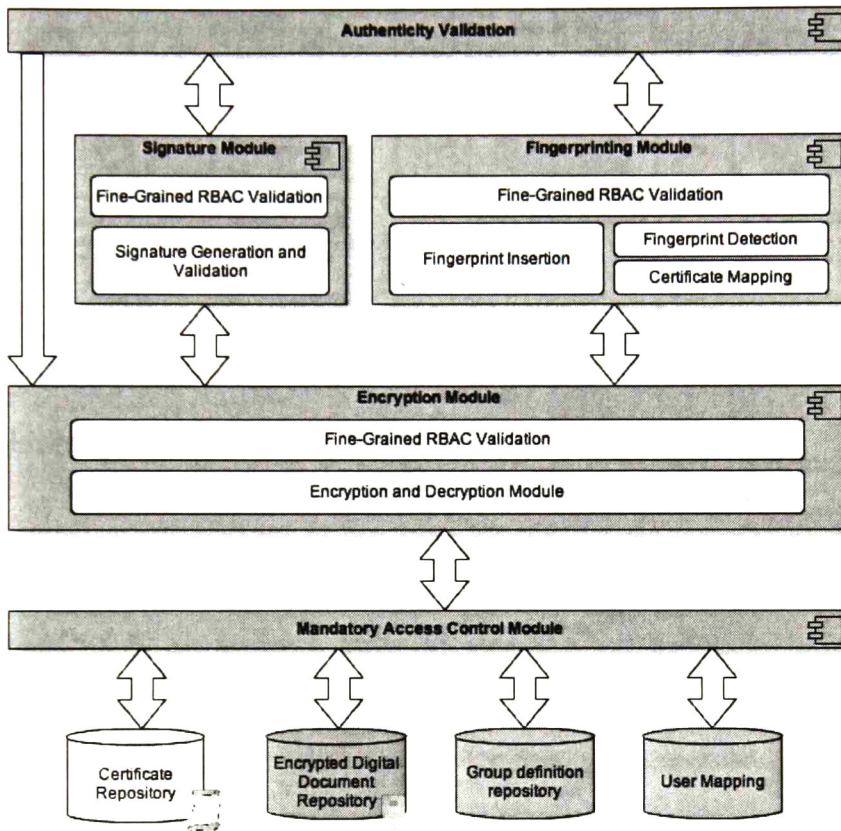


Figure 5.11: High level view of the SDMS server integrating Information Security Services and User Tracing.

The adjustments were made in the server side of the SDMS to integrate the Fingerprinting Module with the ISS submodules as it is illustrated in Figure 5.11. Only authorized users can invoke the Fingerprinting Module and the insertion and detection of fingerprints can be only invoked for certain users passing through a RBAC mechanism. The Encryption Module provides to the Fingerprinted Module digital documents in clear and then encrypts them again before sending them to the final users.

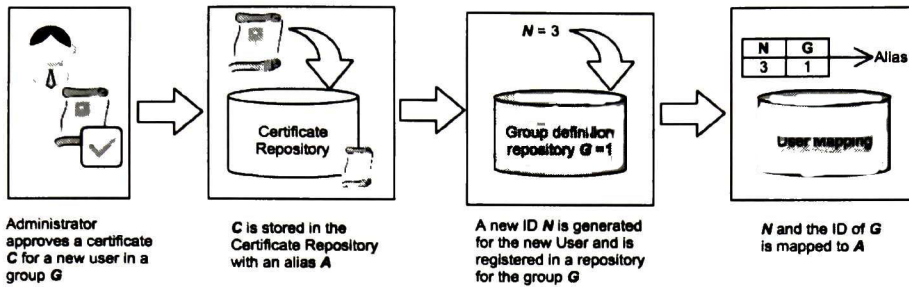


Figure 5.12: Certificate - User ID mapping process.

5.3.2 User Identity

The most important consideration for the integration is that the user identity must be preserved between modules, that is: an authorized user in the ISS Module must have a representation in the Fingerprinting Module, and viceversa. In the ISS Module, a user is identified by means of its digital certificate, whereas in the Fingerprinting Module identification is by means of a user ID. Because each module has a different way to represent the user identity, it is necessary a component that maps the user identity between the ISS and Fingerprinting modules. Also, it is necessary to define the user groups of the Fingerprinting Module. Since no digital documents are shared among Departments in the SDMS, Departments are the best option to be mapped to user groups.

The mapping component requires a persistence mechanism to store the identity mappings. Two repositories were defined: one for user mapping and another for group definition. This mechanism must be protected from access of unauthorized users. If the persistence mechanism is destroyed or altered by malicious users, it will not be possible to insert fingerprints or detect them anymore. Therefore, this is the most critical component of the integration. Similar to the Digital Document Repository access, a MAC validation will be performed each time the mapping mechanism is invoked.

The registration of a new mapping will be performed when a certificate is approved and a new user is added. A new user ID will be generated per certificate and the generated user ID will belong to an area defined by the certificate data. The mapping process is illustrated in Figure 5.12.

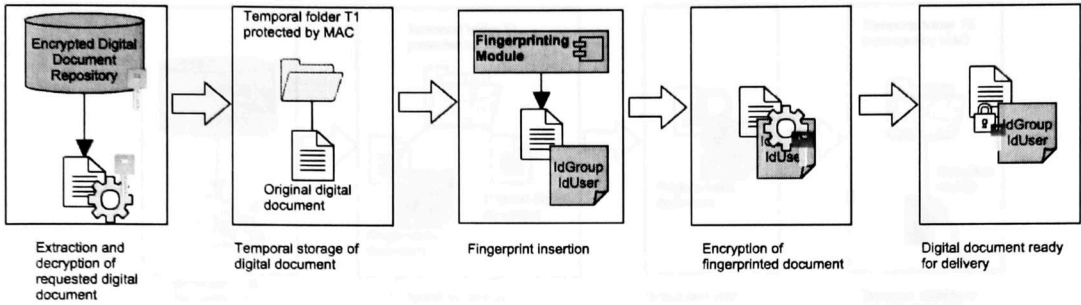


Figure 5.13: Secure Fingerprint Insertion process.

5.3.3 Secure Fingerprint Insertion

The fingerprint insertion must be done each time an authorized user U requests a digital document D to the SDMS in the process shown in Figure 5.6. Once the ISS Module obtains and decrypts a D copy from the DDR, the fingerprint of U will be inserted in D . Then, the downloading process of the fingerprinted document D will continue.

Since the Fingerprinting Module requires digital document in clear to perform the insertion, information security services are required during the insertion operation of the Fingerprinting Module. The digital document in clear is stored temporarily in the DDR protected by MAC. Once the Fingerprinting Module gets the content of the digital document, the document is deleted. In the same way, when the Fingerprinting Module generates a fingerprinted document, the content of that document is accessed by the ISS Module and deleted. The secure fingerprint insertion is shown in Figure 5.13.

5.3.4 Secure Fingerprint Detection

Only Administrator users are able to perform the detection of fingerprints in the Fingerprinting Module. Detection of fingerprints requires two digital documents: the original digital document in clear and its fingerprinted pirate copy. Since original digital documents stored are encrypted, the ISS Module performs the decryption of the original digital document and provides it to the Fingerprinting

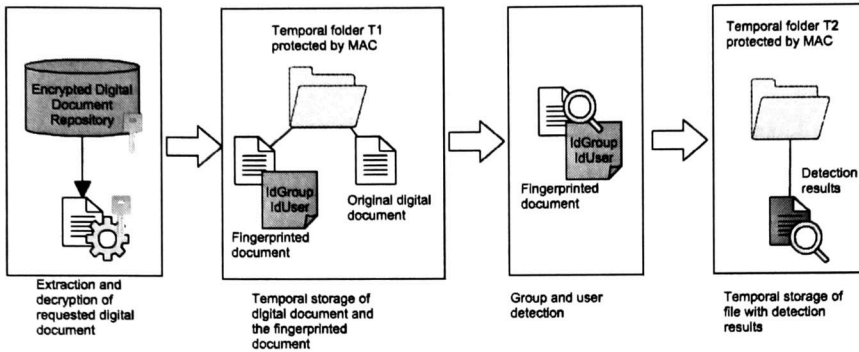


Figure 5.14: Generation of the detection results file after the fingerprint detection.

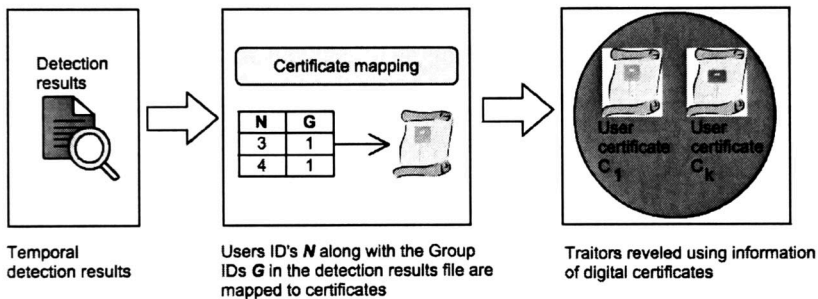


Figure 5.15: Mapping user IDs to digital certificates.

Module. Once the detection of fingerprints has finished, a file with user and groups IDs detected is provided to the Mapping component, which reveals the identity of the traitor users by querying the Certificate Repository. After the ISS Module accesses the generated results file, the file is deleted. This process is graphically shown in Figures 5.14 and 5.15.

In the Fingerprinting Module, just the group that have access to the digital document will be considered as suspicious, and only the ID's assigned to each user in that group will be searched in the fingerprinted pirate copy.

5.4 Summary

This Chapter described the proposed SDMS and its two main modules: an ISS Module to provide information security services to digital documents during the Approval, Storage and Distribution stages of the lifecycle of digital documents, and a Fingerprinting Module to trace traitor users that distribute pirate documents. The ISS Module was designed to avoid that unauthorized users access to digital documents and the Fingerprinting Modules was designed to trace users from pirate document copies and furthermore to generate fingerprinted documents robust to collusion attacks. The integration of the ISS and the Fingerprinting module was designed taking into account the need to preserve the identity of the user between modules. This is needed because identity of users is expressed by different means in each module. The implementation results and evaluation of the integrated systems SDMS is presented in the next Chapter.

6

Results of the Proposed Secure Document Management System

This Chapter presents the Implementation results of the Information Security Services Module (ISS Module), the Fingerprinting Module, and the integration of these two modules in a Secure Document Management System. Results of experimentation are provided showing that the proposed SDMS can be implemented in a production environment.

6.1 Information Security Services Module

With the purpose of validate the correctness of the ISS Module, the operations an performance of this module were validated and evaluated.

