

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université des Sciences et de la Technologie HOUARI BOUMEDIENE

INSTITUT D'INFORMATIQUE

Mémoire du projet de fin d'études

Pour l'obtention du diplôme d'ingénieur d'état en informatique

Option : Réseau

Sujet :

Outil de Protection de
Documents Textuels distribués
sur les Réseaux de
Télécommunication

Proposé par :

Mr. KHELALFA Halim

Etudié par :

BENZAID Chafika
BOUABACHE Fatiha

Encadré par : C.E.R.I.S.T

Soutenu le :

Devant le jury composé de :

M. ZAFFOUNE Président
M^{me} DAHMANI Membre

PROMOTION : 70/2000

RESUME

Les récents développements des technologies dans le domaine de l'informatique et de télécommunications ont révolutionné l'accès à l'information. Aujourd'hui, nous pouvons accéder de manière quasi instantanée à des masses d'informations ; cela était inimaginable il y a quelques années. Cette révolution a aussi son côté négatif : la sécurité de l'information est devenue une préoccupation prioritaire. Dans ce mémoire, nous nous intéressons à un aspect de cette sécurité de l'information : la protection des droits d'auteurs des publications électroniques.

L'étude développée dans ce mémoire porte sur une application du *Watermarking* sur les documents textuels pour protéger les droits d'auteurs "*fingerprinting*" qui consiste à assigner à chaque récepteur un mot de code unique, ce mot de code sera inséré par la suite dans les fichiers avant leurs envois. Ainsi, si on récupère une copie illicite, il suffit d'extraire la marque pour découvrir l'origine du fraude.

A cet effet, pour insérer la marque nous avons combiné entre deux méthodes de marquages : la méthode de décalage de lignes et celle de décalage de mots et ceci pour pouvoir par la suite extraire la marque même en présence de déformation.

Mots clés : Text watermarking, Text hiding, Steganography, fingerprinting, Digital watermarking, Centroid detection, Correlation detection.

Table des matières

Introduction.....	9
Partie I: Concepts de base.....	12
I Document électronique.....	13
I.1 Introduction.....	14
I.2 Définitions	14
• Document.....	14
• Document électronique.....	14
• Format.....	15
I.3 Les formats d'un document électronique.....	15
I.3.1 Format structuré.....	15
• SGML (Standard Generalized Markup Language).....	16
• LATEX.....	16
• HyTime.....	17
I.3.2 Format présentationnel.....	17
• Langage PostScript.....	18
• Format PDF (Portable Document Format).....	18
• TEX.....	19
• TROFF.....	19
I.3.3 Format Bitmap.....	19
I.4 Conclusion	20
II Sécurité informatique.....	21
II.1 Introduction.....	22
II.2 Les problèmes de la sécurité «les menaces »	22
II.3 Les outils cryptographiques	23
II.3.1 Termes et définitions.....	24
II.3.2 Les techniques de la cryptographie	25
II.3.2.1 Chiffrement à clé secrète	25
II.3.2.2 Chiffrement à clé publique	27
II.3.2.3 La signature digitale.....	28
II.3.2.4 Les fonctions de scellement.....	29
II.4 Les mécanismes de sécurité	30
II.4.1 L'authentification et l'identification	30
II.4.2 Le contrôle d'accès	31
II.4.3 Sécurité des communications.....	32
II.5 Les outils de la sécurité.....	34
II.5.1 Les outils cryptographiques.....	34
II.5.2 DSS (Digital Signature Standard).....	35
II.6 Conclusion	35
III Stéganographie et Watermarking.....	37
III.1 Introduction.....	38
III.2 La stéganographie.....	38
III.2.1 Définition	38
III.2.2 Où et comment cacher l'information secrète ?.....	39
III.2.2.1 Texte.....	40
III.2.2.2 Son.....	40
III.2.2.3 Image.....	40
III.2.2.4 Autres	40
III.3 Le watermarking	41
III.3.1 Définitions.....	41
III.3.2 Les objectifs du watermarking.....	41
III.3.3 Caractéristiques d'une bonne marque	42
III.3.4 Les modèles de marquage.....	43
III.3.5 Les champs d'application du watermarking	45
III.3.5.1 Le texte	45
III.3.5.2 L'image.....	48
III.3.5.3 Audio.....	51

