الجمـــهوريـــة الجـــزائريـــة الديـمقراطيـــــــة الشـــعبيـــــة
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعـلـيـم الـعـالـــي و الـبـحـث الـعـلـمي
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

جـامـعـة عمّـــار ثـلـيـــجـــي بالأغـــواط
UNIVERSITE AMAR TELIDJI LAGHOUAT

كـلـيـــــة الـعـلـــــوم
FACULTE DES SCIENCES

DEPARTEMENT DE MATHEMATIQUES ET INFORMATIQUE

Mémoire de MASTER

Domain : Mathématiques et Informatique

Filière : Informatiques

Option : Réseaux, Systèmes et Applications Réparties

Par :
Laiza Elhouiti

THEME

Nouvelles méthodes sémantiques pour la stéganographie dans la langue arabe

Soutenu publiquement devant le jury composé de :

Mr. ***Youcef Ouinten***	*M.C.(A)*	*Président*
Mr. ***Noureddine Chaib***	*M.A.(A)*	*Examinateur*
Mr. ***Attia Nehar***	*M.A.(A)*	*Examinateur*
Mr. ***Mohamed Lahcen Bensaad***	*M.C.(B)*	*Encadreur*

Remerciements

Tout d'abord, je remercie **Dieu** *le tout puissant qui m'a donné la force et la patience d'accomplir ce modeste travail.*

Je tiens à remercier en second lieu mon directeur de recherche ***Dr.Bensaad Lahcen***
pour toute sa gentillesse, son aide, son entière disponibilité et ses précieux conseils, sans lesquels ce travail n'aurait pu aboutir.

Mes vifs remerciements vont également aux ***membres du jury****, qui ont accepté l'évaluation de ce modeste travail et de l'enrichir par leurs propositions.*

Je remercie tous mes enseignants du département de mathématiques et d'informatique

Enfin, je tiens également à remercier toutes Les personnes qui ont participé de près ou de loin à la réalisation de ce travail.

Merci à toutes et à tous

DEDICACES

Je dédie ce mémoire

À mon père (Que Dieu le protège)

À ma mère (Que Dieu la protège) qui a éclairé mon chemin et qui m'a encouragé et soutenue tout au long de mes études

À mon cher frère et chères sœurs pour leurs aides précieuses et leurs encouragements qu'ils m'attribuaient

À mes oncles et mes tantes

À tous mes fidèles amis

À tous les professeurs et enseignants universitaires qui m'ont permis, par leurs e*ff*orts, d'atteindre un tel niveau de formation.

À tous mes collègues de promotion master informatique.

ملخص

منذ الأزل، والإنسان دائما يحاول الحفاظ على أغلب معلوماته خاصة المهمة ويبقيها سرية قدر الإمكان. لذلك نجد أنه خاصة مع تطور التكنلوجيا اقترحت العديد من الطرق من أجل الحفاظ على سرية التواصل، ومن بين أهم الطرق هي إخفاء المعلومات باستخدام خصائص الصورة والصوت والنص، ويعتبر هذا الأخير أصعب الأنواع، لأن أي تغير فيه يكون ملاحظا وجالبا للانتباه لذلك كان العمل على النصوص قليل جدا خاصة العربية حيث لم يبدأ العمل عليها إلا منذ سنة 2006 وتستخدم أغلب الطرق المقترحة خصائص الحروف والتشكيل.

على الرغم من الكمية العالية للمعلومات التي يمكن اخفاؤها إلا أنها تعتبر أقل صلابة واقل أمن لأنها غالبا ما تجلب الانتباه.

لذلك قمنا باقتراح طرق جديدة تعمل على إخفاء المعلومات في معنى النص دون شكله، باستغلال بعض قواعد اللغة التي تمكننا من التغيير في تركيب الجملة دون أن تغير في المعنى المقصود.

بينت النتائج بعد تطبيق الطرق المقترحة على نصوص عربية مختلفة أنها أكثر صلابة وأكثر أمن من الطرق المقترحة قبل في اللغة العربية على الرغم من أن كمية المعلومات المخفية كانت قليلة بالنسبة للطرق الأخرى.

الكلمات المفتاحية

إخفاء المعلومات، الكتابة الخفية، أمن المعلومات، العلامة الرقمية المائية، النص العربي

Résumé

Depuis les temps anciens, les hommes ont toujours voulu préserver leurs informations secrètes aussi privée que possible. Pour cela, nous constatons qu'avec le développement de la technologie, plusieurs techniques ont été proposées afin de maintenir la confidentialité des communications. L'une des méthodes proposées est la dissimulation d'information en utilisant les propriétés de l'image, du son ou du texte. Ce dernier est considéré comme étant le genre de dissimulation le plus difficile à appliquer car tout changement dans le texte peut facilement attirer l'attention. C'est pour cette raison que Les travaux dans les textes sont peu, en particulier dans la langue arabe, où le travail a commencé seulement à partir de 2006 et la plupart des méthodes proposées travaillent sur les caractéristiques des lettres ou les diacritiques.

Malgré les grandes capacités qu'offrent ces méthodes de dissimulations, elles ne restent néanmoins pas aussi robustes et sécurisées qu'on le souhaiterait.

C'est pour cette raison nous allons proposer des nouvelles méthodes en nous basant sur la sémantique du texte, en utilisant quelques règles de grammaire arabe qui nous a permis de changer la structure de la phrase sans changer le sens.

Après le teste des méthodes proposées sur des différents corpus, il nous semblent que nos méthodes proposées sont plus robustes et plus sécurisé que les autres proposées avant dans la langue arabe malgré que la capacité est faible.

Les mots clés

Dissimulation de l'information, Stéganographie, Sécurité de l'information, Tatouage numérique, Texte Arabe.

Abstract

Since ancient time, man is always trying to preserve the most important private information and keep it confidential as possible. Therefore, we find that, especially with the development of technology, many ways suggested maintaining the confidentiality of communication.

Among the most important modern ways of concealing information by exploiting the image, sound and text properties of the latter is the most difficult types because any change in it will be observing and bringing to attention. That's why the work in the texts are very few, especially in the Arabic language, where work did not begin only since 2006, and the most of proposed methods used the characteristics of letter and diacritics to hide information.

Although, the high quantity of information that can be hidden with this methods, they are considered less robust and often bring attention.

Therefore, we will propose new methods to hide information in the meaning of the text by using some grammar rules without a change in meaning.

The results showed after applied the proposed methods in the various Arab texts they are more robust and more secure of the proposed methods before in the Arabic language Although the amount of hidden information was a few.

Key words

Information hiding, Steganography, Information security, Watermarking, Arabic text

Liste des Abréviations

AES : Advanced Encryption Standard.

ASCII: American Standard Code for Information Interchange.

BMP: BitMaP.

DCT: Discrete Cosine Transform.

JPEG: Joint Photographic Experts Group.

LSB : Least Significant Bit.

OCR : Optical Characrer Recognition.

PDF : Portable Document Format.

Terminologie

Stégo-médium : le fichier après l'insertion du message secret.

Extraction : le fait de faire sortir le message secret caché dans le Stégo-médium.

Médium de couverture : le médium dans lequel nous voulons cacher Les informations.

Table des matières

Table des figures

Liste des tableaux

Chapitre 1

Introduction

1.1 Préambule

Avec le développement de la technologie et les différents supports de multimédias, il est devenu facile de stocker une grande quantité d'informations (les documents, les images,...) sous des formats numériques (pdf, jpeg...) et cela en utilisant très peu d'espace. Pour la protection des données, elles sont souvent sécurisées par des outils de cryptographie lors de leur transmission, elles peuvent être détectées facilement par un pirate en essayant de casser l'algorithme. En plus, les données ne sont plus chiffrées après la réception et l'affichage. Ainsi, les données sont devenu facile à utiliser. Comme par exemple si une personne récupère un document et le copier sur CD ou internet...et elle l'utilise pour n'importe qu'elle raison qu'elle soit légale ou non car dans ce cas, le droit d'auteur est souvent négligé.

C'est pour cette raison, que dans les années 1990 une importante recherche est apparu sous le nom anglais «watermarking» que nous pouvons traduire en français en : «tatouage numérique».

Le tatouage consiste à insérer une marque invisible à l'intérieur des documents numériques transitant par les réseaux afin de lutter contre le piratage et d'assurer la protection des droits de propriété intellectuelle en insérant une identification imperceptible et indétectable par tout système ignorant son mode d'insertion [1].

1.2 Motivation

Malgré les nombreux avantages de la technologie, elle reste une arme redoutable pour tous ceux qui veulent l'utiliser à des fins malsaines.

Pour cela, plusieurs problèmes ont été posés parce que l'homme veut toujours préserver sa confidentialité aussi secrète que possible alors que d'autres veulent utiliser ses informations contre lui. Plusieurs techniques ont été proposées telle que la cryptographie, mais avec le développement de la technologie, il est devenu possible de casser les algorithmes de cryptographie et de trouver le message caché.

Pour cela, dans le cadre de ce mémoire, nous allons travailler sur une autre technique de protection des informations : « la stéganographie » qui consiste à cacher l'information sans attirer l'attention du pirate sur l'existence du message.

Plusieurs médias, sont utilisés pour dissimuler l'information comme les images, les vidéos, l'audio et les textes. Parmi eux nous avons choisi de travailler sur le texte parce qu'il est non remarquable et ajouter une marque ou une information cachée dans un texte permet aussi de protéger les droits d'auteurs. Ainsi que la petite taille des fichiers textes qui permet d'utiliser plusieurs fichiers pour dissimuler un message.

Parmi les textes choisis, nous avons opté pour des textes arabes parce qu'il y a peu de travaux sur eux. D'autre part, la langue arabe est une langue trés riche et ses caractéristiques lui permettent d'appliquer plusieurs méthodes de stéganographie sans attirer l'attention.

1.3 Objectif

Notre principal objectif dans ce choix d'étude est de proposer des nouvelles méthodes de dissimulation d'information dans le texte arabe. Ces méthodes utilisent la semantique de la langue arabe, afin de cacher des messages et de protéger les droits de propriétés intelectuelles.

1.4 Plan du mémoire

Ce mémoire s'intéresse aux méthodes de stéganographie dans le texte arabe. Il est structuré en quatre chapitres :

Le chapitre en cours est le chapitre 1.

Le deuxième chapitre 2 a pour but de définir la stéganographie, de retracer son historique ainsi que ses types, les différentes techniques et supports utilisés. Nous allons pour la suite comparer les différentes techniques de dissimulation avec la cryptographie, après nous allons parler de la sécurité dans la stéganographie et enfin une conclusion.

Le troisiéme chapitre 3 s'intéresse à la stéganographie dans le texte arabe. Nous avons commencé par donner une vue générale sur la langue arabe. Ensuite, nous avons éxpliqué quelques méthodes utilisée dans la stéganographie texte, détaillé les méthodes utilisées dans le texte arabe depuis 2006 et enfin nous avons fait une étude comparative entre ces méthodes selon leurs capacités, robustesses et visibilités et nous avons terminé par une conclusion.

Le quatrième chapitre 4 présente notre contribution qui consiste à proposer des nouvelles méthodes de stéganographie dans le texte arabe en utilisant les grammaires de la langue arabe afin de montrer après le test sur différents corpus qu'elles sont plus robustes et plus sécurisées que les autres méthodes proposées avant elles, ainsi que les avantages et les inconvénients de nos méthodes et les améliorations proposées et enfin une conclusion.

Nous avons términé par une conclusion générale 4.9 et quelques perspéctives
.

Chapitre 2

Stéganographie

2.1 Introduction

L'homme a toujours utilisé des techniques de dissimulation pour protéger et préserver des informations qu'il juge importantes. La stéganographie est une des plus anciennes techniques utilisées pour dissimuler un message à l'intérieur d'un autre message de manière à ce que l'on ne puisse pas le détecter.

Ainsi, la stéganographie n'a rien d'une technique récente puisqu'elle a déjà été utilisée par les anciens grecs, romains, chinois, arabes et a beaucoup servi durant les deux guerres mondiales. Mais contrairement aux anciens stéganographes qui dissimulaient leurs messages sous des tatouages sur la peau humaine ou l'utilisation d'animaux, de tablettes de cire, de jeux de mots, de peinture, de l'encre secrète ou de poupées ...etc. L'ère numérique a pris le relais et de nos jours, les messages sont dissimulés sous des images, des audios, des vidéos, des textes ... etc. C'est ce qu'on appelle la stéganographie moderne ou la stéganographie numérique.

Dans ce chapitre, nous allons dans un premier temps définir la stéganographie et les différentes méthodes de dissimulation de données, retracer l'historique de la stéganographie, définir la cryptographie et la comparer avec les méthodes de dissimulation, nous allons aussi expliquer les types de stéganographie, les supports et les techniques, la sécurité et nous allons terminer par une conclusion.

2.2 Stéganographie

La stéganographie est l'art de la dissimulation. Le nom stéganographie a été créé par JOHANNES TRITHEMIUS [1] (1462-1516) dans STEGANOGRAPHIA l'un des tout premiers travaux sur la cryptographie et la stéganographie.

Le terme stéganographie est dérivé des mots « steganos » : couverts et « graphia » :écriture [2]. Elle consiste à cacher un message au sein d'un autre message qui en apparence, il semble anodin, de sorte qu'on ignore l'existence même du secret. La stéganographie suit un processus de dissimulation d'information qui repose sur les deux opérations suivantes[3] :

1. **La dissimulation** :elle consiste à insérer l'information dans le médium.
2. **L'extraction** :elle consiste à récupèrer l'information cachée.