6.1.1 Validation

The sequences of approval, distribution and signature validation of digital documents presented in Sections 5.1.4, 5.1.5 and 5.1.6 of Chapter 5 were validated by testing the possible execution paths obtaining successful results in each case. Furthermore, unit tests on the Signature Module and the Encryption Module were performed to validate that the result of each function was the expected for a set of given entries. The unit testing results are shown in Table 6.1.

User actions	Result
Signature validation by an Auditor user over digital document in Department A	Successful
Decryption of Digital documents in Department A by a Consumer user of Department B	Successful
Encryption of digital documents in Department A by a Reviewer user of department A	Successful
Access to digital documents in Department A by a Consumer user of department B	Denied
Encryption of digital documents in Department A by a Reviewer user of department B	Denied
Encryption of digital documents in Department A by a user of department A without privileges	Denied
Access to digital documents by a user without a digital certificate (no authenticated)	Denied
Signature validation of digital documents in Department A by a user of department A without privileges	Denied
Signature validation of corrupted digital documents	Failed
Signature validation of digital documents with corrupted signatures	Failed

Table 6.1: Tests performed for validation of the Encryption and Signature modules.

Also, browsers were enabled to establish a secure channel between the server and the users by using the SSL protocol and mutual authentication during the Approval and Distribution stages of the document lifecycle.

With the validation of the execution paths and the encryption and signature modules, the ISS Module proved that it provides the required security services required for the system developed in this thesis. Then, a performance evaluation was carried out to determine if the ISS Module was able to be implemented in a production environment.

6.1.2 Performance Evaluation

The most time consuming tasks in the ISS Module are those related to the encryption and decryption of digital documents, along with the read and write operations. The execution time of these functions must be feasible in a production environment. To determine if that is achieved, performance of encryption and decryption were evaluated over 30 digital documents with similar characteristics that are defined in Table 6.2, in a desktop computer with Ubuntu 12.10 OS, with a processor Intel Core i5 CPU at 2.7GHz and 4 GB RAM. This evaluation considered the server operations and the transmission of digital documents.

Characteristic	Value
Format	JPEG
Width (Px)	1900
Height (Px)	2700
Size (MB)	1.1
Number of pages	1

Table 6.2: Characteristics of digital documents used to evaluate the Information Security Service Module.

6.1.2.1 Digital Documents Server Operations

Encryption time was obtained along with the timing of write operations since these tasks are performed always together (see Figure 5.4 in Chapter 5). It was found that the average execution time of these tasks is 924.93ms. In the same way, decryption time was obtained along with the timing of read operation (see Figure 5.6 in Chapter 5) resulting an average execution time of 918.0ms. Similar results were found for the encryption and decryption of the digital document's signature. Regarding the timing for signature verification, that timing involves two decryption operations, one for the digital document and another for decrypting its associated digital signature. This way, signature verification of digital documents is more time consuming, requiring 2015.97ms in average. However, if decryption is not considered, signature verification requires 179.97ms in average.

6.1.2.2 Digital Documents Transmission

When the digital document is transmitted from the server to the client, the operations of encryption and decryption are performed by the SSL protocol. The transmission time was obtained by downloading the digital documents using the browser Firefox version 24.0¹ with the Firebug plugin² that computes the download time. The digital documents were downloaded from the server in which the system was implemented to the same server in order to avoid networking delays. Since the SSL protocol was considered as a black box, encryption in the server side, decryption in the client side and operations of the protocol were evaluated all together. The average time of all these operations performed during the transmission of digital documents is 231.32ms.

6.1.3 Comparison

Previous works in the literature (see Chapter 3) have considered the provision of most of the security services required by a SDMS. These works are compared against the SDMS developed in this thesis work in Table 6.3, in which the information security services of each work are shown along with their maintainability that is considered as the capacity of a SDMS to be enhanced by a user.

For the scenario addressed in this thesis, the works proposed in [1], [2], [3], [99] and [98] do not protect the digital documents in an adequate way when they are stored in the Storage stage due to integrity, confidentiality and nonrepudiation are not provided. The work presented in [96] and [95] are oriented to face two specific scenarios: contact management and hardware encryption respectively. These works present a lack of information security services required for the problematic presented in this thesis. Other similar cases occur with the works presented in [97], [17] and [4]. On the other hand, the work proposed by Adobe [16] provides all the required information security services during the distribution and storage of digital documents. However, the security mechanisms proposed in [16] cannot be modified to implement new schemes for confidentiality, integrity, etc, therefore,

¹<https://www.mozilla.org/en-US/firefox/24.0>

²<https://addons.mozilla.org/es/firefox/addon/firebug>

Work	Transmission			Storage				Maintainable
	C	I	A	C	I	X	N	
Casey <i>et al.</i> [1]	•	•	•			•		•
Liroy <i>et al.</i> [2]	•	•	•			•		•
Gerasimov [94]	•	•	•					
Shi and Ouyang [3]		•	•					
Bai [98]			•			•		
Odagiri <i>et al.</i> [97]	•	•	•	•	•	•		
Zhao <i>et al.</i> [96]	•		•	•				•
Kwok and Nguyen [95]		•	•		•		•	
Munier <i>et al.</i> [17]	•	•		•	•	•		•
Kamara and Lauter [4]	•		•	•	•	•		•
Adobe [16]	•	•	•	•	•	•	•	
Chieu <i>et al.</i> [5]	•	•	•	•	•	•	•	•
Proposed work	•	•	•	•	•	•	•	•

Table 6.3: Comparison of the proposed Information Security Services Module against representative works in the literature. Information Security Services: Confidentiality (C), Integrity (I), Authenticity (A), Access control (X), and Non-Repudiation (N).

security updates depend on the vendor. In contrast with the Adobe solution, our proposal isolates the implementation of security mechanisms in submodules that can be modified without affecting Other submodules, for example, a new Encryption Module implementation can be achieved without affecting the Signature Module. Another work that provides all the required information security services is the one presented by Chieu *et al.* [5]. However, that work was designed with several modules to support the document lifecycle of contracts and security modules adapted to contract management process. Despite of the general architecture of the work presented in [5] it cannot be fully reused since it considers modules for contracts lifecycle, some ideas of that work were taken for the organization of submodules in this thesis work, as mentioned in the software implementation section of the ISS Module.

6.2 Fingerprinting Module

Perceptual transparency and resistance to collusion attacks are properties that must be ensured for the Fingerprinting Module, as it was indicated in Chapter 3. The next sections describe the performed experiments to determine the values of PSNR and SSIM Index metrics in fingerprinted digital documents to achieve perceptual transparency. Then, the best configuration of parameters for the robustness factor of user (β_u), robustness factor of group (β_g), fingerprint length (L) and insertion position (P_w) are identified for the maximum number of colluder detection for lossless and lossy digital documents.

6.2.1 Perceptual Transparency

Perceptual transparency must be satisfied by the Fingerprinting Module to allow retrieving information of digital documents. To achieve this property, adequate values of β_u , β_g , L and P_w were determined through experimentation using as input a set of 1000 digital documents³ with similar characteristics that are defined in Table 6.4. Statistical significance is achieved for this sample size as random t -tests results (for significance level equal to 0.05) showed a p -value about 9.9514×10^{-209} for the biggest.

Characteristic	Value
Format	JPEG
Width (Px)	1900
Height (Px)	2700
Size (MB)	1.1
Number of pages	1

Table 6.4: Characteristics of digital documents used to evaluate the Fingerprinting Module.

³These documents were obtained from Biblioteca Digital del Ateneo de Madrid <http://goo.gl/EloO6>.

6.2.1.1 Sets of Parameters Evaluation

Values of β_u , β_g , L and P_w were combined as defined in Table 6.5 to generate fingerprints that were inserted in the set of 1000 digital documents. Parameter's range was set to generate from highly transparent fingerprinted documents to totally distorted fingerprinted documents. The perceptual transparency was evaluated using the PSNR and SSIM Index metrics.

Parameter	Initial value	Increment	Maximum value
β_u	50,000	50,000	1,200,000
β_g	50,000	50,000	1,200,000
L	50,000	50,000	1,200,000
P_w	1/6	1/6	5/6

Table 6.5: Configuration parameters for fingerprint insertion.

The insertion position P_w showed a minimum impact on the perceptual transparency. When varying P_w while fixing the maximum values of β_g , β_u and L , the PSNR values vary less than one decibel, and only a few tens for SSIM Index. These variations are shown in Figures 6.1 and 6.2. The decreasing of the SSIM Index and PSNR in the maximum value of P_w can be attributed to the contrast between the background and characters in the digital document, generating abrupt color changes that increase the energy in the high frequencies region. Therefore, altering the values in this region can lead to characters distortion. The little variation of SSIM Index and PSNR values is attributed to the generated distortion that is allocated near to the characters, but the uniform background of the digital document mitigates the distortion. In a similar way, the parameter L also was not a relevant factor introducing distortion to the digital document.

The values of β_g and β_u were identified as the most important due to they affect directly the perceptual transparency. Figures 6.4 and 6.3 show the impact of varying these parameters with the maximum values of P_w and L . PSNR and SSIM Index values present a similar behavior when β_g and β_u vary, so for digital documents, the values of SSIM Index can be estimated from the values of PSNR and viceversa. This is consistent with the findings of Hore and Ziou [84].

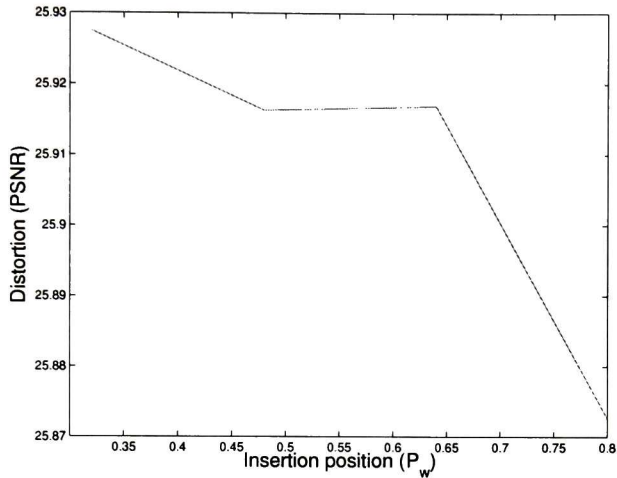


Figure 6.1: Distortion (PSNR) generated in digital documents due to fingerprint insertion varying the insertion position (P_w), with fixed values of robustness factor assigned to users (β_u), groups (β_g) and fingerprint length (L). Low PSNR values implies high distortion.