III.3.6	Attaques sur les marques.....	52
III.3.6.1	Selon le but de l'attaque.....	52
III.3.6.2	selon le type de l'attaque.....	54
III.4	Conclusion.....	55
Partie 2:	Etat de l'art du watermarking des documents textuels.....	56
IV	Le marquage du texte.....	57
IV.1	Introduction.....	58
IV.2	Les techniques de marquage.....	58
IV.2.1	Marquage par décalage de lignes.....	58
IV.2.2	Marquage par décalage des mots.....	59
IV.2.3	Marquage des caractères.....	59
IV.2.4	Comparaison.....	60
IV.3	Les déformations.....	60
IV.3.1	Le bruit (<i>noise</i>).....	60
IV.3.2	L'inclinaison (<i>skewing</i>).....	61
IV.3.3	Flou (<i>blurring</i>).....	61
IV.3.4	Atténuation (<i>Fading</i>).....	61
IV.3.5	Expansion/Rétrécissement (<i>Expansion/Shrinkage</i>).....	61
IV.4	La détection.....	61
IV.4.1	Les profils.....	62
IV.4.1.1	Le profil horizontal.....	63
IV.4.1.2	Le profil vertical.....	63
IV.4.2	Les méthodes de détection.....	64
IV.4.2.1	La méthode de corrélation.....	64
IV.4.2.2	La méthode du barycentre.....	66
IV.4.3	Comparaison entre les performances des deux méthodes.....	68
IV.4.3.1	Modèle.....	68
IV.4.3.2	Le décalage de ligne vs. mot - méthode du barycentre.....	70
IV.4.3.3	Le décalage ligne vs. mot - méthode de corrélation.....	71
IV.4.3.4	Détection par la méthode du barycentre vs. de corrélation - décalage ligne.....	72
IV.4.3.5	Détection par la méthode du barycentre vs. de corrélation - décalage mot.....	73
IV.4.4	Autres règles de décision.....	73
IV.4.4.1	La détection par utilisation des lignes de base.....	74
IV.4.4.2	La détection par utilisation des barycentres.....	74
IV.5	Les attaques.....	74
IV.5.1	Déjouement de la méthode de codage par décalage de ligne.....	75
IV.5.2	Déjouement de la méthode de codage par décalage de mot.....	76
IV.5.3	Déjouement de la méthode de codage des caractères.....	77
IV.6	Conclusion.....	78
V	Analyse du document.....	79
V.1	Introduction.....	80
V.2	Notions de base.....	80
V.2.1	Seuillage global.....	82
V.2.2	Seuillage local.....	82
V.2.3	Seuillage adaptatif.....	82
V.3	Estimation du bruit <i>salt and pepper</i>	83
V.3.1	Filtres morphologiques.....	83
V.3.2	Filtre k-Fill.....	84
V.3.3	Filtre k-Fill amélioré.....	85
V.4	L'inclinaison.....	86
V.4.1	Algorithme de Baird.....	87
V.4.2	Algorithme de Hinds et al.....	87
V.4.3	Algorithme de O'Gorman.....	88
V.4.4	Algorithme de Le et al.....	89
V.4.5	L'algorithme de Amin et al.....	89
V.5	Segmentation.....	91
V.6	Conclusion.....	92
Partie 3 :	Conception et Implémentation.....	93
VI	Conception.....	94
VI.1	Introduction.....	95

VI.2	L'encodeur.....	95
VI.2.1	Génération des marques	96
VI.2.2	Le fichier PostScript.....	97
VI.2.2.1	La structure générale d'un fichier PS.....	97
VI.2.3	Le prétraitement.....	100
VI.2.4	L'insertion de la marque.....	100
VI.2.5	Conversion en BMP	102
VI.3	Le décodeur.....	102
VI.3.1	La phase de prétraitement.....	104
VI.3.1.1	Binarisation.....	104
VI.3.1.2	Filtrage et désinclinaison	104
VI.3.1.2.1	Le filtrage du bruit " <i>Salt and Pepper</i> ".....	104
VI.3.1.2.2	La désinclinaison du document.....	104
VI.3.1.3	Segmentation.....	105
VI.3.2	La phase de détection.....	105
VI.3.2.1	Calcul des profils horizontaux et segmentation des lignes(Image originale)	106
VI.3.2.2	Calcul des profils verticaux et segmentation des mots(Image originale)	106
VI.3.2.3	Calcul des probabilités d'erreurs.....	107
VI.3.2.4	Détection par la méthode de barycentre (décalage de ligne).....	108
VI.3.2.5	Détection par la méthode de corrélation (décalage de mots).....	108
VI.3.2.6	Consultation de la base de donnée.....	108
VII	implémentation.....	109
VII.1	l'encodeur.....	110
VII.1.1	Environnement de programmation.....	110
VII.1.2	La phase de prétraitement.....	110
VII.1.3	La phase d'insertion	111
VII.2	Le décodeur	114
VII.2.1	Environnement de programmation.....	114
VII.2.2	Binarisation	115
VII.2.3	Calcul des profils horizontaux et segmentation des lignes	116
VII.2.4	Calcul des profils verticaux et segmentation des mots	116
VII.2.5	calcul des barycentres.....	117
VII.2.6	le détection	117
VII.2.6.1	Détection par la méthode des barycentres.....	117
VII.2.6.2	Détection par la méthode de corrélation	117
Conclusion	119
Annexes	122
Références bibliographiques	140