On en distingue trois principaux schémas de dissimulation d'information, comme nous la montrons dans la figure 2.1

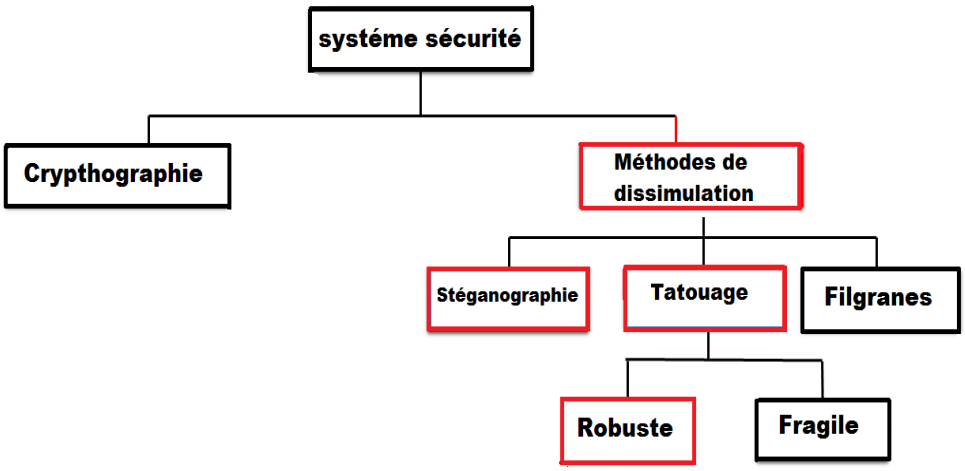


FIGURE 2.1 – Le Systéme de sécurité

La figure montre les deux principaux systémes de sécurité : cryptographie et les méthodes de dissimulation, ainsi les trois types utilisés pour la dissimulation et les deux types du tatouage.

1. Johannes Trithemius : est un abbé bénédictin allemand célèbre pour ses découvertes en cryptologie

2.2.1 La stéganographie

Cette technique cherche à cacher un message secret (texte, image, audio...) dans un médium (texte, image, audio...) de sorte que personne ne puisse distinguer un médium vierge d'un stégo-médium[4].

2.2.2 Le tatouage numérique

Le tatouage numérique est une technique qui est utilisé généralement pour protéger les droits d'auteurs on ajoutant une marque invisible au document sans modifier leur contenu.

2.2.2.1 Les types de tatouage numérique

Les techniques du tatouage peuvent être classées en deux types : fragile et robuste [3].

Pour **le tatouage fragile**, la modification du stégo-medium mène à la destruction d'une donnée cachée [5]. Ce type est utilisé surtout pour vérifier l'intégrité du document. Il est appliqué par exemple dans : les photos d'identités (carte d'identité, passeport, permis de conduire), les images médicales (scanner, IRM,...) ou encore militaires[6].

Mais si les changement appliqués dans un objet stégo pour éliminer les données cachées le rendrait inutile on parle du **tatouage robuste** où la destruction de la donnée cachée méne à la destruction de l'objet stégo[5].

2.2.2.2 Les approches du tatouage

Pour tatouer des données, trois grandes approches ont été utilisés, et chaque approche dépend de type des données et de l'utilisation que l'on en fait[7]. Nous pouvons :

1. Introduire de nouvelles informations.
2. Réorganiser des informations selon un ordre défini.
3. Altérer ou modifier des informations existantes.

2.2.2.3 Les conditions d'algorithmes de tatouage

Chaque algorithme de tatouage doit satisfaire les contraintes suivante :l'imperceptibilité, la robustesse et la capacité ou le débit utile.

1. **Imperceptibilité**

   Elle consiste à cacher la marque dans le document de sorte que le tatouage soit invisible pour ne pas dégrader la qualité du document[8].

2. **Robustesse**

   Elle consiste à cacher l'information de telle sorte que la destruction des données cachées ne soit pas facile, sauf dans le cas du tatouage fragile dédié au contrôle d'intégrité.

3. **la capacité ou le débit utile**

   La capacité sert à la quantité des bits qui peuvent être tatouer dans un document. Elle est généralement faible surtout si nous prenons en concidération la robustesse et l'imperceptibilité.

   "Typiquement, elle est de l'ordre de 64 bits pour les applications de protection de la propriété intellectuelle, ce qui correspond à la taille d'un numéro standard délivré par une société d'auteurs (par exemple un numéro ISAN [1])"[9].

2.2.2.4 Les applications du tatouage numérique

Parmi les applications classique du tatouage[10] :

1. **Protection des droits d'auteurs**

   La protection des droits d'auteur, est l'application la plus utilisée aujourd'hui, elle sert à insérer dans un document une marque spécifique pour protéger les droits d'auteur.

2. **Protection contre la copie**

   Le tatouage protège aussi les CD,DVD ,..contre la copie, afin d'interdire une circulation de média illégale.

3. **Sécurité médicale**

   Le tatouage dans le domaine médicale sert à insérer une marque ou identifiant dans la radio du patient afin d'assurer la correspondance entre le patient et la radio, pour éviter toutes confusion.

1. ISAN : International Standard Audiovisual Number

4. **Indexation**
   Le tatouage est utilisé aussi pour faciliter une recherche dans une base de donnée, en insérant une marque dans le document, il est utilisé plus dans les images. L'indexation des images permet aussi de les classer selon leurs contenu de manière automatique.

2.2.3 Les filigranes numériques

Selon[3], les filigranes numériques ou ce qu'on nous appelons couramment les empreintes digitales sont les marques laissées par les sillons des pulpes digitales. Pour protéger les informations contre le vol, la falsification ou un accès non autorisé. Elles visent aussi à vérifier l'identité des personnes à l'aide d'une ou de plusieurs modalités (voix, empreinte digitale, visage, iris...).

2.3 Historique de la stéganographie

La stéganographie a été utilisé à travers l'histoire depuis 2500 ans par différentes civilisations, en particulier avant la naissance de cryptographie et de nombreuses techniques ont été utilisées. En voici quelques exemples de l'utilisation de la stéganographie dans l'histoire :

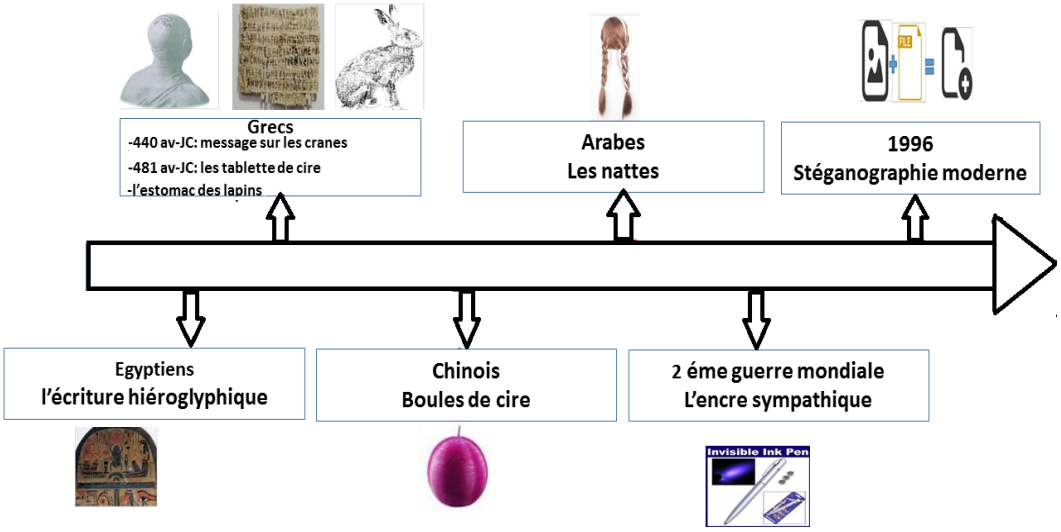


FIGURE 2.2 – Historique de la steganographie

La figure 2.2 résume l'historique de stéganographie depuis son apparition et les méthodes utilisées par les différents civilisation.

2.3.1 La stéganographie chez les Égyptiens

Les égyptiens ont utilisé l'écriture hiéroglyphique pour cacher leurs messages. Certains hiéroglyphes étaient travaillés de telle sorte que seules ceux qui savaient ce qu'il faut chercher pouvaient les lire correctement[11].

2.3.2 La stéganographie chez les Grecs

Les grecs ont utilisé plusieurs techniques pour cacher leurs messages. Parmi ces techniques, nous allons citer trois d'entre elles :

1. **Les crânes des esclaves :** Les esclaves étaient utilisés en tant que messager. Les grecs inscrivaient le message sur leurs crânes rasés et une fois les cheveux poussés, ils pouvaient être envoyés[12].
2. **La tablette de cire :** Cette technique a été utilisée pour la première fois par Demarate [1], ancien roi de Sparte. Lorsqu'il se réfugia auprès du roi des perses, Xerxès [2] 1er, Demarate fut mis au courant d'un projet d'invasion de la Grèce. Il décida des lors de prévenir les siens en toute discrétion. Il utilise ainsi le stratagème suivant :

   « Il prit une tablette double, en gratta la cire, puis écrivit sur le bois même les projets de Xerxès ; ensuite il recouvrit de cire son message : ainsi le porteur d'une tablette vierge ne risquait pas d'ennuis. Les tablettes étant arrivées à Sparte, la reine Gorgô qui connaissait la technique fit gratter la cire et découvrit ainsi le message de Demarate »[12].
3. **l'estomac des lapins chassés :** Les grecs utilisaient beaucoup cette technique pour dissimuler les messages à l'intérieur de l'estomac du lapin. Le chasseur pouvait ainsi se déplacer avec l'animal sur son épaule, sans prendre le risque d'être soupçonné.

2.3.3 La stéganographie chez les chinois

Contrairement aux grecs qui utilisaient l'estomac des animaux, les chinois ont pour leur part utilisé l'estomac des êtres humains pour cacher une

1. Demarate : est un roi de Sparte (515 à 491 av. J.-C.), de la maison des Eurypontides.
2. Xerxès : né vers -519, mort en -465, grand roi perse, membre de la dynastie des Achéménides.

information qui pouvaient être primordiale pour la sécurité de leurs pays. Ainsi, ils dissimulaient le message à l'intérieur de petites boules en soie cirées. Ces dernières étaient avalées par un messager et une fois arrivé à destination, il suffisait d'attendre que la nature fasse son œuvre pour pouvoir découvrir le message envoyé[12].

2.3.4 La stéganographie chez les Arabes

Plusieurs techniques ont été utilisées par les arabes dans ce domaine, mais nous avons choisi une technique qui n'a pas été mentionné dans les livres de stéganographie[5]. Il s'agit de la méthode utilisée par HATIB IBN ABI BALTA'AH [1]qui a utilisé les nattes d'une femme pour dissimuler un message qu'il voulait envoyer à Koraiche. Les détails de cette technique ont été décrite dans le **HADITH** suivant :

« Directement de 'Ali ben Abd Alah directement de Sufyane qui dit 'Amru ben Dinar nous a apporté, je l' ai entendu par deux fois en disant :Hassan ben Muhammad m'a apporté en disant : j'ai entendu 'Ali(r) dire "le prophète (que la paix et le salut d'Allah soient sur lui)m'envoya [en mission] avec az-Zubayr et al Muqdad ben al-Aswad il [nous] dit "Allez vers la mare de khakh. Il y a là-bas une femme qui a avec elle une missive. Prenez d'elle cette lettre !" "en effet, nous démarrâmes au galop de nos chevaux. A notre arrivée à la mare, nous trouvâmes la femme, -nous lui dimes alors "sors la lettre" -"je n'ai pas de lettre !"nia t'elle -"soit que tu sortes la lettre soit que nous allons jeter les vêtements" sur ce ,elle retira la missive du cordon de ces cheveux[et nous la remit].nous l'apportâmes alors au messager de dieu (ç),c'était un message de Habit ben Abu Balta 'a à quelques gens polythéistes de la Mecque par lequel il les informait de quelque décision du messager de dieu(que la paix et le salut d'Allah soient sur lui) »[13].

Les anciens arabes ont également beaucoup utilisés la stéganographie dans leurs poèmes et mots pour cacher les messages et avec la richesse de leurs langue et l'utilisation des métaphores, ils ont pu y cacher leurs intentions réelles derrière des mots explicites.

2.3.5 La stéganographie durant la deuxième guerre mondiale

La stéganographie a été énormément utilisée durant les deux guerres mondiales particulièrement pour la seconde guerre mondiale qui a vu de

1. Hatib ibn Abi Balta'ah : était l'un des sahaba

nombreuses formes de stéganographie, comme ce message qui a été envoyé par un espion allemand :

« **Apparently neutral's protest is thoroughly discounted and ignored. Ismam hard hit. Blockade issue affects pretext for embargo on by-products, ejecting suets and vegetable oils.** ».

En français, nous avons :

« **Apparemment, la protestation des pays neutres est totalement ignorée. Is-man frappe fort. L'issue du blocus donne des prétextes pour un embargo sur certains produits, mis à part graisses animales et huiles végétales.** »

Ce message qui en apparence inoffensif est pourtant porteur de sens parce que si nous prenons la deuxième lettre de chaque mot, nous aurons le message suivant :

« Pershing sails from NY June 1 »

qui en français nous donne :

« le Pershing part de New-York le 1er, juin. »

Les allemands ont également utilisé l'encre sympathique qui consiste à écrire un message à l'aide du jus de citron, de lait ou certains produits chimiques qui invisible à l'œil, ils ont besoin que d'une simple flamme pour révéler le contenu du message[12].