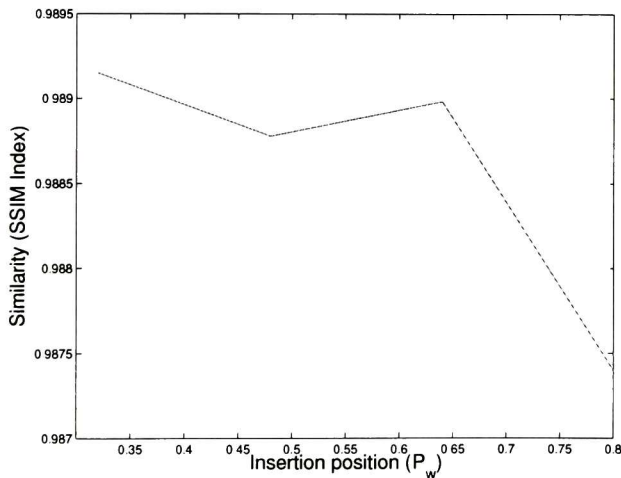


Figure 6.2: Similarity (SSIM Index) between an original document and its fingerprinted copy varying the insertion position (P_w) with fixed values of the robustness factor assigned to users (β_u), groups (β_g) and fingerprint length (L). High SSIM Index values implies a high similarity.

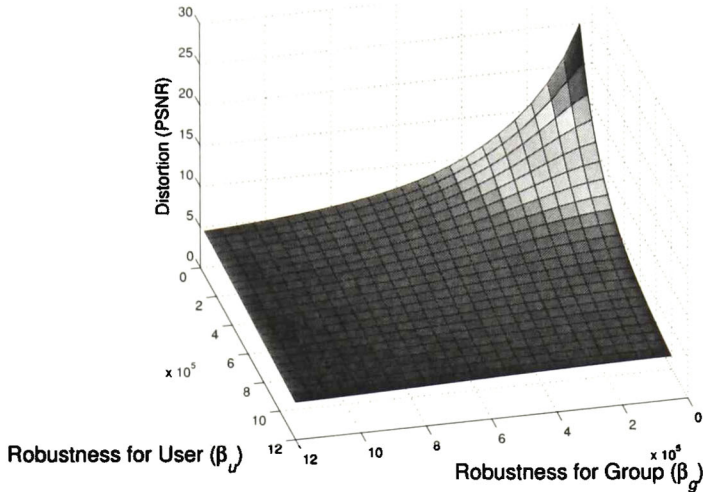


Figure 6.3: Distortion (PSNR) generated in digital documents due to fingerprint insertion varying the robustness factor for users (β_u) and groups (β_g), with fingerprint length (L) and insertion position (P_w) fixed.

The lost of quality in the digital documents with the reduction of the PSNR and SSIM Index values are shown in Figures 6.5a, 6.5b, 6.5c, 6.5d, 6.5e and 6.5f. For the characteristic of perceptual transparency, the values of β_g and β_u are interchangeable, due to distortions are given by the total energy that is defined by Equation 5.3 during the design of the Fingerprinting Module.

Under a subjective assessment, fingerprinted documents with PSNR = 30dB and SSIM Index = 0.996 are easily readable and distortion cannot be perceived. The PSNR value is consistent with the distortion tolerated for natural images reported in previous works [109, 110]. It is possible to reduce even more the PSNR and SSIM Index values since it is considered that perceptual transparency of fingerprints in digital documents is achieved if the documents are legibles. However, minimum values of PSNR and SSIM Index have not been reported. Hence, these values are determined in this thesis work. Reducing the PSNR and SSIM Index values will allow to increase the robustness factors B_u and B_g .

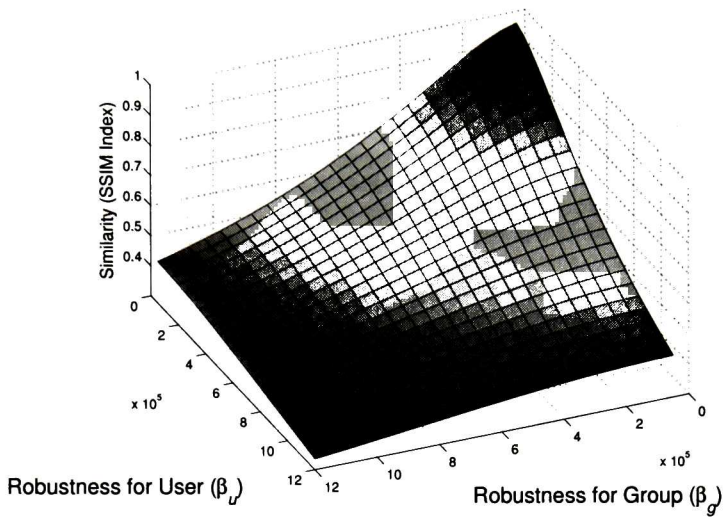


Figure 6.4: Similarity (SSIM Index) between an original document and its fingerprinted copy varying the robustness factor for users (β_u) and groups (β_g), with fingerprint length (L) and insertion position (P_w) fixed.

6.2.1.2 Inquest Evaluation

To validate the PSNR and SSIM Index values that meet perceptual transparency for digital documents, an inquest was applied to 100 respondents. The inquest consisted in the evaluation of digital documents with a PSNR value in a range of 4dB to 30dB. The possible answers available for the respondents were the following:

1. I do not perceive distortion in the digital document.
2. I perceive distortion in the digital document but the text is easily readable.
3. I perceive distortion in the digital document and the text is hardly readable.
4. The digital document is not readable.

The results of the inquest are shown in Table 6.6 and plotted in Figure 6.6 showing the change in the perception of the respondents while the PSNR of digital documents decreases. Most of the

Abandonemos estas consideraciones generales, y pasemos á la lectura de las cartas de que al comenzar hemos hablado. Estas cartas van dirigidas á un antiguo camarada de colegio, Cristian Sothe, hombre que ocupó más tarde cargos importantes en la magistratura prusiana, y de carácter recto y severo, que por su seriedad y buen sentido era llamado desde muy joven Staatarrath (consejero de Estado).

Heine describe el aprecio y estima en que le tiene diciendo:

(a) Original digital document.

Abandonemos estas consideraciones generales, y pasemos á la lectura de las cartas de que al comenzar hemos hablado. Estas cartas van dirigidas á un antiguo camarada de colegio, Cristian Sothe, hombre que ocupó más tarde cargos importantes en la magistratura prusiana, y de carácter recto y severo, que por su seriedad y buen sentido era llamado desde muy joven Staatarrath (consejero de Estado).

Heine describe el aprecio y estima en que le tiene diciendo:

(b) Fingerprinted document using SSIM Index = 0.99 and PSNR = 30.28dB.

Abandonemos estas consideraciones generales, y pasemos á la lectura de las cartas de que al comenzar hemos hablado. Estas cartas van dirigidas á un antiguo camarada de colegio, Cristian Sothe, hombre que ocupó más tarde cargos importantes en la magistratura prusiana, y de carácter recto y severo, que por su seriedad y buen sentido era llamado desde muy joven Staatarrath (consejero de Estado).

Heine describe el aprecio y estima en que le tiene diciendo:

(c) Fingerprinted document using SSIM Index = 0.97 and PSNR = 22.16dB.

Abandonemos estas consideraciones generales, y pasemos á la lectura de las cartas de que al comenzar hemos hablado. Estas cartas van dirigidas á un antiguo camarada de colegio, Cristian Sothe, hombre que ocupó más tarde cargos importantes en la magistratura prusiana, y de carácter recto y severo, que por su seriedad y buen sentido era llamado desde muy joven Staatarrath (consejero de Estado).

Heine describe el aprecio y estima en que le tiene diciendo:

(d) Fingerprinted document using SSIM Index = 0.92 and PSNR = 16.68dB.

Abandonemos estas consideraciones generales, y pasemos á la lectura de las cartas de que al comenzar hemos hablado. Estas cartas van dirigidas á un antiguo camarada de colegio, Cristian Sothe, hombre que ocupó más tarde cargos importantes en la magistratura prusiana, y de carácter recto y severo, que por su seriedad y buen sentido era llamado desde muy joven Staatarrath (consejero de Estado).

Heine describe el aprecio y estima en que le tiene diciendo:

(e) Fingerprinted document using SSIM Index = 0.87 and PSNR = 14.21dB.

Abandonemos estas consideraciones generales, y pasemos á la lectura de las cartas de que al comenzar hemos hablado. Estas cartas van dirigidas á un antiguo camarada de colegio, Cristian Sothe, hombre que ocupó más tarde cargos importantes en la magistratura prusiana, y de carácter recto y severo, que por su seriedad y buen sentido era llamado desde muy joven Staatarrath (consejero de Estado).

Heine describe el aprecio y estima en que le tiene diciendo:

(f) Fingerprinted document using SSIM Index = 0.50 and PSNR = 6.00dB.

Figure 6.5: Distortion of digital documents due to fingerprint insertion.

Metrics		Type of answer			
Distortion (PSNR)	Similarity (SSIM Index)	#1	#2	#3	#4
30.28dB	0.996	52	41	7	0
28.11dB	0.991	15	69	16	0
26.03dB	0.989	14	74	12	0
24.21dB	0.978	47	45	3	5
22.16dB	0.975	43	55	1	1
20.16dB	0.953	21	64	13	2
18.08dB	0.933	21	58	16	5
16.68dB	0.922	7	79	13	1
14.21dB	0.877	12	73	13	2
12.32dB	0.857	4	55	39	2
10.88dB	0.783	0	36	62	2
8.85dB	0.669	4	7	61	28
6.00dB	0.500	4	6	34	56
4.54dB	0.347	2	6	13	79

Table 6.6: Inquest results of perceptual transparency of fingerprinted digital documents at different levels of distortion (PSNR) and similarity (SSIM Index).

respondents considered that fingerprinted documents with PSNR values greater than 14dB and SSIM Index greater than 0.887 are legibles, otherwise those documents were considered as nonlegibles. Therefore, configurations of β_u , β_g , L and P_w that generate fingerprinted documents with PSNR and SSIM Index equal to or greater than these values satisfy perceptual transparency.

6.2.2 Collusion Resistance

Having identified that a PSNR value of 14dB and a SSIM Index of 0.887 achieve perceptual transparency, it is possible to select configurations of β_u , β_g , L and P_w (see Table 6.5) that generate fingerprinted documents that meet perceptual transparency. Configurations that generate fingerprinted documents with a PSNR value of 16dB were selected. The selected values of parameters of β_u , β_g and L are shown in Table 6.7. Since the best value for P_w is the lower one (1/6), this value is selected as the most appropriated and is fixed for simulations described in the next Section.

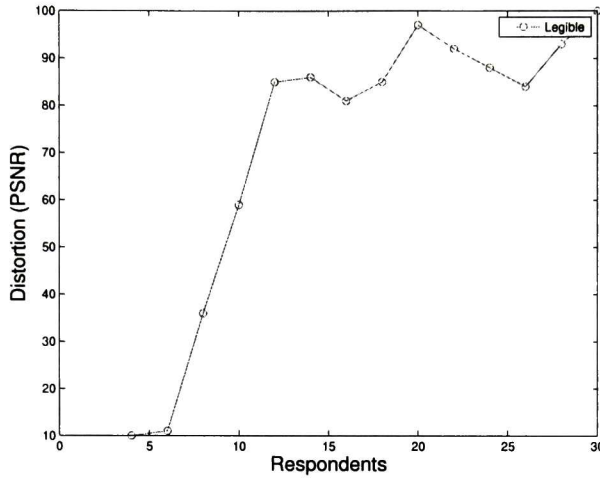


Figure 6.6: Perception of 100 respondents about fingerprinted digital documents with different distortion levels (PSNR).

6.2.2.1 Robustness Factors

In Table 6.7, there are only five configurations of β_u and β_g for each value of L with a PSNR of 16dB. In order to determine the values of β_u and β_g that allow the higher colluder detection ratio, the first five configurations of β_u and β_g were chosen with $L = 50,000$. Then, average collusion attacks were simulated over 50 digital documents, from 2 to 300 colluders that belong to the same group. The fingerprinted documents were generated in the lossless format TIFF with a RGB color scheme.

The results of the collusion attacks simulation are graphically shown in Figures 6.8 and 6.9. Configuration 1 ($\beta_u = 200,000$, $\beta_g = 50,000$, $L = 50,000$) has the higher number of detected colluders, whereas configuration 5 ($\beta_u = 50,000$, $\beta_g = 200,000$, $L = 50,000$) has the lower one. This is significant because the values of β_u and β_g in configuration 1 and 5 are inverted because in the average collusion attack the value of β_u of each colluder is reduced proportionally to the number of colluders. However as β_g is the same for each user, the robustness factor of β_g is first accumulated as many times as the number of colluders, and then it is divided by the same value, having no changes. Therefore,

Configuration	β_u	β_g	L	Configuration	β_u	β_g	L
1	200,000	50,000	50,000	19	100,000	200,000	200,000
2	200,000	100,000	50,000	20	50,000	200,000	200,000
3	150,000	150,000	50,000	21	200,000	50,000	250,000
4	100,000	200,000	50,000	22	200,000	100,000	250,000
5	50,000	200,000	50,000	23	150,000	150,000	250,000
6	200,000	50,000	100,000	24	100,000	200,000	250,000
7	200,000	100,000	100,000	25	50,000	200,000	250,000
8	150,000	150,000	100,000	26	200,000	50,000	300,000
9	100,000	200,000	100,000	27	200,000	100,000	300,000
10	50,000	200,000	100,000	28	150,000	150,000	300,000
11	200,000	50,000	150,000	29	100,000	200,000	300,000
12	200,000	100,000	150,000	30	50,000	200,000	300,000
13	150,000	150,000	150,000	31	200,000	50,000	350,000
14	100,000	200,000	150,000	32	200,000	100,000	350,000
15	50,000	200,000	150,000	33	150,000	150,000	350,000
16	200,000	50,000	200,000	34	100,000	200,000	350,000
17	200,000	100,000	200,000	35	50,000	200,000	350,000
18	150,000	150,000	200,000				

Table 6.7: Configurations of β_u , β_g and L that satisfy perceptual transparency for a fixed P_w .

the robustness factor assigned to users must be the highest possible to resist a big amount of colluders. Figure 6.7 shows the detection sequence from a pirate document generated from the collusion of two colluders with ID=300 and ID=600. The fingerprints of both colluders were defined by configuration 1, and despite the reduction in energy, the threshold T_u could still detect them. Simulations performed for lossless digital documents present a considerable amount of noise under T_u . Much of this noise was due to the pixels in the host image after fingerprint insertion, which were not in the range 0 - 255. Therefore, negative values were set to 0 and values higher than 255 were set to 255. These roundings of were reflected as noise.

6.2.2.2 Fingerprint Length

After defining the values of $\beta_u = 200,000$ and $\beta_g = 50,000$ as the configuration with the higher detection rate, a new simulation of collusion attacks was performed to determine the value of L

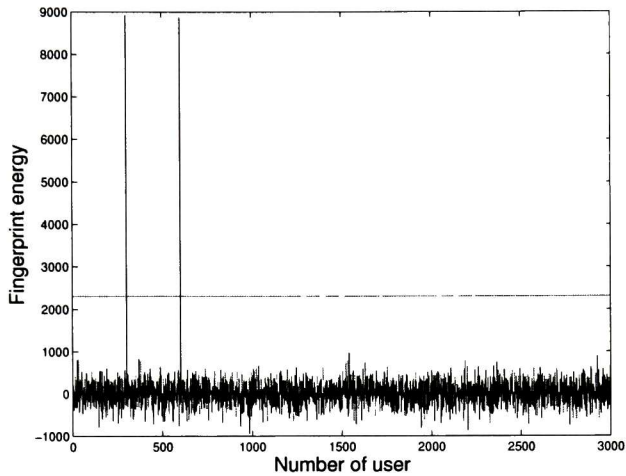


Figure 6.7: Detection of 2 colluded users with IDs 300 and 600, with $B_u = 200,000$, $B_g = 50,000$, $P_w = 1/6$ and $L = 350,000$. The horizontal line defines the threshold value.

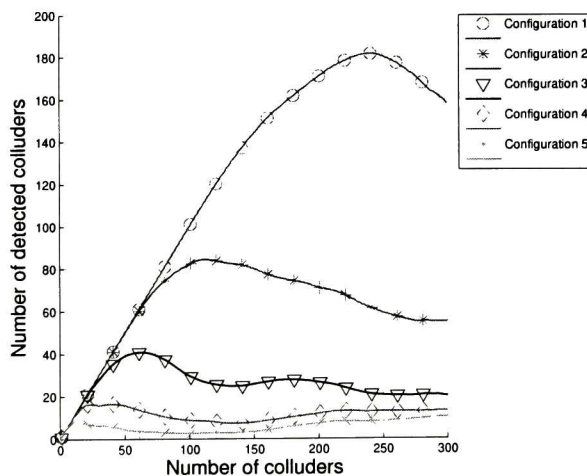


Figure 6.8: Number of detected colluders after a collusion attack for fingerprints generated with configurations 1, 2, 3, 4 and 5 from Table 6.7 and $P_w = 1/6$.

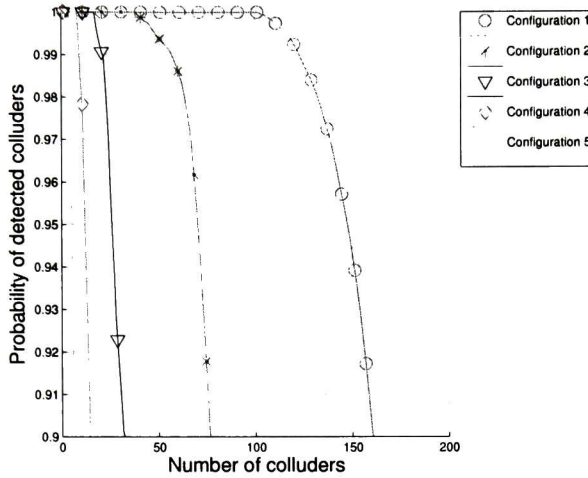


Figure 6.9: Ratio of detected colluders after a collusion attack for fingerprints generated with configurations 1, 2, 3, 4 and 5 from Table 6.7 and $P_w = 1/6$.

that provides the best detection rate. Configurations with different values of L , $\beta_u = 200,000$ and $\beta_g = 50,000$ were selected from Table 6.7, and these configurations were 1, 6, 11, 16, 21, 26 and 31. The simulation results are graphically shown in Figures 6.10 and 6.11. As L is increased, the detection ratio also increased. In [60], Kuribayashi had already mentioned this behavior for the images in general, but in that work it was indicated that the value of L is limited by the image size. It is possible to calculate the highest value of L for the document samples, as shown in Equation 6.1 where L_{max} is the maximum value of L , D_w is the document width, D_h is the document height, D_{com} is the number of color components and P_w is the insertion point of the fingerprint.

$$L_{max} = (D_w * D_h * D_{com}) - P_w \quad (6.1)$$

Using Equation 6.1, for the sample documents L_{max} is 12,150,000. Using $L = 350,000$, it was possible to detect 270 colluders in a collusion attack with 300 colluders. From results shown in Figures 6.10 and 6.11, it was determined that configuration 31 in Table 6.7 allows the maximum amount of colluder detection.

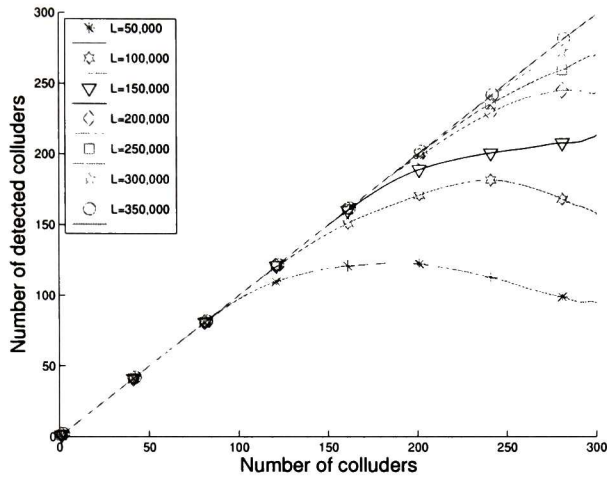


Figure 6.10: Number of detected colluders after a collusion attack for fingerprints generated with $B_g = 50,000$, $B_u = 200,000$ and $P_w = 1/6$.

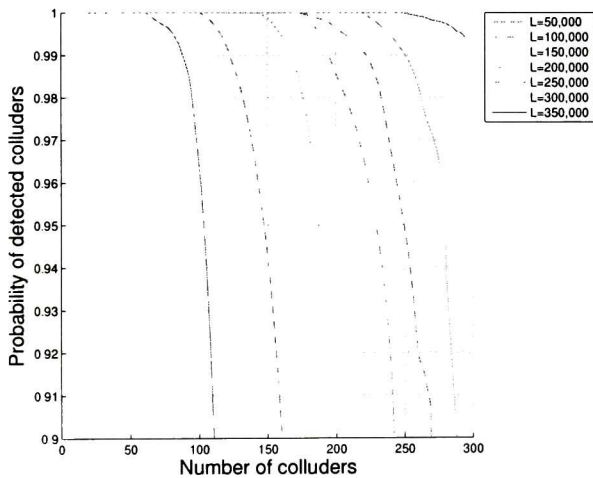


Figure 6.11: Probability of detected colluders after a collusion attack for fingerprints generated with $B_g = 50,000$, $B_u = 200,000$ and $P_w = 1/6$.