2.3.6 La stéganographie informatique

Depuis 1996, la stéganographie a énormément évolué et ceux grâce à l'informatique qui a permis aux stéganographes d'exprimer toutes leurs créativités tout en évitant d'éveiller l'attention sur les messages dissimulés[14].

Dans l'informatique numérique, les données subirent des nombreuses compressions destructives par élimination des données inutiles. Ces derniers Peuvent être remplacé par des données utiles qui seront les données que nous voulons cacher. Pour cela, nous pouvons utiliser tous les fichiers numériques comme les images, les sons, les vidéos,...etc[3].

2.4 Cryptographie

La cryptographie ou l'art du chiffrement est une science utilisée pour transformer un message pour le rendre illisible. Cette transformation est appelée le chiffrement, le déchiffrement est l'action qui permet de le reconstruire. Pour chiffrer un message, nous avons besoin d'un algorithme qui dépend d'un paramètre appelé clef.

La cryptanalyse, à l'inverse, est l'étude des procédés cryptographiques dont l'objectif principal est de trouver des faiblesses et, en particulier, de pouvoir décrypter des textes chiffrés[5].

Plusieurs techniques peuvent être utilisées. Nous avons choisi le chiffrement de CÉSAR. Ce crypto système consiste à remplacer chaque lettre du texte clair, par une lettre différente, située x lettres après dans l'alphabet, où x est la valeur de la clé passée en argument[15].

Exemple

Si nous codons le mot « MASTER »en utilisant la valeur 3 pour la clé du César, l'alphabet est décalé de manière à commencer à la lettre D.

L'alphabet

« **ABCDEFGHIJKLMNOPQRSTUVWXYZ** »

Si nous décalons le début de 3 lettres, nous obtenons :

« **DEFGHIJKLMNOPQRSTUVWXYZABC** »

Où D = A, E = B, F = C... etc.

Comme montré dans la figure2.3 [15] qui explique le décalage des lettres par trois positions, comme le montre la figure

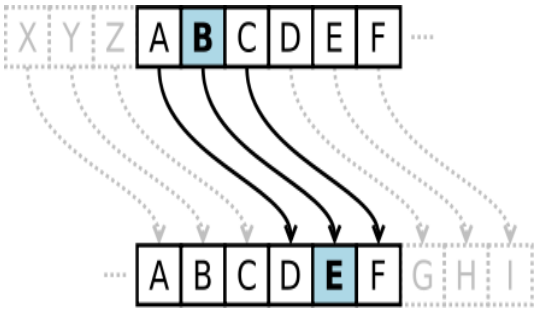


FIGURE 2.3 – chiffrement de césar

Avec ce procédé, le texte en clair « MASTER »est crypté en « PDVWHU ». Pour autoriser un autre utilisateur à lire le texte chiffré, indiquez-lui que la valeur de la clé est égale à 3.

2.5 Comparaison

Dans cette partie, nous allons comparer les trois méthodes de dissimulation que nous avons cité précédemment avec la crypthographie[16, 17].
Les critères de Comparaison que nous avons choisis sont :

— **Donnée** : sert au message que l'on veut le protéger.
— **Médium** : le support utilisé.
— **Clé** : l'outil utilisé pour cacher le message.
— **Outils** : les outils utilisées pour sécuriser la communication.
— **Visibilité** :la détection de message par l'œil humain.
— **Détection** :la possibilité de piratage.
— **Imperceptibilité** :la précision de sécurité.
— **Robustesse** :la solidité de la méthode.
— **Type attaque** :la façon de trouver le message secret.
— **Objectif** :les buts visés par la méthode.
— **Résultat** :l'obtenu après l'application de la méthode .

TABLE 2.1 – Comparaison des méthodes de systéme de sécurité

	Les méthodes de dissimulation			cryptographie
	Stéganographie	Tatouage	Filigranes	
Donnée	le message à transmettre	dépend de medium	Une empreinte dépendant du médium et son utilisation	le message à transmettre
Medium	sans importance	le medium dont on veut protéger les droits	le médium dont on souhaite prévenir la diffusion de copie illégale	/
Clé	optionnel	pour insérer la marque	pour insérer l'empreite	pour coder le message
Outils	Steghide...etc	eZy watermark..	visage, iris..	RSA,DES,...
Visibilité	non	oui	oui	oui
Détection	Difficile	Difficile	Difficile	Difficile
Imperceptibilité	élevé	élevé	élevé	élevé
Robustes	dépend au type de steganographie utilisé	dépend au type de tatouage utilisé	oui	oui
Type d'attaque	Stéganalyse	Traitement d'image	Traitement d'image	cryptanalyse
Objectif	Communication secrète	Protection de droits	Protection de documents	Protection des données
Résultat	Stego medium	Fichier tatoué	dépend de médium	Fichier chiffré

2.6 Types de stéganographie

Sur la base des techniques utilisées dans la stéganographie. Elles peuvent être classées en deux types [14, 18, 19]. Comme le montre la figure 2.4 qui résume les deux types de stéganographie, ainsi les méthodes utilisées dans chaque type :

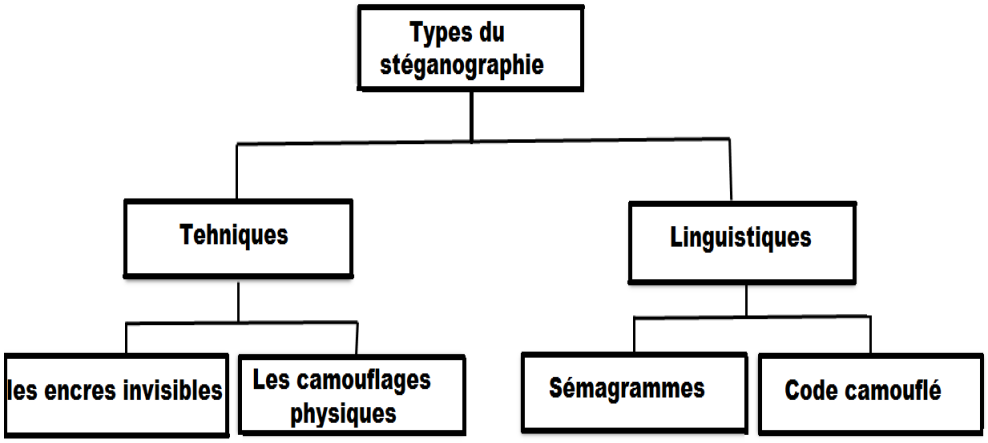


FIGURE 2.4 – Les types de steganographie

2.6.1 La stéganographie linguistique

Cette technique sert à dissimuler le message en utilisant un objet de façon à ce que le pirate ne puisse pas remarquer la présence du message. Elle peut se diviser en deux types :

2.6.1.1 Le sémagramme

Technique qui permet de transmettre le message sans utiliser aucune lettre ou chiffre, seulement des symboles qui expliquent le sens. Exemple :

Selon D. Kahn[19], le premier sémagramme utilisé consiste à passer un fil à travers des trous qui représentent l'alphabet dans un osselet, un morceau de bois et chaque trou représente une lettre ...etc.

2.6.1.2 Le code camouflé : les nulles

Elle consiste à marquer avec un signe particulier d'une façon non remarquable les lettres qui composent le message. Plusieurs méthodes ont été utilisées. Par exemple :

Insérer dans un bagage un livre ou un autre document sur lequel les lettres d'une ligne avaient été piquées de piqûres minuscules et tout à fait invisibles sauf pour le destinataire.

2.6.2 La stéganographie technique

Contrairement à la stéganographie linguistique, la stéganographie technique utilise des moyens physiques pour dissimuler le message. Elle a été utilisée par les premiers sur plusieurs formes expliquées dans l'historique. Comme :

— Les encres invisibles
— Les camouflages physiques

2.7 Supports et techniques de stéganographie

Dans cette partie nous expliquons quelques supports numériques les plus utilisés pour dissimuler des données : l'image, le son et le texte...etc. La figue 2.5 résume les différents supports et techniques de stéganographie expliqués dans cette partie.

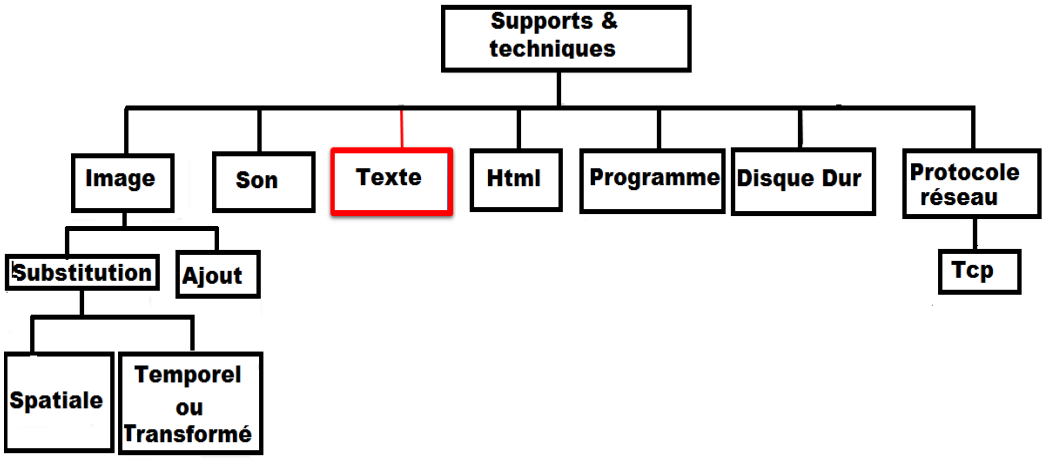


FIGURE 2.5 – Supports et techniques de la stéganographie

2.7.1 Image

Plusieurs algorithmes ont été développés pour cacher les données dans une images depuis l'apparence du stéganographie numérique.

Nous pouvons classer les algorithmes proposés selon les altérations faites sur l'image originale pour inserer le message secret[20, 21, 22] en :

— Des algorithmes qui se basent sur la substitution qui consiste à remplacer les parties insignifiantes du message originale (image,audio,..) avec des données secrètes. Les parties insignifiantes sont les bits de poids faible, et peuvent être considérés dans le Domaines spatial (LSB...) ou dans le domaine temporel ou transformé (DCT...).
— La deuxième catégorie se base surl'ajout des données supplémentaire au stégo-medium. Comme par exemple la génération par étalement de spectre d'un signal.

2.7.2 Son

Pour cacher l'information dans le son, ils se sont basés sur les faibles variations qui sont imperceptible pour l'oreille. le bruit de fond ou Les basses fréquences peuvent contenir une grande quantité d'information.

Évidemment, ce bruit doit de préférence être transmis de façon numérique pour que les pertes de transmission n'efface pas le message caché. Pour que le bruit artificiel reste indécelable,il doit posséder les propriétés statistiques d'un vrai bruit de fond [3, 21, 23].

2.7.3 Texte

La dissimulation des données dans le texte est une chose bien particulière, comme nous allons le voir en détail dans le chapitre suivant.

2.7.4 HTML

Certains algorithmes du stéganographie proposent d'utiliser le code HTML pour cacher le message, ils font des modification au code source d'une page HTML pour camoufler le fichier secret en insérant des espaces entre les balises, variant minuscules et majuscules dans les balises,... Mais le probléme est que cela attire l'attention et peut etre détecter seulement par une analyse du code source ou même par un coup d'oeil [3].

2.7.5 Programmes

Une autre méthode qui utilise les "zones mortes" du code du programme comme les commentaires, branche morte d'un organigramme ou dans le programme lui-même à l'aide d'une commande jamais utilisée (quintuple clic) pour cacher l'information [3].

2.7.6 Disques dur ou CD

Il est aussi possible de cacher des fichiers à l'intérieur de l'espace d'un disque dur ou d'une disquette.« Car il faut bien différencier 2 choses : ce qui est inscrit sur le disque et ce que y est inscrit dans la table d'allocation du disque géré par le système d'exploitation. En effet pour répertorier tous les fichiers d'un disque, le système d'exploitation accède et met à jour une table d'allocation des fichiers : seuls les entrées de fichiers figurant dans cette table sont accessibles par le système d'exploitation. L'idée est alors d'inscrire un fichier physique sur le disque dur sans que la table d'allocation en ait connaissance : ainsi le fichier est sur le disque mais le système d'exploitation ne le voit pas. Pour ce faire il suffit d'utiliser un logiciel spécialisé écrivant et lisant directement sur le disque

sans passer par le système d'exploitation. Un tel système de stéganographie est efficace si l'intercepteur n'en a pas conscience, par contre si celui-ci est au courant alors il n'aura aucune difficulté pour retrouver le fichier. L'autre point faible de cette technique est que le disque ne doit subir aucune modification en écriture par le système d'exploitation une fois que la stéganographie a été effectuée car sinon le système d'exploitation pourrait écraser sans en avoir connaissance le fichier caché car les secteurs du disque abritant ce fichier sont considérés comme espace libre pour le système d'exploitation »[3].