6.2.2.3 Attack on Digital Documents in Lossy Format

After determining the configuration of parameters that allow the maximum colluder detection, the detection rate of fingerprints in digital documents with lossy formats was determined. Frequently, image-based digital documents in organizations are stored in lossy formats in order to save space and reduce the bandwidth consumption during their transmission. Hence, determining how the compression affects the fingerprint detection under a collusion attack is important to support the feasibility of implementation. Collusion attacks were simulated using configuration 31 in Table 6.7 to generate fingerprinted documents. The resulting document was stored again in JPEG format. This implies that fingerprints in the digital document were affected twofold by the compression process. Figure 6.12 shows the results of the simulated collusion attacks from 2 to 250 colluders, over 15 digital documents compressed to 80, 60 and 30% quality factor of JPG. Comparing the maximum amount of colluders detected using documents in JPEG format (Figure 6.12) with those using documents in TIFF format (Figure 6.10), it was observed that detection rate in documents in JPEG format is diminished drastically. However, a considerable amount of colluders were still detected in digital documents in the lossy format JPEG, making this an attractive characteristic for organizations.

6.2.3 Performance Evaluation

Having validated that the Fingerprinted Module can achieve perceptual transparency and is resistant to collusion attacks, it was determined if the insertion and detection of fingerprints hold feasible execution times, and therefore, if the Fingerprinting Module could be implemented in real work environments. Performance of insertion and detection of fingerprints were evaluated over 100 digital documents in JPEG and TIFF format, in a desktop computer with Ubuntu 12.10 OS, with an Intel Core i5 processor at 2.7GHz and 4 GB RAM. The insertion evaluation considers the document read time, the fingerprint insertion time and the fingerprinted document write time. Read and write

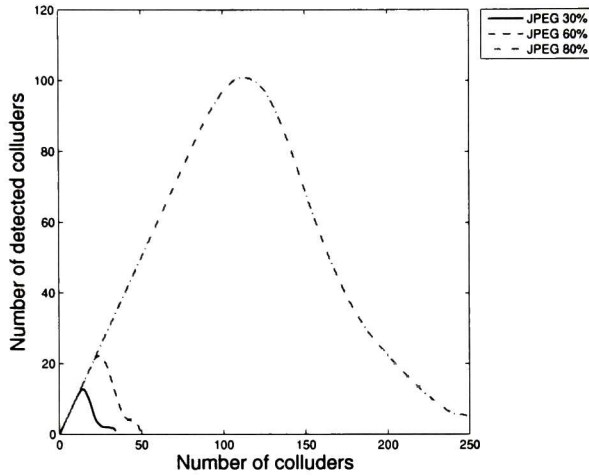


Figure 6.12: Amount of detected colluders, after a collusion attack with lossy digital documents for fingerprints generated with $B_g = 50,000$, $B_u = 200,000$, $L = 350,000$ and $P_w = 1/6$.

times consider the timing to execute the DCT/IDCT transformations over the digital documents. The results obtained are shown in Table 6.8. It is noticed that the time of insertion and detection does not differ significantly between the formats. The most time-consuming task in the insertion and detection of fingerprints is the execution of the DCT/IDCT, with an execution time higher than 3500ms. This can be observed in the relation between the insertion and detection time of fingerprints in Table 6.8. The insertion function requires two transformations and requires twice the time of detection, where only one DCT transformation is needed since fingerprint detection was evaluated with the original digital documents already in the frequency domain. Considerable time is spent during fingerprint insertion due to the DCT transforms applied over digital documents of large dimensions (1900×2700 pixels), however this time is still feasible for real-life applications. The time for performing the detection of users after the digital document has been read and transformed with the IDCT is very low, around 74ms, allowing a wide amount of users per group and having an acceptable detection time.

Format	Insertion time (read, insert, write)	Detection time (read, detect)	Detection time (detect)
JPEG	7574.14ms	3764.19ms	76.53ms
TIFF	7311.45ms	3853.06ms	73.55ms

Table 6.8: Performance evaluation of insertion and detection of fingerprints.

6.2.4 Comparison

In contrast with previous works presented by Brassil et al. in [6] and Schick et al. in [8], the Fingerprinting Module presented in this thesis work is resistant to collusion attacks which is supported by the experimentation reported in previous sections. Although, the fingerprinting scheme presented in the work of Darwish in [7] is resistant to collusion attacks, that work is oriented to XML documents, while in this thesis work the fingerprinting module is oriented to digital documents represented as images. Also, the fingerprinting scheme presented in [7] is vulnerable to simple attacks such as the deletion of white spaces characters that represent the fingerprint, whereas the Fingerprinting Module in this thesis work is robust to these common attacks since it is based on Spread Spectrum. Other works have been proposed based in the watermarking techniques for digital documents described in Chapter 3 (line shifting encoding, word shifting encoding and character space encoding). However, the works derived from these techniques have not been proved for fingerprinting neither their robustness to collusion attacks has been reported. Furthermore these works have not evaluated the impact of watermark insertion over perceptual transparency. Table 6.9 shows a comparison of the Fingerprinting Module developed in this thesis work against the most representative works reviewed in Chapter 3. The Fingerprinting module developed in this thesis was validated regarding perceptual transparency and robustness to collusion attacks through experimentation.

Work	Domain	Strategy	Robustness		Perceptual transparency	Number of users
			Collusion-resistant	Lossy compression		
Low <i>et al.</i> [68]	space	Line shifting encoding	not reported	yes	not reported	not reported
Low <i>et al.</i> [70]	space	Line shifting encoding	not reported	yes	not reported	not reported
Alattar [73]	space	Line shifting encoding	not reported	yes	not reported	not reported
Kim <i>et al.</i> [74]	space	Word shifting encoding	not reported	yes	not reported	not reported
Yawai and Hiransakolwong [77]	space	Word shifting encoding	not reported	yes	not reported	not reported
Huang and Yan [71]	space	Character space encoding	not reported	yes	not reported	not reported
Chotikakamthorn [72]	space	Character space encoding	not reported	yes	not reported	not reported
Proposed	frequency	Spread spectrum	yes	yes	guarantee	L^2

Table 6.9: Comparison of the Fingerprinting Module in this thesis against other insertion techniques in the literature applicable to fingerprinting.

6.3 Integrated System

The proper functioning of the SDMS was validated considering the integration of the ISS and the Fingerprinting Modules. Validation consisted in testing the functions of the SDMS in which the ISS and the Fingerprinting Modules interact and measuring their timing performance. These validations are described in the next Sections.

6.3.1 Operations Validation

The execution paths in the SDMS were validated performing unit tests ensuring that the result of each cryptographic function along with the fingerprinting functions worked properly. The unit testing results are shown in Table 6.10.

User actions	Result
Signature validation of digital documents in Department A by an Auditor user	Successful
Decryption of digital documents in Department A by a Consumer user of department A	Successful
Encryption of digital documents in Department A by a Reviewer user of department A	Successful
Fingerprint insertion on a digital document during Distribution stage	Successful
Fingerprint detection by an Administrator user	Successful
Fingerprint detection under a collusion attack by an Administrator user	Successful
Mapping of user digital certificate to user ID	Successful
Access to digital documents by a user without a digital certificate (no authenticated)	Denied
Access to digital documents in Department A by a Consumer user of department B	Denied
Encryption of digital documents in Department A by a Reviewer users of department B	Denied
Encryption of digital documents in Department A by a user of department A without privileges	Denied
Access to the User Mapping repository by a non-authorized user	Denied
Fingerprint detection by a non-authorized user	Denied
Signature validation of digital documents in Department A by a user of department A without privileges	Denied
Signature validation of corrupted digital documents	Failed
Signature validation of digital documents with corrupted signatures	Failed
Fingerprint detection in a non-fingerprinted document	Failed

Table 6.10: Test performed to validate the correct functioning of the system integrating the ISS and fingerprinting modules.

Unit test were performed obtaining the expected results. Hence, it was determined that the ISS and Fingerprinting Modules work properly considering the actions that the SDMS must allow users to perform. Having determined that the SDMS works properly, a performance evaluation was carried out to determine if the SDMS is suitable for a production environment.

6.3.2 Performance Evaluation

The ISS and Fingerprinting Modules interact each other when a digital document is downloaded and when fingerprints are detected from a pirate document. Hence, timing performance is re-evaluated for the fingerprint detection and the downloading of digital documents integrating the fingerprint insertion function. Tests were carried out over 30 digital documents in JPEG, in a desktop computer with Ubuntu 12.10, with a processor Intel Core i5 CPU at 2.7GHz and 4 GB RAM, with the SDMS and client in the same computer to avoid networking delays. Performance of other allowed user actions listed in Table 6.10 that do not require interaction between modules were already evaluated in Sections 6.1.2 and 6.2.3 of this Chapter.

6.3.2.1 Download Digital Documents

During the download of digital documents, digital documents are decrypted, then fingerprinted and finally encrypted and sent to a user using the SSL protocol. The average execution times of these functions are shown in Table 6.13.

The average time that takes a user to obtain a fingerprinted document is 8652.1ms. It can be observed in Table 6.13 that the fingerprint insertion is the most time consuming function requiring around 7500ms to be performed, whereas the task of the ISS Module is around 1100ms. The time it takes to download a digital document is acceptable since not real time responses are required generally in a SDMS. In the case that better response times are required, the fingerprint insertion time can be improved. The fingerprint insertion requires to perform a DCT function and a IDCT function taking around 3500ms each one. However, performing the DCT function on digital documents when

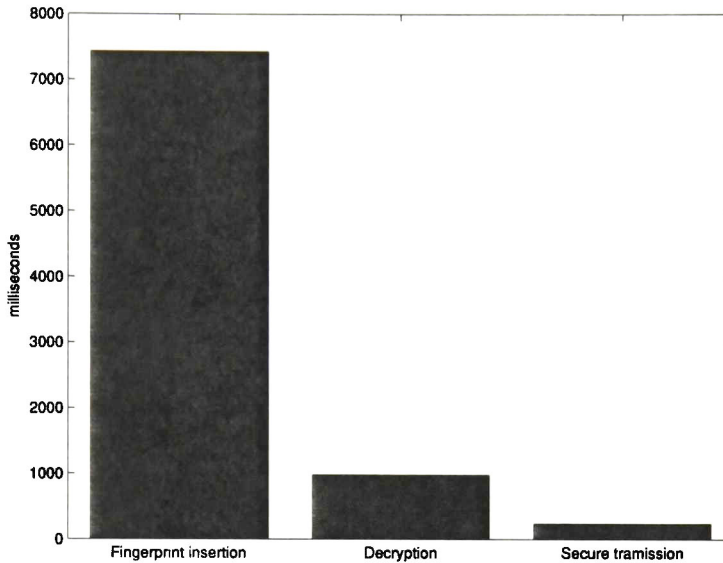


Figure 6.13: Average time of fingerprint insertion, decryption and secure trasmission, required to download a digital document.

are Approved, implies only to perform the IDCT function during their download, reducing insertion time by a half. Although, this strategy requires to store the digital documents with no compression, increasing the space required to store them.

6.3.2.2 Fingerprint Detection

During fingerprint detection in a pirate document copy, the ISS Module obtains the encrypted original document from the digital document repository. Then, the encrypted original document is decrypted and stored temporary. After that, the original digital document in clear text is read by the Fingerprinting Module and it is transformed to the frequency domain. These operations require an average time of 4034.54ms. This time is slightly higher than the time reported only for the Fingerprinting Module because the required interaction with the ISS Module. Once the original and pirate documents are in the frequency domain, average detection time is about 78.03ms. This

time does not differ significantly from the detection time determined only by the Fingerprinting Module since at this point, interaction with the ISS Module is not more required. Detection times of fingerprints with the integrated modules do not differ significantly from the times presented during the Fingerprinting Module performance evaluation.

The SDMS, integrating the ISS and Fingerprinting modules, exhibited an acceptable performance, proving to protect digital documents from unauthorized access during the Approval, Storage and Distribution stages of the document lifecycle. Also, it was possible to determine the identity of traitor users from fingerprinted documents. Since performance evaluations were done in a desktop equipment, it is expected that performance of the proposed SDMS improves if a specialized server is used.

6.4 Summary

This Chapter presented results of the the ISS Module, Fingerprinting Module and the integration of the two modules in a SDMS.

The ISS Module was validated by performing successfully all the allowed execution paths and denying the invalid paths during the Approval, Storage and Distribution stages of the document lifecycle in the SDMS. The performance of this module was evaluated obtaining execution times that are feasible for a production environment. Finally, this module was compared with the works reviewed in the state-of-the-art, showing that this module meets the required information security services whereas other works do not do it because their implementation scenario is different that the one addressed in this thesis.

During the validation of the Fingerprinting Module permissible levels of distortion in terms of legibility of fingerprinted documents were defined. It was found that, in contrast with fingerprinting applied in natural images, lower values of PSNR and SSIM Index are tolerated by fingerprinted

digital documents without affecting drastically the perceptual transparency, which was achieved with a minimum PSNR value of 14dB and a minimum SSIM Index value of 0.877. Then, appropriated values of parameters were determined to achieve the higher amount of colluders detection. It was shown that the robustness factor assigned to users must be higher than the robustness factor assigned to groups in order to have a higher colluder detection probability. The Fingerprinting Module is able to fully detect until 270 of 300 colluders for lossless compressed digital documents. Furthermore, the detector performance after lossy compression remains competitive for real work environments. Finally, the Fingerprinting Module was compared with the works presented in the state-of-the-art, and works of watermarking for digital documents that could be used in a fingerprinting context. From that comparison, it was concluded that this is the first implementation of a fingerprinting scheme for image-based digital documents that is robust to collusion attacks. The number of users available, and the high quality lossy compression robustness make the Fingerprinting Module proposed in this thesis suitable for implementation in a production environment.

In order to state that the integration of modules and the resulting SDMS is able to be implemented, all the execution paths of the SDMS, integrating the ISS and Fingerprinting Modules, were validated by performing successfully all the allowed execution paths and denying the invalid paths. Finally, the performance of downloading digital documents and detecting fingerprints in fingerprinted documents were evaluated. These are the functions that require interactions between the ISS and Fingerprinting Module. It was found that performance on these functions integrated of the SDMS are feasible for a production environment.

The development of the SDMS integrating the ISS and Fingerprinting Modules responds the investigation question tackled in this thesis, achieving all its research objectives. To the best of author's knowledge, this is the first work that integrates information security services and user tracing techniques to create a SDMS.

7

Conclusions and Future Work

In this thesis, a Secure Document Management System (SDMS) that allows tracing users was developed. To achieve this, two modules were developed: a Information Security Services Module and a Fingerprinting Module.

The Information Security Services Module was designed and implemented to provide information security services to digital documents during the stages of Approval, Storage and Distribution of the digital document lifecycle in a Document Management System. The selected techniques to provide the required information security services were: i) digital signatures for integrity, authentication and non-repudiation; ii) symmetric encryption and digital envelopes for confidentiality; iii) access control was implemented using Role-Based Access Control and Mandatory Access Control. These techniques required a Public Key Infrastructure which was implemented using cryptographic techniques to provide protection for 30 years following the NIST standards and recommendations.

RSA was used as the asymmetric cryptographic algorithm with a key length of 2048 bits, AES was used as the symmetric cryptographic algorithm with a key length 256 bits and SHA-2 with a length of

256 bits is used as hash function. The execution of all the paths of the Secure Document Management System proved that the Information Security Services Module protects the digital documents with feasible performance times in the entirely stages of Approval, Storage and Distribution.

For the Fingerprinting Module, it was selected a state-of-the-art fingerprinting technique resistant to collusion attacks based on Spread Spectrum, to protect digital documents represented as images from illegal distribution. In particular, the system proposed by Minoru Kuribayashi was studied. Larger fingerprint lengths than in natural images can be used for digital document's fingerprinting without affecting drastically the perceptual transparency of fingerprints. In contrast with fingerprinting in natural images that are considered perceptually transparent with a PSNR near to 30dB, fingerprinted digital documents are perceptually transparent with a minimum PSNR value of 14dB and a minimum SSIM Index value of 0.877, allowing to increase the robustness of fingerprints maintaining their perceptual transparency. The minimum PSNR and SSIM Index values to maintain perceptual transparency of fingerprints in digital documents were validated through an inquest to 100 respondents. The energy associated to fingerprints was the most relevant value found, as it generates distortion over all the digital document. Parameters configurations composed by robustness factor of user, robustness factor of group, fingerprint length and, fingerprint insertion position were determined to achieve perceptual transparency for fingerprints. From these configurations, the configuration that detects the higher number of colluders under a collusion attack was determined. It was shown that energy assigned to users must be higher than energy assigned to groups to have the higher colluder detection probability. The best configuration of parameters for the selected fingerprinting scheme has proved to be resistant to collusion attacks in lossless compressed digital documents, allowing the full detection of 270 users from a collusion attack with 300 colluders. Furthermore, the detector performance after lossy compression stills detect a considerable amount of colluders. Performance of this module was suitable for real work environments.

Finally, the integration of the Information Security Services Module and the Fingerprinting Module was achieved by developing a component that maps the user identity between Modules. Also, it was

evidenced the need to protect the insertion and detection of fingerprints with information security services. The developed Secure Document Management System proposed in this thesis allows the protection of digital documents during their Approval, Storage and Distribution, and also allows the tracing of traitor users. Furthermore, the system holds suitable times for being implemented in a production environment.

7.1 Main Contributions

The main contribution in this thesis is the development of the first Secure Document Management System that is able to trace traitor users that distribute pirate documents. Specially, the contributions of this work are:

- The development of an Information Security Services Module to protect digital documents during their document lifecycle, tackling an scenario for which previous works had not considered.
- The development of a Fingerprinting Module that generates fingerprinted documents to trace traitor users, being this the first collusion resistant fingerprinting work for image-based digital documents.
- The evaluation of a state-of-the-art fingerprinting scheme in the context of restricted distribution of digital documents.
- The establishment of permissible values of PSNR and SSIM Index for fingerprinted digital documents, being this to author's knowledge the first time these values are reported.
- It was proved that fingerprinting techniques and information security services can be integrated to create a system that in a comprehensive manner solves the secure storage and distribution of digital documents and allows to trace traitor users.

7.2 Future Work

Future work can be done to extend the robustness of the fingerprints generated by the Fingerprinting Module to resist other attacks than collusion and lossy compression. For the developed system, timing performance when downloading digital documents can be improved transforming the document to their frequential representation since they are uploaded. Also, new ways of integration between information security services and fingerprinting techniques can be explored, such as the embedding of fingerprinting techniques with cryptographic techniques by designing an asymmetric cryptographic algorithm that requires a user ID at the moment of digital documents decryption to generate a fingerprint that is inserted in the decrypted digital document. Another way of integration of information security services and fingerprinting techniques are related to the integrity validation of a fingerprinted pirate copy, comparing it with the non-fingerprinted original digital document in order to determine if besides the pirate copy was distributed illegally, it has been modified by malicious users. Validation of integrity of fingerprinted digital documents can be used to extend the protection provided to digital documents in the Usage stage of the document lifecycle.