2.7.7 les protocoles de réseaux

Une autre technique utilise les protocoles réseaux tel que le protocole IP, le Protocol VoIP (Voice over Internet Protocol) et TCP. Par exemple dans le protocol TCP la dissimulation est faite à l'aide du numéro de séquence pour cacher 1 ou 2 octect par connexion. exemple :

dans le coté client, avant d'envoyé un octect, on le multiplie par un multiplicateur K constant pour obtenir le nouveau numéro de séquence, le multiplicateur K est partagée entre le client et le serveur.

Dans le côté client le numéro de séquence est calculé par la formule suivante :

$$\textbf{Valeur ASCII} \times \textbf{nombre K} = \textbf{Numéro de Séquence}$$

Du côté du serveur, pour récupérer les caractères cachés, il suffit de diviser le numéro de sequence recuperer par K.

$$\textbf{Numéro de séquence DIV K} = \textbf{valeur ASCII}$$

Pour plus de sécurité, on choisi le numéro de séquence aléatoirement et l'opérateur XOR peut être utilisé à la place de la multiplication [24].

2.7.8 Autres

Plusieurs autres techniques peuvent être utilisées pour cacher le message commes les systémes d'exploitation, les dispositifs d'un mobile et les machine virtuelle...etc [24].

2.8 Sécurité

La steganalyse est une technique permettant d'attaquer les méthodes de stéganographie. Elle peut être appliquée par deux types de personnes :

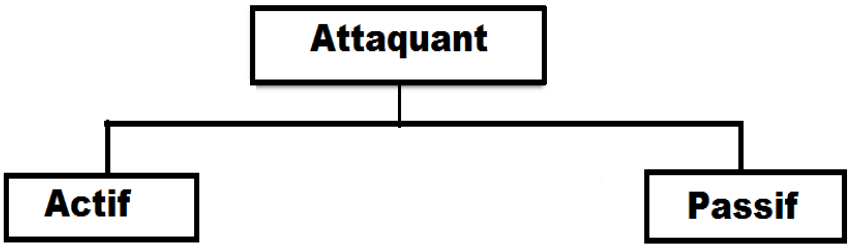


FIGURE 2.6 – Les types d'attaquants

La figure 2.6 montre les deux types d'attaquants.

2.8.1 L'attaquant actif

Qui connaît la présence de l'information et tente de la modifier ou de l'extraire. Exemple des méthodes utilisées :

Compression et décompression

Elle est utilisée dans les fichiers stéganographiques codé d'une façon simple pour inciter les pertes.

Modification des propriétés de l'image

En modifiant sa taille, faire une rotation...pour rendre le message totalement invisible.

Autres algorithmes

Il existe aussi certains programmes qui utilisent des algorithmes avancés. Comme l'outil stirMark qui manipule et teste le marquage d'image. Actuellement une autre attaque sert à couper l'image en des sous images et les rassemblant dans une seule page html [20].

2.8.2 L'attaquant passif

Le personne qui arrive à détecter la présence du message et qui fait que constater sa présence. Ce type d'attaque est utilisé dans :

— **Règles de charte**
Au niveau de l'entreprise, il est possible d'utiliser la stéganographie dans les images ou sons pour autoprotection.

— **Logiciel de détection**
Plusieurs logiciels comme StegDetect ont été développé pour détecter la présence ou non d'une information cacher sans la modifier. Mais il reste néanmoins nécessaire de se donner des limites dans l'utilisation de ces softwares[20].

2.9 Conclusion

Malgré l'efficacité de la stéganographie dans la protection des informations, les droits d'auteur...etc. Elle a été utilisé d'une façon illégale dans plusieurs domaines. Pour cela, nous trouvons que les chercheurs de stéganographie sont toujours en concurrence infinie avec les méthodes de stéganalyse, ce qui oblige les chercheurs à toujours créer des nouvelles techniques qui surpassent les essais de l'équipe adverse.

Chapitre 3

Etat de l'art

3.1 Introduction

Dans la stéganographie numérique plusieurs médias ont été utilisés tel que les images, les audios, les vidéos et les textes...etc.

Contrairement aux autres médias, la stéganographie texte est le genre le plus difficile parce que la structure du document texte est identique à ce que l'on obtient, alors que dans les autres types du document comme l'image la structure est différente de ce que nous observons. Pour cela, nous pouvons cacher l'information en introduisant les changement dans la structure du document sans apporter une modification notable.

La stéganographie dans le texte contrairement aux autres médias a été utilisée depuis des temps très anciens, cela s'est prolongé jusqu'à aujourd'hui et encore[3].

L'utilisation du texte pour dissimuler l'information est préféré parce que le texte occupe moins d'espace, communique plus d'information...ainsi que d'autres avantages.

Différentes approches et méthodes ont été tentées pour mettre en œuvre la stéganographie texte, la plupart de ces approches cachent les données en utilisant les caractéristiques des lettres de la langue. Ces approches sont généralement remarquables et attirent l'attention de l'être humain, alors que d'autres approches utilisent la syntaxique et la sémantique du texte ont été proposées dans les autres langues, mais peu de travail a été fait sur la langue arabe.

Dans ce présent chapitre, nous allons dans un premier lieu parler de la langue arabe et de ses caractéristiques dans la section 3.2. Aprés nous allons parler de les différents types de stéganographie utilisés dans le texte en expliquant quelques exemples des algorithmes proposés dans les autres langues dans la sémantique et la syntaxique dans la section 3.3.

Ensuite nous allons détaillés les différentes méthodes du stéganographie proposées dans le texte arabe depuis 2006 dans la section 3.4.

Ainsi, dans la section 3.5, nous allons faire une étude comparative entre ces derniers afin de démontrer leurs degrés de capacités, de robustesses et de visibilités . Enfin nous allons terminer par une conclusion.

3.2 La langue Arabe

L'arabe est la quatriéme langue la plus parlée dans le monde. C'est également la langue officielle dans 26 pays[25]. L'arabe est également l'une des plus anciennes langues du monde. Langue du Coran et langue liturgique de l'Islam, elle est également parlée par des millions dans le monde[26].

La langue arabe se caractérise par un système alphabétique orienté de droite à gauche et exclusivement écrit en caractères cursifs, sans majuscules. L'alphabet arabe se compose de 28 lettres qui ne s'associent jamais en bi-grams ou tri-grams. La plupart d'entre elles, changent de forme selon leurs position dans le mot (initiale, médiane, finale) et suivant les règles d'attachement de la lettre qui la précède . 6 lettres(montré dans la figure 3.1 par étoile) se singularisent, en effet, par une absence de lien graphique à la lettre suivante, de sorte que même au milieu d'un mot, une lettre peut s'écrire sous sa forme initiale ou isolée[27][28].

La figure 3.1 montre les 28 lettres arabes, et les lettres qui se singularisent sont montrées par étoile.

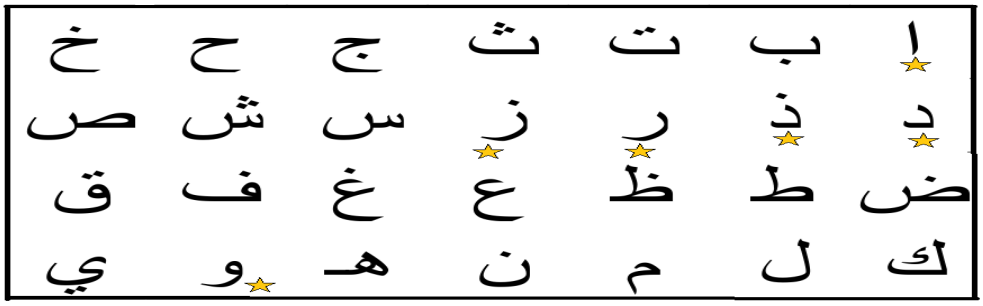


FIGURE 3.1 – Les lettres Arabes

Les lettres arabes classées en 15 pointées et 13 non pointées comme représenté dans la figure 3.2 :

Lettres pointées	Lettres non-pointées
ب، ت، ث، ج، خ، ذ، ز، ش، ض، ظ، غ، ف، ق، ن، ي	أ، ح، د، ر، س، ص، ط، ع، ك، ل، م، هـ، و

FIGURE 3.2 – Classification des lettres

Outre ces lettres, la langue arabe utilise deux autres formes graphiques particulières qui n'apparaissant qu'en fin de lemme : le 'alif maqṣûra sorte de 'alif courbé qui indique, comme lui, un prolongement de la voyelle [a] ; le t$\bar{a}$' marbûṭa dérivé de la lettre t$\bar{a}$ porté exclusivement par des substantifs, il correspond souvent à la marque du féminin. En cas d'ajout d'un suffixe, sa forme est identique au t$\bar{a}$' en position médiane.

Cet alphabet, essentiellement consonantique, peut être complété de signes diacritiques, placées au-dessus ou en dessous des lettres[29]. Les principaux sont[30] :

— la chadda : marquant le double de la consonne ou semi-consonne qui le porte.
— le soukoun : correspondant à l'absence de voyelle.
— la fatha : codant la prononciation de la voyelle brève [A].
— la kasra : codant la prononciation de la voyelle brève [i].

— la dhamma : codant la prononciation de la voyelle brève [u].
— les trois signes de tanwin : lorsque les trois voyelles brèves (fatha, kasra et dhamma) sont doublées, elles prennent un son nasal, comme si elles étaient suivies de " n " et on les prononce respectivement :
— an pour la fathatan.
— in pour la kasratan .
— un pour la dhammatan.
Les différent diacritiques sont montrés dans la figure 3.3 :

Fatha	◌َ	**Tanween fatha**	◌ً
Dhamma	◌ُ	**Tanween dham**	◌ٌ
Kasrah	◌ِ	**Tanween Kasrah**	◌ٍ
Chadda	◌ّ	**Sukun**	◌ْ

FIGURE 3.3 – Les diacritiques

Un autre caractère est le caractère d'extension arabe ou **'Tatwil'** est utilisé pour élargir l'espace entre les lettres connectées. Le Tatwil est représenté par (-) qui augmente la longueur d'une ligne de texte. Il ne peut pas être ajouté au début ou à la fin des mots. Il est utilisé pour ajuster le texte sans aucune modification dans le contenu du texte. Comme l'exemple d'extention entre la lettre "haa" et "mim" montré dans la figure 3.4 :

FIGURE 3.4 – Tatweel

3.3 Méthodes de la stéganographie dans le texte

Plusieurs méthodes ont été utilisées pour cacher l'information secrète dans un texte. La figure 3.5 résume les méthode utilisées dans le texte, ainsi quelques exemples des méthodes proposées en arabe.

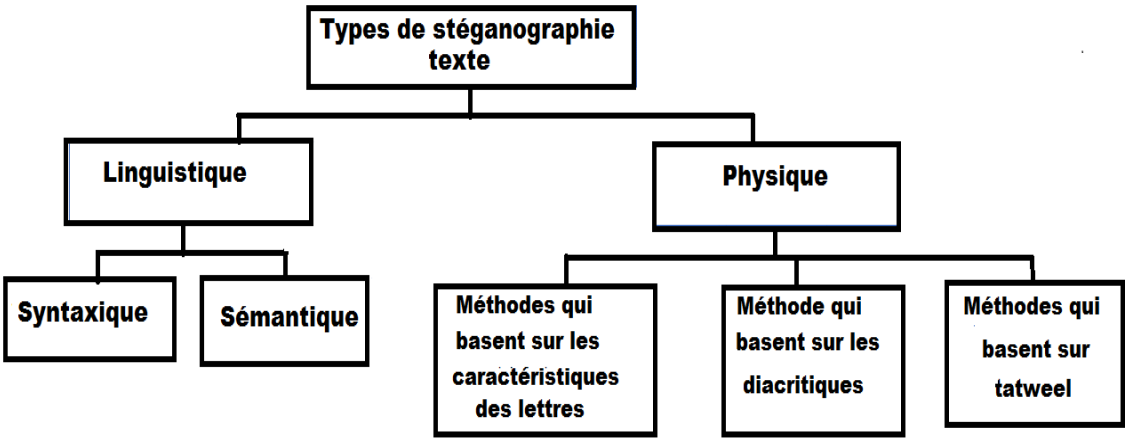


FIGURE 3.5 – Types de stéganographie texte

3.3.1 linguistique

les méthodes linguistique se basent sur la compréhension des composants principaux du texte dans une langue, nous pouvons distinguer deux types :

3.3.1.1 Syntaxique

Elle consiste à modifier la structure du texte pour insérer le message secret, exemples : méthodes de décalage des lignes et des mots[31], méthode des éspaces [32], les ligatures[33], une autre qui utilise l'horthographe des mots[34]...etc.

3.3.1.2 Exemples

Méthodes de décalage des lignes et des mots

Brassil et al proposent dans [31] de jouer sur la mise en forme d'un texte pour cacher l'information, en déplaçant les lettres des mots horizontalement. Le message secret est apparait seulement lorsque les deux textes sont superposés.

Ainsi, cette méthode n'attire pas l'attention du lecteur mais elle permet de cacher seulement 1 bit par mot. la figure 3.6 montre un exemple de cette méthode :

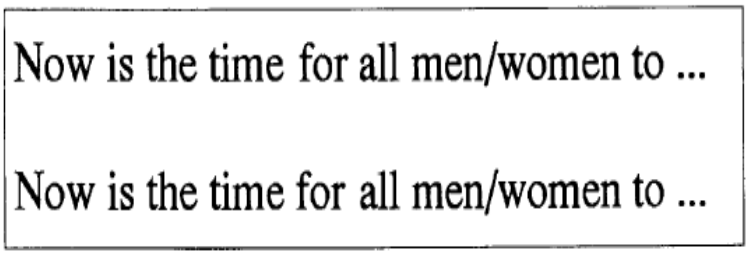


FIGURE 3.6 – Exemple de décalage des mots

Dans la figue 3.6 [31], les lettres des mots de la deuxiéme phrase sont déplacer un peu vers la droite. La superposition des deux phrases met en relief la méthode.