Bibliography

- [1] T. Casey, M. Roe, B. Tuck, and S. Wilbur, "Secure automated document delivery," in *Fifth Annual Computer Security Applications Conference, 1989*, (University College London), pp. 348–356, 1990.
- [2] A. Lioy, F. Maino, and M. Mezzalama, "Secure document management and distribution in an open network environment," in *ICICS'97 International Conference on Information and Communications*, (Beijing, China), pp. 109–117, 1997.
- [3] J. Shi and J. Ouyang, "eSign: an enterprise portal for secure document management," in *2005 IEEE International Conference on Information Reuse and Integration 2005*, pp. 481–486, 2005.
- [4] S. Kamara and K. Lauter, "Cryptographic cloud storage," in *Proceedings of the 14th international conference on Financial cryptography and data security, FC'10*, (Berlin, Heidelberg), pp. 136–149, 2010.
- [5] T. Chieu, T. Nguyen, and L. Zeng, "Secure search of private documents in an enterprise content management system," in *IEEE International Conference on e-Business Engineering, 2007*, (Hong Kong, China), pp. 105–112, 2007.
- [6] J. Brassil, S. Low, and N. Maxemchuk, "Copyright protection for the electronic distribution of text documents," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1181–1196, 1999.
- [7] S. Darwish, "New system to fingerprint extensible markup language documents using winnowing theory," *Signal Processing, IET*, vol. 6, no. 4, pp. 348–357, 2012.
- [8] R. Schick and C. Ruland, "Document tracking - on the way to a new security service," in *2011*

- Conference on Network and Information Systems Security (SAR-SSI)*, (La Rochelle, France), pp. 1–5, 2011.
- [9] D. Ruiu, "Learning from information security history," *IEEE Security Privacy*, vol. 4, no. 1, pp. 77 – 79, 2006.
- [10] K. Goda and M. Kitsuregawa, "The history of storage systems," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1433 –1440, 2012.
- [11] A. Q. Shazia Akram, Mehraj-Ud-Din Dar, "Document image processing a review," *International Journal of Computer Applications*, vol. 10, no. 5, pp. 35–40, 2010.
- [12] Adobe, "Making the case PDF/A and Adobe Acrobat," white paper, Adobe Systems Incorporated. <http://goo.gl/F4ZkL> - Last access 28/09/2013, 2010.
- [13] R. Gupta, A. Karayil, and R. Rajendran, "Contract lifecycle management," white paper, Infosys. <http://goo.gl/aEQj9> - Last access 28/09/2013, 2008.
- [14] S. Feldman and R. Villars, "The information lifecycle management imperative," white paper, IDC Information and data. <http://goo.gl/4aOqo> - Last access 28/09/2013, 2006.
- [15] K. Kimberland, "2010 Cybersecurity Watch Survey," survey, Computer Emergency Response Team (CERT), 2010.
- [16] Adobe, "A primer on electronic document security," white paper, Adobe Systems Incorporated, 2007.
- [17] M. Munier, L. Munier, Vincent, and R. Magali, "Self-protecting documents for cloud storage security," in *IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, (Liverpool), pp. 1231 –1238, 2012.

- [18] H. Berghel, "Wikileaks and the matter of private manning," *IEEE Computer*, vol. 45, no. 3, pp. 70 – 73, 2012.
- [19] L. Coles-Kemp and M. Theoharidou, "Insider threat and information security management," in *Insider Threats in Cyber Security* (C. W. Probst, J. Hunker, D. Gollmann, and M. Bishop, eds.), vol. 49 of *Advances in Information Security*, pp. 45–71, Springer US, 2010.
- [20] G. Kessler, "Information security: New threats or familiar problems?," *Computer*, vol. 45, no. 2, pp. 59 –65, 2012.
- [21] B. M. Gaff, R. A. Loren, and E. A. Spinney, "Intellectual Property, Part II," *Computer*, vol. 45, no. 2, pp. 9 –11, 2012.
- [22] W. Stallings, *Cryptography and Network Security: Principles and Practice*. Upper Saddle River, NJ, USA: Prentice Hall Press, 5th ed., 2010.
- [23] E. Schaefer, "An introduction to cryptography and cryptanalysis," tech. rep., Santa Clara University, 1999.
- [24] F. P. Ramos, I. L. Perez, J. P. G. Moran, and A. A. R. Varon, *Hacking y Seguridad en Internet*. AlfaOmega, 2007.
- [25] M. Agrawal and P. Mishra, "A comparative survey on symmetric key encryption techniques," *International Journal on Computer Science and Engineering (IJCSE)*, vol. 4, pp. 877–882, 2012.
- [26] W. Diffie and M. E. Hellman, "New Directions in Cryptography," *IEEE Transactions on Information Theory*, vol. IT-22, no. 6, pp. 644–654, 1976.
- [27] S.-C. Cheung and D. Chiu, "A watermarking infrastructure for enterprise document management," in *Proceedings of the 36th Annual Hawaii International Conference on System*

- Sciences (HICSS'03)*, vol. 4 of *HICSS '03*, (Washington, DC, USA), pp. 105.2–, IEEE Computer Society, 2003.
- [28] N. Koblitz, "Elliptic Curve Cryptosystems," *Mathematics of Computation*, vol. 48, no. 177, pp. 203–209, 1987.
- [29] V. S. Miller, "Use of elliptic curves in cryptography," in *Advances in cryptology Lecture notes in computer sciences 218*, (New York, NY, USA), pp. 417–426, Springer-Verlag New York, Inc., 1986.
- [30] D. Brown, "Sec 2: Recommended elliptic curve domain parameters," tech. rep., Certicom Corp., 2010.
- [31] N. Koblitz and A. J. Menezes, "A survey of public-key cryptosystems," *SIAM*, vol. 46, no. 4, pp. 599–634, 2004.
- [32] K. Schmeh, "A taxonomy of cryptographic techniques for securing electronic identity documents," in *Securing Electronic Business Processes* (N. Pohlmann, H. Reimer, and W. Schneider, eds.), pp. 349–356, Vieweg Teubner, 2010.
- [33] G. Zhao, X. Yang, B. Zhou, and W. Wei, "RSA-based digital image encryption algorithm in wireless sensor networks," in *2nd International Conference on Signal Processing Systems (ICSPS)*, vol. 2, (Dalian, China), pp. 640–643, 2010.
- [34] K. Gupta, S. Silakari, R. Gupta, and S. Khan, "An ethical way of image encryption using ecc," in *First International Conference on Computational Intelligence, Communication Systems and Networks.*, pp. 342–345, 2009.
- [35] T. A. Z. S. Lala Krikor, Sami Baba, "Image Encryption Using DCT and Stream Cipher," *European Journal of Scientific Research*, vol. 32, no. 1, pp. 48–58, 2009.
- [36] B. Schneier, *Applied Cryptography*. Wiley and Sons, Incorporated, second ed., 1995.

- [37] S. Preiya, R. Pavithra, and J. Jakkulin, "Secure role based data access control in cloud computing," *International Journal of Computer Trends and Technology*, pp. 146–151, 2011.
- [38] W. Al-Hamdani, "Cryptography based access control in healthcare web systems," in *Information Security Curriculum Development Conference, InfoSecCD 10*, pp. 66–79, ACM, 2010.
- [39] R. S. Chad Dougherty, Kirk Sayre and D. Svoboda, "Secure design patterns," tech. rep., Software Engineer Institute, 2009.
- [40] N. Liu, "Cloud technology in the security management of enterprise document," in *Second International Conference on Innovations in Bio-inspired Computing and Applications (IBICA)*, pp. 267–269, 2011.
- [41] A. A. Syed, "Digital watermarking," Report 1000614216, The University of Texas at Arlington, 2009.
- [42] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2 ed., 2008.
- [43] V. Licks and R. Jordan, "Geometric attacks on image watermarking systems," *MultiMedia, IEEE*, vol. 12, no. 3, pp. 68–78, 2005.
- [44] S. Voloshynovskiy, S. Pereira, T. Pun, J. Eggers, and J. Su, "Attacks on digital watermarks: classification, estimation based attacks, and benchmarks," *IEEE Communications Magazine*, vol. 39, no. 8, pp. 118–126, 2001.
- [45] M. Wu, W. Trappe, J. Wang, and K. Liu, "Collusion-resistant fingerprinting for multimedia," *IEEE Signal Processing Magazine*, vol. 21, no. 2, pp. 15–27, 2004.
- [46] A. Bamatraf, R. Ibrahim, and M. N. M. Salleh, "A new digital watermarking algorithm using combination of least significant bit (LSB) and inverse bit," *CoRR*, vol. 3, no. 4, 2011.

- [47] J. Fridrich, M. Goljan, and R. Du, "Detecting lsb steganography in color, and gray-scale images," *IEEE MultiMedia*, vol. 8, no. 4, pp. 22–28, 2001.
- [48] N. M. Walter Bender, Daniel Gruhl and Y. Lu, "Techniques for data hiding," *IBM Systems Journal*, vol. 35, pp. 313–336, 1996.
- [49] I.-K. Yeo and H. J. Kim, "Generalized patchwork algorithm for image watermarking," *Journal: Multimedia Systems - MMS*, vol. 9, pp. 261–265, 2003.
- [50] M. Arnold, "Audio watermarking: Features, applications, and algorithms," in *IEEE International Conference on Multimedia and Expo (III)*, pp. 1013–1016, 2000.
- [51] H.-J. Kim and I.-K. Yeo, "Modified patchwork algorithm: A novel audio watermarking scheme," tech. rep., Kangwon National University, Department of Control and Instrumentation Engineering, 2008.
- [52] B. Chen and G. W. Wornell, "Quantization index modulation methods for digital watermarking and information embedding of multimedia," *VLSI Signal Processing*, vol. 27, no. 1/2, pp. 7–33, 2001.
- [53] J. Prins, Z. Erkin, and R. Lagendijk, "Anonymous fingerprinting with robust QIM watermarking techniques," *EURASIP Journal on Information Security*, vol. 2007, pp. 1–7, 2007.
- [54] I. S. Jinshen Wang and R. Lagendijk, "Scale estimation in two-band filter attacks on qim watermarks," in *Security, Steganography, and Watermarking of Multimedia Contents VIII*, 2006.
- [55] J. D. Nayerlaan, "Spread Spectrum (SS)," techreport, Hogeschool Voor Wetenschap & Kunst, 1999.
- [56] R. Chandramouli, N. Memon, and M. Rabbani, "Digital watermarking," in *Encyclopedia of Imaging Science and Technology*, 2002.