Dans le même article [31], une autre méthode qui consiste à déplacer les lignes verticalement vers le haut ou vers le bas pour cacher l'information. Comme le montre l'exemple ci-dessous :

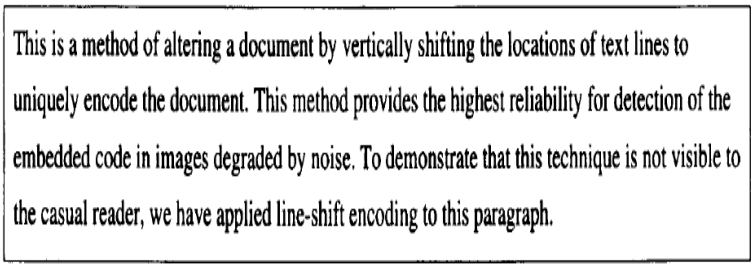


FIGURE 3.7 – Exemple de décalage des lignes

Dans la figure 3.7 [31], nous constatons que la première ligne est décalé un peu vers le haut, ainsi la troisième ligne.

Méthodes des éspaces

Le processus de dissimulation dans cette méthode [32] se fait par l'ajout d'espace supplémentaire dans le texte soit à la fin de chaque ligne, paragraphe ou entre les mots. Cette méthode peut être appliquée a tous texte de sorte qu'elle ne soulève pas l'attention de lecteur, la figure 3.8 montre un exemple de dissimulation dans un texte en ajoutant un espace entre les mots :

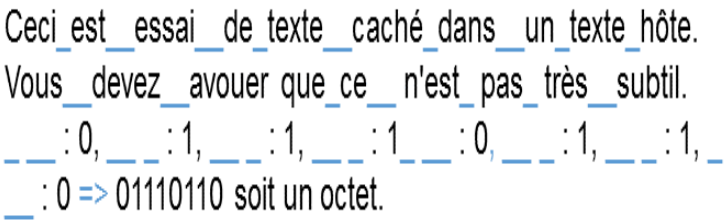


FIGURE 3.8 – Exemple de méthode des éspaces

Dans la figure 3.8 [32] , nous constatons que la méthode utilise un espace entre deux mots suivi de deux espaces, pour cacher le bit 0, et deux espaces suit d'un espace pour cacher le bit 1.

Malgré que cette méthode est facilement implémenté, il faut énormément de lignes pour cacher peu de bit et elle est super-visible par les personnes.

Méthode des ligatures

Certains groupes spéciaux de lettres peuvent être réunis pour créer un seul glyphe comme représenté dans la figure 3.10 . Dans cette méthode [33], L'algorithme de dissimulation cherche les ligatures disponibles dans le texte pour cacher 0 si non 1. Par exemple, si nous voulons cacher 1 nous écrivons fi on mettons un espace entre f et i. Si non, nous écrivons fi sans éspace entre f et i. Exemple :

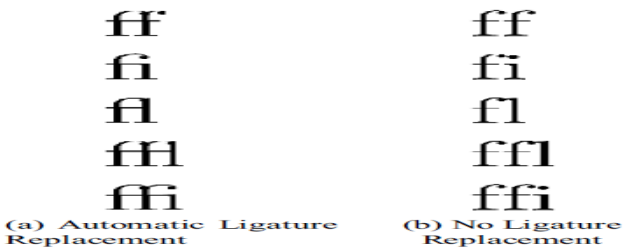


FIGURE 3.9 – Exemple des ligatures

La figure 3.10 [33], montre dans le côté gauche l'écriture de "ff" et "fi"...avec une ligatures. Et dans le côté droit montre l'ecriture des même mots sans ligatures.

Méthode du stéganographie texte en utilisant l'orthographe

Les chercheurs dans [34] présente une nouvelle méthode de stéganographie pour dissimuler des données dans les textes anglais.

Cette méthode se base sur le concept de substitution des orthographes des mots qui ont des orthographes différents au Royaume-Uni et aux États-Unis. Par exemple "dialogue" a des termes différents au Royaume-Uni (Dialog) et USA (Dialogue). Donc, nous pouvons cacher des données dans le texte en remplaçant ces mots. Le tableau 3.1 montre quelques mots qui ont des orthographes différents au Royaume-Uni et des États-Unis :

TABLE 3.1 – tableau des quelques mots au Royaume-Uni et États-Unis

American Spelling	British Spelling
Favorite	Favourite
Criticize	Criticise
Fulfill	Fulfil
Center	Centre

Le tableau 3.1 montre, dans le côté gauche quelques mots dans le dialecte américan, et les mêmes mot dans le dialecte de l'angleterre.

Méthode des émoticônes

En 2009 Wang et Al [35], proposent une méthode qui utilise les propriétés de chat, ils se basent sur l'utilisation des émoticônes pour cacher l'information secréte.

Plusieurs idées ont été proposées, l'une d'entre elle consiste à utiliser l'emplacement d'émoticône. Par exemple : si l'émoticône est insérer dans la troisième et la quatrième position dans le message, la donnée secrète est : **0011**.

le problème avec cette méthode qu'il faut un grand nombre d'émoticône pour cacher peu de bit.

3.3.1.3 Sémantique :

Dans les méthodes sémantiques les changement sont faites dans le contenu du texte pour insérer le message secret, elles ont été utilisées plus dans les textes anglais. Exemple : La méthode des synonymes[36], et une autre qui se base sur quelques mot clés dans la grammaire anglaise [37]...etc.

3.3.1.4 Exemple

Méthode des synonymes

Les auteurs dans[36] proposent un nouveau algorithme dans la langue Behase Melayu qui utilise les synonymes des mots.

Le processus de dissimulation consiste a créer la même liste de synonymes des mots dans le coté de l'émetteur et le récepteur pour pouvoir cacher et extraire le message secret.

Pour cacher 0, le mot dans le texte reste tel qu'il est. Mais si nous voulons cacher 1 le mot est remplacer par son synonyme dans la liste.

Méthode de dissimulation en utilisant la grammaire anglaise

Cet algorithme étudie la possibilité d'intégrer un message secret dans un texte en utilisant quelques règles du grammaire anglaise sans modifier le sens ou la syntaxe du document original.

En général, l'ordre des mots dans la langue anglaise est important mais certaine manipulation de quelques mots clés peuvent donner un sens similaire[37]. Par exemple la phrase : « Because she was tired, she went to sleep » En français : «Parce qu'elle est fatiguée, elle veut dormir» peut être écrite sous une autre forme, tout en préservant la structure et le sens : «She went to sleep, because she was tired » En français «Elle veut dormir, parcequ'elle est fatiguée »

Dans cet exemple, nous avons changé la première phrase en commençant par le mot clé -Because- avec la deuxième phrase, afin de cacher un peu d'information.

Autres techniques possible en utilisant quelques mots clés qui apparaissent fréquemment dans un texte anglais, des transformations sont faites sur la base de ces mots. Nous allons citer quelques exemples dans le tableau 3.2[37] :

TABLE 3.2 – tableau des méthodes utilisées dans la langue anglaise

Transformation Types	Examples of Transformations		
	Maneuverable Words/Clause	Sample Sentence	Bit
Splitting Separable Phrasal Verbs	Adverb particle of phrasal verb	**Take off**those bright yellow sandals. **Take**those bright yellow sandals**off**.	0 1
Swapping Operands of Boolean Operator	Operands around "**and**", "**or**", "**nor**"	He enjoys reading both **magazines** and **novels**. He enjoys reading both**novels** and **magazines**.	0 1
Re-arranging Conditional Statement	independent and dependent clauses	**If you pay using cash**, you will get a discount.. You will get a discount,**if you pay using cash**.	0 1
Adverb Placement	including modifiers of auxiliary verbs	**Also**, we can start on the research project tonight. We, **also**, can start on the research project tonight. We can **also** start on the research project tonight. We can start on the research project tonight, **also**.	00 01 10 11

Le tableau 3.2 montre quelques exemples des méthodes utilisées dans la langue anglaise, comme par exemple l'utilisation des phrases conditionnelle comme le montre la troisième ligne ou quelques mots clés comme les mots "Also", "And", "OR"...etc.

3.3.2 Physique

Cette approche apporte certaines modification physique dans le texte pour cacher le message secret, les modifications sont faites en ajoutant, supprimant des caractères...etc[38].

3.4 Méthodes dans le texte arabe

Pour organiser le travail, dans cette partie nous allons commencer dans un premier temps à classer les méthodes, après, nous tenterons d'expliquer les algorithmes proposés dans chaque type de méthode en respectant leurs ordres chronologiques.

3.4.1 Des méthodes qui utilisent les caractéristiques des lettres

La figure montre les différents méthodes de stéganographie texte qui utilisent les caractéristiques des lettres. Les auteurs dans [39] en **2006**, ont

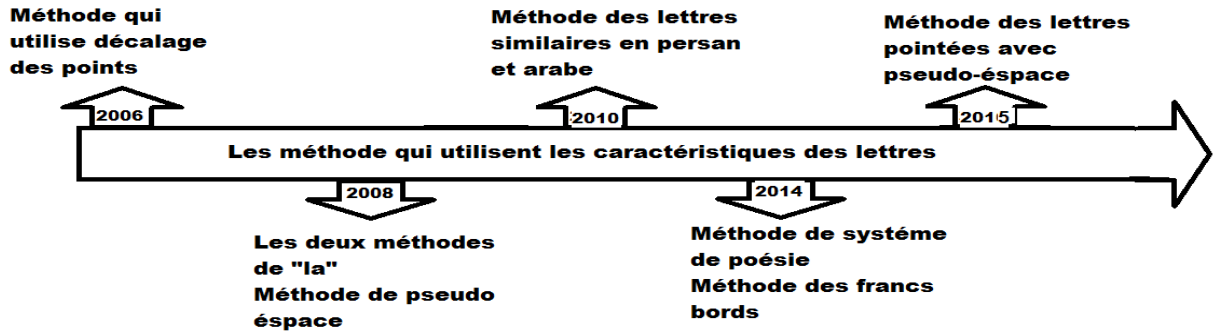


FIGURE 3.10 – Les méthodes des caractéristiques des lettres

proposés une approche qui dépend de l'existence des points sur la majorité des lettres de la langue **Arabe, Ardu et Persan**.

Ces points sont utilisées pour cacher les informations secrètes dans un texte, si le bit à cacher égale à 1 le point dans la lettre pointée est décalée un peu vers le haut, mais si le bit à cacher est 0, l'emplacement du point ne change pas. La figure 3.11 [39] montre un exemple de déplacement vertical du point pour la lettre « NOON » :

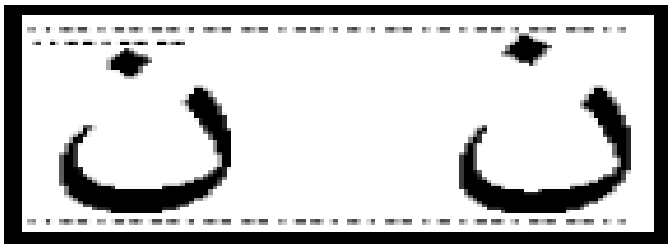


FIGURE 3.11 – La lettre NOON

Dans cette figure, nous remarquons que le points de la lettre "NOON" écrit à droite est décalé un peu vers le haut par rapport a sa position normale pour cacher le bit 1, tandis que le point dans l'autre "NOON" à gauche est utilisé pour cacher 0.

Avantages et inconvenients

Cette approche à une grande capacité dans la langue arabe où elle y a 15 lettres pointées parmi 28.

Toutefois, le processus de retaper va supprimer l'information cachée, Ainsi, cette méthode nécessite la création d'une police spéciale.

Une autre méthode de stéganographie dans le texte **Arabe/Persan** a été proposée en **2008** par Shirali-Shahreza et Mohammad [40], les chercheurs dans cette approche se sont basés sur le mot « La » . Ce mot est obtenu par la connexion de la lettre « Lam »avec « Alef » .

Le processus de dissimulation se fait sur la base de l'existence de deux formes de ce mot :

la forme spéciale sans extention entre "lam" et "alef"" qui est représenté par un seule Unicode et **la forme normale** en insérant une extension entre « Lam » et «Alef ».

Pour cacher le bit 0, la forme normale est utilisée, et la forme spéciale pour cacher le bit 1.

Avantages et inconvenients

Cette méthode a une faible capacité parce que l'utilisation du « la » est limitée dans le texte, Aussi l'utilisation de l'extension augmente la taille du fichier et donne une apparence anormale au texte.