- [57] I. Cox, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, pp. 1673–1687, 1997.
- [58] T. Todorov, "Spread spectrum watermarking technique for information system securing," *International Journal "Information Theories & Applications"*, vol. 11, pp. 405–408, 1996.
- [59] W. Jane, M. Wu, W. Trappe, and L. Ray, "Group-oriented fingerprinting for multimedia forensics," *EURASIP Applied Signal Processing*, vol. 2004, pp. 2153–2173, 2004.
- [60] M. Kuribayashi, "Hierarchical spread spectrum fingerprinting scheme based on CDMA technique," *EURASIP Journal on Information Security*, vol. 2011, p. 16, 2011.
- [61] Z. Jalil and A. Mirza, "A review of digital watermarking techniques for text documents," in *International Conference on Information and Multimedia Technology. ICIMT '09.*, (Jeju Island, South Korea), pp. 230–234, 2009.
- [62] X. Zhou, Z. Wang, W. Zhao, S. Wang, and J. Yu, "Performance analysis and evaluation of text watermarking," in *International Symposium on Computer Network and Multimedia Technology.*, (Wuhan, China), pp. 1–4, 2009.
- [63] H. M. Meral, B. Sankur, A. S. Özsoy, T. Güngör, and E. Sevinç, "Natural language watermarking via morphosyntactic alterations," *Computer Speech & Language*, vol. 23, no. 1, pp. 107 – 125, 2009.
- [64] M. Atallah, C. McDonough, V. Raskin, and S. Nirenburg, "Natural language processing for information assurance and security: an overview and implementations," in *Proceedings of the 2000 workshop on New security paradigms, NSPW '00*, pp. 51–65, ACM, 2000.
- [65] U. Topkara, M. Topkara, and M. Atallah, "The hiding virtues of ambiguity: quantifiably resilient watermarking of natural language text through synonym substitutions," in *Proceedings of the*

- 8th workshop on Multimedia and security, MM#38;Sec '06, (New York, NY, USA), pp. 164–174, ACM, 2006.*
- [66] X. Sun and A. J. Asiimwe, "Noun-verb based technique of text watermarking using recursive decent semantic net parsers," in *Proceedings of the First international conference on Advances in Natural Computation Volume Part III, ICNC'05, (Berlin, Heidelberg), pp. 968–971, Springer-Verlag, 2005.*
- [67] J. Brassil, S. Low, L. O'Gorman, and N. Maxemchuk, "Electronic marking and identification techniques to discourage document copying," in *IEEE Journal on Selected Areas in Communications*, pp. 1278–1287, 1995.
- [68] S. Low and N. Maxemchuk, "Document identification for copyright protection using centroid detection," *IEEE Transactions on Communications*, vol. 46, no. 3, pp. 372–383, 1998.
- [69] J. Brassil, S. Low, N. Maxemchuk, and L. O'Gorman, "Hiding information in document images," in *Conference on Information Sciences and Systems (CISS-95)*, pp. 482–489, 1994.
- [70] S. Low, N. Maxemchuk, J. Brassil, and L. O'Gorman, "Document marking and identification using both line and word shifting," in *INFOCOM '95, Proceedings of the Fourteenth Annual Joint Conference of the IEEE Computer and Communication Societies*, vol. 2, (Boston, MA), pp. 853–860, 1995.
- [71] D. Huang and H. Yan, "Interword distance changes represented by sine waves for watermarking text images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 12, pp. 1237–1245, 2001.
- [72] N. Chotikakamthorn, "Electronic document data hiding technique using inter-character space," in *The 1998 IEEE Asia-Pacific Conference on Circuits and Systems. IEEE APCCAS 1998., (Chiang Mai, Thailand), pp. 419–422, 1998.*

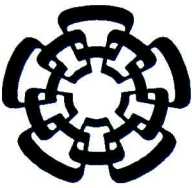
- [73] A. Alattar and O. Alattar, "Watermarking electronic text documents containing justified paragraphs and irregular line spacing," tech. rep., Digimarc Corporation, 2004.
- [74] Y. won Kim, K. ae Moon, and I. seok Oh, "A text watermarking algorithm based on word classification and inter-word space statistics," in *Proceedings Seventh International Conference on Document Analysis and Recognition 2003*, pp. 775–779, 2003.
- [75] S. Low and N. Maxemchuk, "Capacity of text marking channel," *IEEE Signal Processing Letters*, vol. 7, pp. 345–347, 2000.
- [76] H.-G. Choo and W.-Y. Kim, "Data-hiding capacity improvement for text watermarking using space coding method," in *Digital Watermarking* (T. Kalker, I. Cox, and Y. Ro, eds.), vol. 2939 of *Lecture Notes in Computer Science*, pp. 593–599, Springer Berlin Heidelberg, 2004.
- [77] W. Yawai and N. Hiransakolwong, "Increase the hiding-bit capacity and strength for text watermarking with the line intersection on text image," in *8th International Conference on Computing Technology and Information Management (ICCM)*, vol. 1, (Seoul, South Korea), pp. 427–433, 2012.
- [78] Y.-L. Tang and Y.-T. Huang, "Print-and-scan resilient watermarking for authenticating paper-based certificates," in *2010 First International Conference on Pervasive Computing Signal Processing and Applications (PCSPA)*, pp. 357–361, 2010.
- [79] F. Lefebvre, A. Gueluy, D. Delannay, and B. Macq, "A print and scan optimized watermarking scheme," in *2001 IEEE Fourth Workshop on Multimedia Signal Processing*, pp. 511–516, 2001.
- [80] B. Chor, A. Fiat, and M. Naor, "Tracing traitors," in *Advances in Cryptology - CRYPTO '94* (Y. Desmedt, ed.), vol. 839 of *Lecture Notes in Computer Science*, pp. 257–270, Springer Berlin Heidelberg, 1994.

- [81] Safavi-Naini and Y. Wang, "Sequential traitor tracing," *IEEE Transactions on Information Theory*, vol. 49, pp. 1319 – 1326, may 2003.
- [82] Mediahedge, "Digital fingerprinting," white paper, Mediahedge, Civolution, Gracernote, 2010.
- [83] J. Ku and B. Girod, "On the robustness and imperceptibility of digital fingerprints," in *IEEE International Conference on Multimedia Computing and Systems.*, vol. 2, pp. 530 –535, 1999.
- [84] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *20th International Conference on Pattern Recognition (ICPR)*, (Washington, DC, USA), pp. 2366–2369, 2010.
- [85] B. S. Ismail Avcibas, "Statistical analysis of image quality measures," tech. rep., Department of Electrical and Electronic Engineering, Bogazici University, Istanbul, Turkey, 1999.
- [86] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [87] G. Strang, "The discrete cosine transform," *SIAM Rev.*, vol. 41, no. 1, pp. 135–147, 1999.
- [88] J. J. Garcia-Hernandez, C. Feregrino-Urbe, and R. Cumplido, "Collusion-resistant audio fingerprinting system in the modulated complex lapped transform domain," *PLoS ONE*, vol. 8, p. e65985, 06 2013.
- [89] Z. Wang and B. Hunt, "The discrete w transform," *Applied Mathematics and Computation*, vol. 16, 1985.
- [90] N. Ahmed, T. Natarajan, and R. Rao, "Discrete cosine transform," *IEEE Transactions on Computers*, vol. C-23, no. 1, pp. 90–93, 1974.
- [91] H. Kekre and J. Solanki, "comparative performance of various trigonometric unitary transforms for transform image," *International Journal of Electronics*, vol. 44, no. 3, 1978.

- [92] M. Frigo and S. Johnson, "The design and implementation of FFTW3," *Proceedings of the IEEE*, vol. 93, no. 2, pp. 216–231, 2005.
- [93] S. Goyal, "A survey on the applications of cryptography," *International Journal of Science and Technology*, vol. 1, no. 7, pp. 137–140, 2012.
- [94] D. Gerasimov, "Creation of the system of secure document circulation for state organization," in *Proceedings of the 6th Annual 2005 International Siberian Workshop and Tutorials on Electron Devices and Materials*, (Eragol, Altai, Russia), pp. 248–249, 2005.
- [95] T. Kwok and T. Nguyen, "An enterprise electronic contract management system using dual XML and secure PDF documents," in *IEEE 10th International Enterprise Distributed Object Computing Conference Workshops, 2006. EDOCW '06.*, (Hong Kong, China), p. 57, 2006.
- [96] G. Zhao, X. Hu, Y. Li, and L. Du, "Scheme for digital documents management in networked environment," in *IEEE International Conference on Network Infrastructure and Digital Content, 2009. IC-NIDC 2009.*, (Beijing, China), pp. 995–998, 2009.
- [97] K. Odagiri, N. Ishii, R. Yaegashi, and M. Tadauchi, "A distribution system of document medium with copyright protection," in *10th ACIS International Conference on Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing, 2009. SNPD.*, (Daegu, South Korea), pp. 207 – 212, 2009.
- [98] Y. Bai, "Access control for XML document," in *New Frontiers in Applied Artificial Intelligence* (N. Nguyen, L. Borzemeski, A. Grzech, and M. Ali, eds.), vol. 5027 of *Lecture Notes in Computer Science*, pp. 621–630, Heidelberg, Berlin: Springer Berlin Heidelberg, 2008.
- [99] P. Sevinc, D. Basin, and E.-R. Olderog, "Controlling access to documents: A formal access control model," tech. rep., University of Oldenburg, 2006.

- [100] S. Schleimer, D. S. Wilkerson, and A. Aiken, "Winnowing: local algorithms for document fingerprinting," in *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, SIGMOD '03, (New York, NY, USA), pp. 76–85, ACM, 2003.
- [101] Brainloop, "End-to-end security for confidential digital documents," White Paper 1, Brainloop Inc., 2009.
- [102] E. B. Barker and A. L. Roginsky, "Transitions: Recommendation for Transitioning the Use of Cryptographic Algorithms and Key Lengths," tech. rep., Gaithersburg, MD, United States, 2013.
- [103] J. de Lavarene, "SSL with oracle JDBC thin driver," Whitepaper 24, Oracle Corporation, 2010.
- [104] Y. Wang, "Public-key cryptography standards: PKCS," report, University of North Carolina at Charlotte, 2004.
- [105] X. Zhou, W. Zhao, Z. Wang, and L. Pan, "Security theory and attack analysis for text watermarking," in *International Conference on E-Business and Information System Security. EBISS '09.*, (Wuhan, China), pp. 1–6, 2009.
- [106] M. W. Frank Warmerdam, Andrey Kiselev and D. Kelly., "TIFF library and utilities. Available: <http://www.libtiff.org/>. last access 28/09/2013," 2013.
- [107] I. J. Group, "Libjpeg. Available: <http://www.ijg.org/>. last access 28/09/2013," 2013.
- [108] T. Distler, "Image quality assessment: IQA. Available: <http://tdistler.com/iqa/>. last access 28/09/2013," 2013.
- [109] A. Shaamala, "Study of the effect DCT and DWT domains on the imperceptibility and robustness of genetic watermarking," *IJCSI International Journal of Computer Science Issues*, vol. 8, pp. 220–225, 2012.

-
- [110] E. Nezhadarya, J. Wang, and R. Ward, "Image quality monitoring using spread spectrum watermarking," in *IEEE 16th International Conference on Image Processing (ICIP)*, (Cairo, Egypt), pp. 2233–2236, 2009.



CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL IPN

UNIDAD TAMAULIPAS

Cd. Victoria, Tamaulipas, a 10 de diciembre de 2013.


Los abajo firmantes, integrantes del jurado para el examen de grado que sustentará el C. Mario Diego Muñoz Hernández, declaramos que hemos revisado la tesis titulada:

“Un sistema seguro para el almacenamiento y distribución de documentos digitales con servicio de rastreo de usuarios deshonestos”


Y consideramos que cumple con los requisitos para obtener el grado de Maestro en Ciencias en Computación.

Atentamente,

Dr. Arturo Díaz Pérez



Dr. Víctor Jesús Sosa Sosa



Dr. José Juan García Hernández





CINVESTAV - IPN
Biblioteca Central



SSIT0012200