Pour résoudre ce problème les chercheurs[41] en **2008** ont proposés une

amélioration. Ils ont utilisés les différents Unicode de « lam » et « «Alef » pour former le mot « la » dans la forme spéciale et normale se basant sur le fait que chaque lettre peut avoir quatre formes différentes qui dépend de sa position dans le mot. La figure 3.12 montre un exemple d'application de deux méthodes de "la" :

Message secret	1001
Texte de couverture	لا تحـزن لأن الحزن لا يردُّ قدراً ولا يجلب نفعاً
Méthode 1 de « La»	لا تحزن لـــان الحزن لـــا يرد قدرا و لا يجلب نفعا
Méthode 2 de « La»	لا تحزن لـان الحزن لـا يرد قدرا ولا يجلب نفعا
	⇩ ⇩ ⇩ ⇩ 1 0 0 1

FIGURE 3.12 – Exemple des deux méthodes de lam

La figure 3.12 montre un exemple des deux méthodes, dans la première méthode pour cacher le bit 1 la forme spéciale est utilisées comme montré dans la première et la dérnière "la" dans la phrase et pour cacher le bit 0 la forme normale est utilisée avec extention entre "lam" et "alif" comme le montre la deuxième et la troisième "la" dans la phrase. Tandis que dans la deuxième méthode la forme spéciale avec les différents Unicode de "la" est utilisée pour cacher 1, et la forme normale avec les différents Unicode de "la" pour cacher 0.

Avantages et inconvenients

le procédé amélioré ne change pas la taille du fichier, et donne un aspect normale au texte avec le même rapport de capacité faible.

En **2008** les chercheurs [42] ont proposés une autres approche. Le processus de dissimulation dans cette méthodes consiste à regarder d'abord si la lettre du mot est reliée à la lettre suivante ou non.

Si elle est connectée avec la lettre suivante, ils ont ajoutés un caractère de pseudo-connexion (Zero Width Joiner : ZWJ) pour cacher le bit 1, et ils n'ont rien ajouté pour cacher le bit 0 et comme les lettres sont reliées, l'ajout de ZWJ n'a pas d'effet sur le texte.

Mais si la lettre n'est pas connecté à la lettre suivante, ils ont ajoutés un caractère de pseudo-espace(Zero Width No Joiner : ZWNJ) entre les deux lettres pour cacher le bit 1 et ils n'ont rien ajouté pour cacher le bit 0. Dans

ce cas aussi l'ajout de ZWNJ n'a aucun effet sur l'apparence du mot dans le texte.

Pour cacher les données dans la dernière lettre d'un mot, ils insèrent toujours ZWNJ après, pour cacher le bit 1 et ils n'ont rien ajouté pour cacher le bit 0. comme le montre l'exemple présenté dans la figure 3.13 :

	Word	Unicode Representation
Input text	شام	0645,0627,0634
Data to hide	101	
Replaced characters	شام	FEE1,0627,FEB7
Adding required ZWJ	شام	FEE1,0627,200D,FEB7
Stego text	شام	FEE1,0627,200D,FEB7

FIGURE 3.13 – Exemple de la méthode de pseudo-espace

Dans la figure 3.13[42], chaque lettre est représenté par un numéro qui indique son Unicode dans sa position. Par exemple l'Unicode de la lettre 'shaa' dans sa position initiale est '0634', après nous allons remplacer chaque caractère qui cache le bit 1 par son Unicode dans une autre position, et pour indiquer qu'il y a un bit caché, nous ajoutons un caractère de pseudo espace , dans ce cas un zwj qui est représenté par '200D' parce que les lettres sont connectées et nous voulons cacher le bit 1.

Avantages et inconvenients

Cette technique a une faible capacité mais elle est indetéctable par l'oeil humain.

Une autre méthode de stéganographie pour les texte **Persan** et **Arabe** est présenté dans [43] en **2010**, les chercheures se sont basé sur l'existence de deux lettres similaires "Ya" et "Kaf" avec des different Unicode dans la langue arabe et persan. Les auteurs proposent d'utiliser les caractère Persans pour cacher le bit 0 et les caractères arabes pour cacher le bit 1.

Avantages et inconvenients

Cette méthode a l'avantage de haute imperceptibilité puisque aucun changement ne se produit en apparence du texte, mais elle a une faible capacité sur la base d'utilisation de ces deux lettres dans le texte et aussi les caractéristiques de la langue ne sont pas respecté .

En **2011**,Dans [44], les francs bords des lettres arabes ont été utilisés pour cacher les bits secrets. La figure 3.14 [44]montre un exemple des francs bords dans lettres arabes :

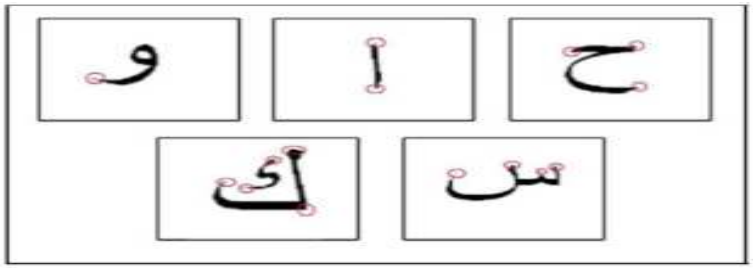


FIGURE 3.14 – Exemple des francs bords

L'algorithme utilise deux clés. La première clé détermine les lettres pointées et les lettres non pointées qui seront utilisées pour cacher les bits secrets. Si la somme des chiffres de la clé est pair, alors les lettres pointées sont utilisées, et les lettres non pointée sont utilisés autrement.

La seconde clé est utilisé pour déterminer les positions des bits secrets dans les lettres du texte de couverture. Le nombre des francs bords montre la possibilité de cacher les bits 1 et 0. Dans les caractères dont le nombre des francs bords est 1, elle est possible de cacher le bit secret 0 ou 1. Par ailleurs, si le nombre des francs bords est deux, la position secrète de bits est en quatre possibilité : 11, 10, 00 ou 01.

Avantages et inconveninets

Cette méthode a une grande capacité, avec une grande invisibilité, ainsi elle résoue le problème de retaper. Le positionnement aléatoire des bits secrets offre une grande sécurité pour le texte.

En **2014** ESAM ALI KHAN[45] a proposé d'utiliser le système de la poésie arabe afin de cacher les bits secrets, puisqu'il a une représentation binaire intégré dans chaque poème arabe, cela peut être utilisé comme base pour cacher les bits secrets. Comme le montre la figure3.15 [45] :

Dans cette figure, le versé du poème est écrit premièrement comme ça ce prononce, après il est représenté en binaire.

L'idée est de présupposer que le bit secret est incorporé dans une position des bits binaire du poème, le bit secret est alors identique à cette position ou son inverse.

The verse	وألاّ أرى غيري له الدهرَ مالكا			ولي وطنٌ آليت ألا أبيعَهُ			
How it is pronounced	رمالكا	لهدده	وألا أرى غيري	أبيعهو	ت ألا	ولي و طنن ءالي	
The corresponding feet	مفاعلن	فعولن	مفاعيلن	فعولن	مفاعلن	فعولن	مفاعيلن	فعول
Binary representation	011011	01011	0101011	01011	011011	01011	0101011	1011
Its classification	Al-Taweel meter							

FIGURE 3.15 – Exemple de poeme arabe

Afin de faire la distinction entre les deux, le caractére de tatweel est ajouté lorsque le bit secret est opposé de bit binaire de la lettre correspondante du poème.

Avantages et inconvenients

Cette méthode a une grande capacité, mais elle n'est pas robuste contre le processus de retaper, et elle a une application limité.

Aussi, en **2014** une autre méthode proposée dans [46] , les chercheurs proposent d'utiliser les trois entités principales : les francs bords, les points, et la typo-proportion.

Le typo proportion a été adopté à partir de la calligraphie. La figure 3.16[46] illustre des exemples de proportion pour les caractères arabe :

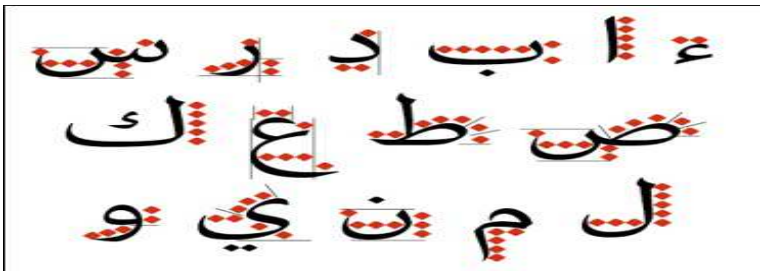


FIGURE 3.16 – Exemples de proportion

Dans cette figure 3.16, les points montre la façon d'ecrire les lettres en utilisant l'ecriture de "EL- Naskh".

La figure 3.17[45] montre des structures primitives de certains caractères Arabes :

Char	Unicode	Primitive Structural Method Entities			Num. Of Potential Hiding Place
		Sharp-Edges	Dot (s)	Typo-graphical Proportion	
ا	\U0627	2	0	1	3
ب	\U0628	2	1	3	6
ت	\U062A	2	2	3	7
ث	\U062B	2	3	3	8
ج	\U062C	3	1	5	9

FIGURE 3.17 – Exemples des structures de quelques lettres

D'après la figure3.17, chaque caractère a plus d'un endroit pour cacher les bits secrets. Par exemple pour un message secret de 16 bits de long (exemple : 10.100.101 10.100.101), nous prenons un caractère du texte de couverture exemple "ba". Le potentielle totale de caractère "ba" est 6.

Nous prenons 6 bits à partir des bits secrets de droite à gauche, exemple : **100101**.

L'algorithme va répéter le processus pour le prochain caractère du texte de couverture et les prochains bits secrets jusqu'à ce que le message secret soit caché dans le texte de couverture.

Avantages et inconvenients

Cette méthode offre plusieurs positions possible pour cacher les bits secrets, elle est sécurisée et offre une grande capacité, parcequ'elle permet de cacher plusieurs bits dans un seul caractére.

En **2015**, les auteurs[38] ont proposés une nouvelle méthode qui utilise les lettres pointées avec le pseudo-espace, Le processus de dissimulation consiste premièrement à vérifier la dernière lettre de chaque mot, si la lettre est pointées et le bit à cacher est 1 alors un pseudo-espace est ajouté entre les mots, si non ils ont rien ajoutés mais si la dernière lettre est non pointée et le bit qu'on veut le cacher égale 0 un pseudo-espace est ajouté, mais si le bit qu'on veut le cacher est 1, ils ont rien ajoutés La figure 3.18 [38]montre un exemple :

Dans la figure le premier bit à cacher est 1, et la dernière lettre est non-pointée alors rien est ajouté, mais dans l'avant dernier mot le bit à cacher est 0 et la dernière lettre est pointée alors un pseudo-éspace est ajouté entre les deux mots.

Watermarking bits	1001011
Original text	البر حسن الخلق والإثم ما حاك في نفسك
Output text	البر حسن الخلق والإثم ما حاك في نفسك 1 1 0 1 0 0 1

FIGURE 3.18 – Exemple de la méthode des lettres pointées avec le pseudo-espace

3.4.2 Méthodes qui se base sur les diacritiques :

La figure montre les différents méthodes de stéganographie qui utilisent les diacritiques.

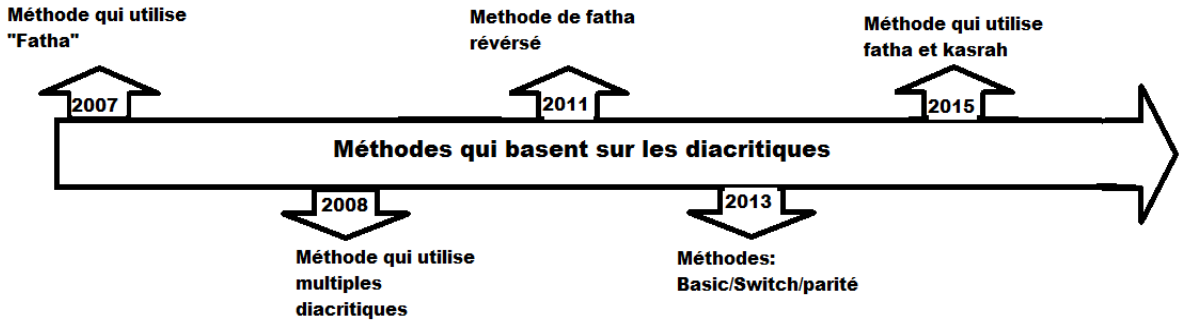


FIGURE 3.19 – Méthodes qui utilisent les diacritiques

En **2007** les auteurs dans [29] ont proposés une technique intéressante où le diacritique **'Fatha'** est utilisé pour cacher 1 tandis que le reste des signes diacritiques sont utilisés pour cacher 0. Ils se sont basé sur le fait que **'Fatha'** représente près de la moitié des signes diacritiques dans un texte arabe. La figure 3.24 [29] montre un exemple :

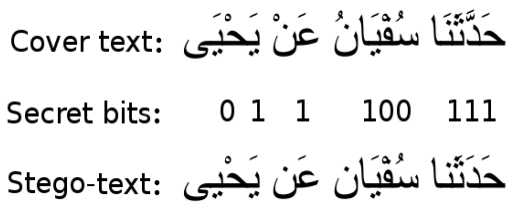


FIGURE 3.20 – Exemple de la méthode de fatha

Dans la figure3.24, les bits sont cachés dans les lettres qui porte le diacritique 'fatha' seulement, et si le bit à cacher est 1 le diacritique 'fatha' reste tel qu'il est, et si le bit est 0 le 'fatha' est supprimé.

Avantages et inconvenients

Cette méthode offre la possibilité de cacher dans chaque caractére un bit, mais elle base sur les diacritiques qui sont raremant utilisés.

les auteurs en **2008**, suggère une méthode [47] qui utilise plusieurs signes diacritiques pour cacher plusieurs bits.

L'idée se base sur le fait que les éditeurs du texte afficherons plusieurs signes diacritiques successifs sur le dessus de l'autre de sorte qu'ils apparaissent comme un seul signe.

Cette méthode a deux approches texte et image avec différents scénarios. L'un des scénario était de prendre n bits secret et l'envisager comme un nombre binaire et puis le convertir en un nombre x décimale, puis le cacher dans un signe diacritique en ajoutant x copie de ce diacritique.

Avantages et inconvenients

Si la capacité est très élevée, la fiabilité de l'impression est faible et elle donne un invisible tatouage. En plus, cette méthode ne fonctionne pas sur certains éditeur (gedit 3.4.1 sur Ubuntu version 12·0·4) [48].

Aussi, elle repose sur des caractéristique qui ne font pas partie la langue mais liées aux éditeurs du texte.

Une autre technique en **2011** dans [49], les chercheures ont utilisés la réversion de diacritique. L'étude renverse la manière originale de 'Fatha' à partir d'une petite ligne inclinant à gauche au-dessus de lettre vers la droite. La régulière fatha est utilisé pour cacher le bit 1 ,et fatha inverse utilisée pour cacher le bit 0.

Avantages et inconvenients

Cette méthode peut être detecter par les programme d'OCR, et les changements sont visible dans le texte. En plus le caractère « fatha inversé » ne fait pas partie des diacritiques de la langue.

La principale contribution en **2013**, les trois méthodes proposées dans [48] qui base sur l'utilisation d'un texte complètement diacritiqué.

Dans la première méthode si le bit à caché est 1 alors le diacritique reste tel qu'il est, mais si le bit à cacher est 0 le diacritique est retiré du texte, Comme représente dans la figure 3.21 [48] :

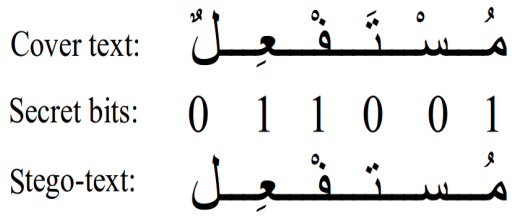


FIGURE 3.21 – Exemple de la méthode basique

Dans le mot representé dans la figure 3.21, nous constatons que dans la première, quatriéme et cinquiéme lettre le diacritique reste tel qu'il est parce que le bit à cacher est 1. Tandis que le diacritique est supprimer dans les autres lettres oú le bit à cacher est 0.

Dans la deuxième méthode ou « switch méthode »le diacritique est apparu dans le texte seulement si nous avons un switch de 0 à 1 ou de 1 à 0, comme le montre la figure 3.22 :

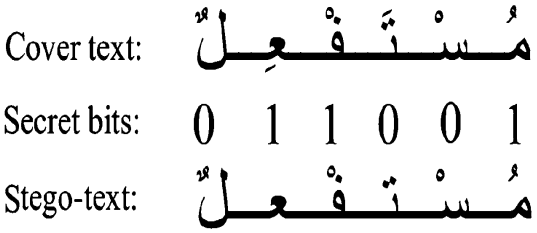


FIGURE 3.22 – Exemple de la méthode de switch

Dans la figure 3.22 [48]qui représente la méthode de swich, nous constatons

que le diacritique dans la deuxiéme lettre reste tel qu'il est parce que nous avons un switch dans les bits secret, mais il est éliminé dans la troisiéme lettre parce que les deux bits sont similaire.

La dernière technique se base sur le bit de parité, un bit de parité est affecté pour chaque lettre dans le texte de couverture, si la position de la lettre dans le texte est pair, le bit de parité de cette lettre est un 0, si non un 1.

Après, nous mettons séquentiellement chaque bit secret au dessous de la lettre du text de couverture, si le bit secret est différent du bit de parité, nous montrons le diacritique et de la cacher s'ils sont égaux, la figure3.23montre un exemple :

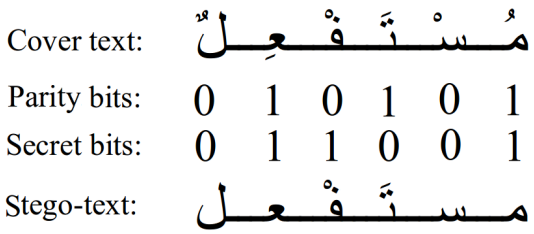


FIGURE 3.23 – Exemple de la méthode de parité

Dans la figure 3.23 [48], la première lettre a une bit de parité égale à 0, et le bit à cacher est 1, alors le diacritique reste tel qu'il est dans la lettre, mais dans la cinquième lettre, le bit de parité et le bit à cacher sont similaire, alors le diacritique est supprimé.

Avantages et inconvenients
Dans ces méthodes, le nombre de bit à cacher est environ 2 fois le nombre des signes diacritique indiqué. En effet, les signes diacritiques supprimés également cachent les bits, qui signifie que la capacité des méthodes est deux fois la première méthode[29], mais elles ont une faible robustesse.

L'algorithme proposé en **2015** dans [50] utilise les deux diacratique "fatha" et "kasrah". Les chercheurs se sont basés sur le processus suivant :

Ils ont commencé par diviser le message secret en deux tableaux pair et impair, aprés ils ont comparé la valeur du tableau pair avec le diacritique **Fatha**, si le bit à cacher est 1 alors garder le fatha dans le texte de couverture, si non retirer le fatha du texte. Aprés en incrémente l'indice du tableau, et nous répétons les mêmes étapes jusqu'a la fin du tableau.

Ensuite, nous utilisons les mêmes étapes précédentes avec le tableau pair en utilisant le diacritique **Kasrah**.

3.4.3 Méthodes qui se base sur Tatweel :

La figure montre les différents méthode de tatweel utilisées dans le texte arabe Cette catégorie dépend de l'existance du caractère d'extension en langue

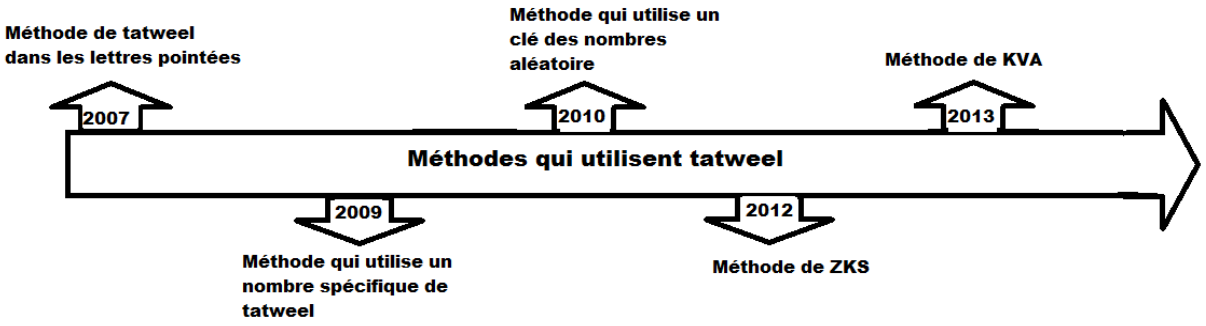


FIGURE 3.24 – Les méthodes de tatweel

arabe "Tatweel". Elle est utilisée dans la stéganographie texte pour contenir les bits secrets dans différents algorithmes.

Afin d'améliorer l'aspect négatif de robustesse dans la méthode proposée par Shirali-Shahreza dans les lettres pointées[39], les auteurs dans [47] en **2007** ont proposés une nouvelle méthode pour cacher l'information dans les lettres au lieu d'utiliser les points seulement, ils ont proposés d'utiliser les lettres pointées avec extension pour cacher le bit 1 et les lettres non pointées avec extension pour cacher le bit 0.

Notons que l'existence de l'extension dans la lettre n'a aucun effet sur le contenu du texte. En effet, ce caractère d'extension dans la langue arabe est utilisé seulement pour arranger le texte. Cette méthode peut avoir l'option d'ajouter l'extention avant ou après les lettres. Comme le montre l'exemple de la figure 3.25

Secret bits	110010
Cover-text	من حسن اسلام المر ء تركه مالا يعنيه
Steganographic text	من حسن اسلام المر ء تـركـه مـالا يـعنـيه 1 1 0 0 1 0

FIGURE 3.25 – Exemple de la methode d'extention

Dans la figure 3.25 [47], nous constatons que les lettres pointées avec extention tel que "ta" et "ya" sont utilisées avec extention pour cacher le bit 1. Tandis que les autres sont utilisés pour cacher le bit 0.

Avantages et inconveniet :

La sécurité est faible dans cette méthode parce qu'un mot peut avoir plusieurs extensions représentant les bits secrets qui reflètent l'attention d'existence d'une information cachée.

Pour cette raison 3.4.3, les chercheurs en **2009** dans [51] ont proposés une autre méthode qui réduit le nombre de Tatweel dans un mots en insérant un nombre spécifique de Tatweel dans les mots qui ont des caractères extensible et essayé d'utiliser seulement l'emplacement de Tatweel possible au sein d'un mot donné dans le texte pour cacher le bit secret. L'existence d'une information caché dans le mot est représenté par insertion de Tatweel après quelques caractères extensible dans le mot.

Pour appliquer cette méthode plusieurs scénarios ont été utilisés, l'un d'eux se base sur le calcul du nombre maximum de Tatweel possible a inséré dans un mot extensible : un, deux, trois. La figure 3.26 montre un exemple d'application de la méthode :

Bit secret	001010
Texte	سنمتعهم
Méthode proposée	سنمتـعـهم

FIGURE 3.26 – Exemple de la méthode qui réduit le nombre de Tatweel

Avantages et inconvenients

Cette approche donne une capacité meilleure et une plus grande sécurité que la méthode précédente, mais elle est non-robuste contre le processus de retaper.

En **2010** dans [52], l'idée proposée était d'utiliser une clé secrète pour générer des nombres aléatoires qui représente la position de Tatweel a ajoutée aux mots de texte électronique pour cacher les bits de message secret. La figure 3.27 [52] montre un exemple d'application

Texte de couverture	متى استعبدتم الناس و قد ولدتهم أمهاتهم أحرارا
Nombres aléatoire	1 1 2 7 0 2 7
Stégo-texte	متـى اسـتـعبدتم النـاسوقد ولـدتــهم أمـهاتهم أحــرارا

FIGURE 3.27 – Exemple de méthodes d'extention qui utilise des nombres aléatoires

Dans le deuxième mot montré dans la figure 3.27, le nombre aléatoire générer est deux, alors l'extention est ajouté seulement dans les deux lettres qui accepte l'extention dans le mot et la même chose pour le reste.

Avantages et inconvenients

Le procédé proposé fait la tâche d'un attaquant beaucoup plus difficile par rapport aux méthodes précédentes, Mais il offre moins de capacité que les méthodes basées sur les diacritiques.

les chercheurs dans [53] en **2012** ont développés un nouveau algorithme pour les textes arabes basés sur le « Tatweel » et « zero width » parce que ces deux derniers ne fontt aucun changement sur le sens du texte.

Le nouveau algorithme propose d'utilise « zero width » et « kashida« (ZKS) pour cacher 2 bit par chaque caractère, comme le montre la figure 3.28 [53] :

Extension	Zero Width	Code	Letter effect
No	No	00	No EFFECT
Yes	No	01	Extension
No	Yes	10	Zero width
Yes	Yes	11	Extension + Width

FIGURE 3.28 – Algorithme de ZKS

Avantages et inconvenients

Cette méthode augmente la capacité des bits à cacher, mais l'utilisation de tatweel augmente la taille de fichier et attire l'attention.

En **2013** les chercheurs dans [54] ont proposés une méthode d'insertion du tatweel après certaines lettres seulement pour cacher 1 et ils n'ont rien ajouter

pour cacher le bit 0. Comme le montre la figure 3.29 :

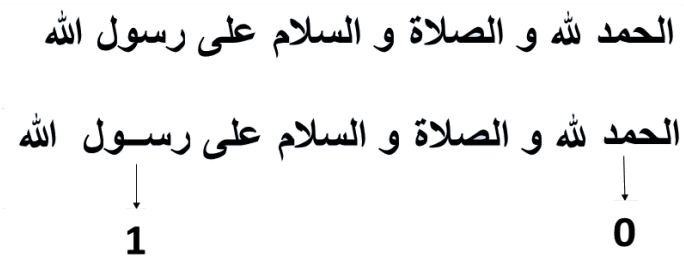


FIGURE 3.29 – Exemple de la methode de tatweel dans certains caractéres

Avantages et inconvenients

La capacité de cette méthode dépend de l'existance des lettres précédente dans le texte.

Un algorithme de variation de kashida (KVA) a était proposé en **2013** par [55] pour augmenter la robustesse. Premierement, le texte est divisé en bloc, puis appliquer au hasard l'un des quatre scénarios suivant pour cacher les bits secrets.

1. Ajouter kashida après une lettre pointue pour cacher 1, et ne rien ajouter pour cacher 0.
2. Ajouter kashida après une lettre non-pointue pour cacher 1, et ne rien ajouter pour cacher 0.
3. Ajouter Kashida après une lettre pour cacher 1, et ne rien ajouter pour cacher 0.
4. Ajouter Kashida après une lettre pour cacher 0, et ne rien ajouter pour cacher 1.

La figure 3.30 [55] montre un exemple de cette application :

3.5 Comparaison des méthodes de stéganographie

Pour comparer les méthodes de stéganographie proposées dans le texte arabe, nous allons les comparer selon les trois critères les plus importantes : capacités, robustesse et sécurité.

Applied Scenario	Hidden Bits	Stego Object
Scenario1	0100	اتق شر من احسنت اليه
Scenario2	1011	اتق شر من احسنت اليه
Scenario3	11000100	اتق شر من احسنت اليه
Scenario4	001011100	اتق شر من احسنت اليه

FIGURE 3.30 – Exemple de l'algorithme KVA

3.5.1 Capacité :

La capacité est la quantité des bits qui peuvent être cachés dans le texte de couverture. Elle est exprimée en pourcentage, par exemple, une capacité de 5% signifie que dans 100 bits, 5 bits sont cachés, ou peut être exprimé également en bits / KB, où, par exemple une capacité de 300 bits/KB veut dire que dans chaque kilobit 300 bit sont cachés[48].

3.5.1.1 Comparaison entre les méthodes qui utilisent les caractéristiques des lettres

SHIRALI-SHAHREZA a proposé plusieurs méthodes du stéganographie dans les textes arabes avec des capacités différentes. Après le teste sur un corpus de 10 différentes pages du sport d'un journal Iranien :

Sa première méthode proposée en 2006 [39] qui se base sur l'utilisation des points pour cacher les bits secrets offre une capacité de 1,3 % qui veut dire que nous pouvons cacher 112,8 bit dans chaque kilobit (112,8 bit/KB).

Les deux autres méthodes proposées en 2008[40, 41] de « La avec extension » et « La avec les différents Unicode de lam » offre une capacité moins que la première qui vaut 0,012 % c'est à dire 1,17 bit/KB.

L'autre méthode proposée par Shirali en 2008 qui utilise les lettres similaires en persan et arabe avec différentes Unicode [43] donne une capacité de 0,40% (33,25 bit/KB), aprés le test sur 8 pages du journal précédent.

La méthode de pseudo-espace[42] a été testée sur un corpus de 2 pages du journal précédent, elle peut cacher 4,18 bit dans chaque kilobit (0,05%).

En 2014 une nouvelle méthode qui se base sur le système de la poésie arabe proposée par Esam Ali Khan[45] donne une capacité meilleures que les

précédentes qui peut arriver jusqu'à 8% ou (655,48 bit/KB).

Jusqu'à maitenant les méthodes qui utilisent les francs-bord offrent les meilleures capacités qui peuvent arriver jusqu'à 750 bits/KB, et une capacité théorique qui vaut 62,5% (100%*5/8)

Aprés cette étude nous pouvons conclure que dans les méthodes qui utilisent les caractéristiques des lettres, les meilleures capacités sont obtenus par les méthodes qui utilisent les francs-bords[44, 56], après par la méthode qui utilise le système de la poésie arabe [45], la méthode proposée par [39] qui utilise les point, offre aussi une bonne capacité mais elle est moins que les précédentes. Ensuite les méthodes proposées dans [43] qui utilise les lettres similaire, la méthodes de pseudo espace [42] et enfin la capacité minimale est obtenue par les deux méthodes de "la"[40, 41].

3.5.1.2 Comparaison des capacités des méthodes qui utilisent les diacritiques

La méthode proposée par Abad en 2007 [29] qui utilise les 08 différents diacritiques offre une capacité théorique qui peut arriver à 6.25 % (Nous trouvons que le diacritique "fatha" forme 50% des diacrtiques du texte, et chaque diacritique prend 8 bit et nous pouvons cacher 1 bit dans le diacritique alors la capacité théorique est 50 % * 1/8), tandis que la capacité pratique après le test sur une partie de Musnad Al-Emam Ahmed qui contient 7305490 caractères et 1037265 mots est 3,28 % (269 bit/KB) qui est la moitié de la capacité théorique.

Une autre capacité très élevée est obtenue par les trois méthodes proposées par Mr.bensaad en 2013 [48] et testées sur des fichiers textes des tailles différentes :

la méthode basique donne une capacité de 6,44% (528b/KB), la méthode de parité offre une capacité de 6,33% (518 bit/KB) et 6,45 % (528 bit/KB) est obtenue par la méthode du switch.

la méthode des multiples diacritiques proposée par Gutub en 2008[47] offre aussi une capacité très élevée parce qu'elle permet de cacher plusieurs bits dans une seule caractère.

la méthode proposée par Esraa Mohammad Ahmadoh et Gutub en 2015 qui utilise « Kasrah » et « Fatha » pour le processus de dissimulation[50] offre une capacité plus grande que la premiére [29], parce que si le diacritique "kasrah" forme 20 % des diacritiques de texte, la capacité théorique obtenue est presque 8,75% (20%*1/8+6.25)

Après cette comparaison, nous pouvons conclure que la meilleure capacité est obtenu par la méthode proposée dans [47] parcequ'elle permet de cacher plusieurs bits dans un seule caractére. Après par la méthode proposée en 2015 [50]parceque cette derniére utilise 2 diacritiques ce qui augmente la capacité. Ensuite les 03 méthodes proposées en 2013[48] et enfin la méthode proposée dans [29] qui utilise "fatha".

3.5.1.3 Comparaison des capacités des méthodes qui utilisent Tatweel

Dans la méthode d'extension des lettres avec les points proposée par Adnan Gutub en 2007[57], la capacité obtenue après le test sur un livre de l'arabe contemporaine (842684 mots de 415 diverse des textes de sites Internet.) est 1,22%(100,32 bit/KB).

L'autre méthode en 2009 [51] qui propose une restriction des nombres d'extensions pour ne pas attirer l'attention offre une capacité de 3,16 % (260 bit/KB) après les essais sur 04 textes différents.

Tandis que la méthode qui utilise une clé pour générer l'emplacement des extensions proposées dans [52] en 2010 a été testée sur des textes électroniques avec des tailles différentes offre une capacité de 2,8% (229 bit/KB).

La plus petite capacité est obtenue par la méthode d'insertion d'extension après certains caractères seulement proposée par Gutub [54] est environ 0,0008% (16,8 b/KB) après le test sur 4 textes de taille différente.

Nous pouvons dire que la meilleure capacité est obtenu par la méthode qui propose une restriction des nombres d'extensions[51], après par la méthode qui utilise une clé pour générer l'emplacement des extensions[52], ensuite la méthode d'extension des lettres pointées[57] et enfin la minimale capacité est obtenue par la méthode qui propose d'inserer l'extension après certains caractères seulement[54].

3.5.2 Robustesse

La Robustesse est la possibilité de résister de modifier ou détruire les données cachées. Une méthode est robuste si l'extraction des données cachées est difficile, pour dire qu'une méthode est robuste, il faut la tester contre :

— Les programmes d'OCR.
— La fonction du copier et coller.
— Le processus de retaper du caractère.

— Le changement de police.

3.5.3 Les méthodes qui utlisent les caractéristiques des lettres

La méthode de dissimulation par le décalage verticale des point dans les lettres pointées[39] est robuste seulement contre le processus d'impression. Mais les programmes d' OCR peuvent détecter ce décalage et le processus de retaper du caractére peut supprimer l'information cachée.

les deux méthodes de « la » [40, 41] sont aussi facilement détectées par OCR parce que les caractéristiques de la langue ne sont pas respectées, mais elles sont robustes contre les autres méthodes.

La méthode des lettres similaires [43] ne respecte pas les caractéristique de la langue, alors, elle est non robuste contre les programmes d' OCR et le processus de retaper de caractère, mais elle est robuste contre la fonction du copier-coller et elle nécessite pas une police spéciale.

La méthode des pseudo-espace[42] est aussi facilemet détecté par l'OCR et l'information cachée pouvait être supprimer par le retape du caractère, mais elle est robuste contre le copier-coller et elle nécessite pas l'ajout d'une police.

la méthode qui utilise le systèmes de la poésie Arabe[45] utilise l'extention, alors elle est non robuste contre le retape du caractère.

les méthodes des francs-bords sont robustes contre les processus précédents, parce-qu'elles respectent les caractéristique de la langue et elles ne necessitent pas un changement.

Alors, nous pouvons conclure que la meilleure robustesse est celle obtenue par les méthodes des francs-bords.

3.5.3.1 Les méthodes de Tatweel

Toutes les méthodes qui utilisent Tatweel, elles sont robustes contre le processus du copier coller et elles nécessitent pas des polices spéciale, nous pouvons insérer tatweel seulement avec (CTRL+J), mais elles sont facilement détécté par les programmes d'OCR et supprimé avec le retape du caractére.

3.5.3.2 Les méthodes des diacritiques

Les techniques des diacritiques offre une bonne robustesse face aux programmes d'OCR et ne pose aucun problème avec la fonction du copier coller mais certaines méthodes nécessitent une police spéciale comme « fatha reversé ».

3.5.3.3 Sécurité :

La sécurité fait référence à la possibilité qu'une personne peut détecter la présence d'un message caché dans le texte de couverture.

Certaines méthodes des codage sont un peu visible par l'être humain comme la méthode du décalage du points et autre sont visibles comme la méthode du « la » et certaines autres ne peuvent pas être détecter par l'œil humain comme la méthode des lettres similaires, pseudo-éspace et les francs-bords et le systéme de poésie Arabe.

Tandis que, les méthodes qui utilisent tatweel sont visibles par l'être humain.

les méthodes des diacritiques aussi sont remarquables parce qu' aujourd'hui peu de texte qui sont diacritiqués.

3.5.4 Résultats des comparaisons

Après la comparaisons des méthodes qui utilisent : les caractéristiques des lettres, les diacritiques, le Tatweel selon leurs : capacités, robustesses et sécurités nous pouvons conclure que les meilleures méthodes sont ceux qui utilisent **les francs-bords** des lettres[44, 56].

Les tableaux (3.3, 3.4, 3.5) résume les méthodes.

3.6 Conclusion

Malgré que l'utilisation du stéganographie est très pratique, son utilisation informatique sur le texte est détectable par certains programmes. Tandis que la sécurité de la stéganographie repose sur le fait que le message ne sera sans doute pas détecté ou détruit. Dans le chapitre suivant, nous allons essayer de proposer de nouvelles méthodes de stéganographie dans le texte arabe qui se base sur la sémantique de la langue arabe.

TABLE 3.3 – Méthodes qui utilisent les caractéristiques des lettres

Méthode	Corpus	Capacité		Robustesse				Securité
		%	bit/KB	OCR	Copier & coller	refrappe	Changement de police	
Methodes qui base sur les caractéristiques des lettres								
Méthode des points [39]	10 différentes pages du sport d'un journal Iranien	1,37	112,8	✗	✗	✗	✗	Un peu sécurisé
Méthode de "La" avec extention [40]	10 différentes pages du sport d'un journal Iranien	0,012	1,17	✗	✓	✗	✓	Non sécurisé
Méthode de "La" avec différent Unicode [41]	10 différentes pages du sport d'un journal Iranien	0,012	1,17	✗	✓	✗	✓	Non sécurisé
Méthode de deux lettres similairesavec différent Unicode [43]	8 différentes pages du sport d'un journal Iranien	0,40	33,25	✗	✓	✗	✓	sécurisé
Méthode de pseudo-espace [42]	2 différentes pages du sport d'un journal Iranien	0,05	4,18	✗	✓	✗	✓	sécurisé
Méthode qui base sur le système de la poésie arabe [45]	3 poémes	8	655,48	✓	✓	✗	✓	sécurisé
Méthodes des francs-bords[56]	non spécifié	élevé	élevé	✓	✓	✓	✓	sécurisé

TABLE 3.4 – Méthodes qui utilisent les diacritiques

Méthode	Corpus	Capacité		Robustesse				Securité
		%	bit/KB	OCR	Copier & coller	refrappe	Changement de police	
Methodes basées sur diacritique								
Méthode des 8 différents diacritiques [29]	une partie de Musnad Al-Emam Ahmed qui contient 7305490 caractères et 1037265 mots	3,28	269	✓	✓	✓	✓	remarquable
Méthode des multiples diacritiques[47]	non spécifié	élevé	élevé	non spécifié	✓	✗	✓	invisible
Méthode de reverse fatha[49]	non spécifié	non spécifié	non spécifié	✓	✓	✗	✗	remarquable
Méthode basique qui utilise les diacritiques [48]	des livres qui sont presque entièrement discrétisées	6,44	528	✓	✓	✓	✓	remarquable
Méthode de switch qui utilise les diacritiques [48]	des livres qui sont presque entièrement discrétisées	6,45	528	✓	✓	✓	✓	remarquable
Méthode de parité qui utilise les diacritiques [48]	des livres qui sont presque entièrement discrétisées	6,33	518	✓	✓	✓	✓	remarquable
Méthode qui utilise deux diacratiques [50]	la poésie arabe Mu'allaqât	71.45	non spécifié	✓	✓	✗	✓	remarquable

TABLE 3.5 – Méthodes qui utilisent Tatweel

Méthode	Corpus	Capacité		Robustesse				Securité
		%	bit/KB	OCR	Copier & coller	refrappe	Changement de police	
Methodes basées sur Tatweel								
Méthode d'extension des lettres avec les points [57]	Le corpus de l'arabe contemporaine (842684 mots de 415 diverse des textes de sites Internet.)	1.22	100.33	✗	✓	✗	✓	remarquable
Méthode de restriction de nombre de tatweel[51]	4 textes différents	3.16	260	✓	✗	✓	✗	remarquable
Méthode qui utilise une clé pour générer tatweel aleatoire[52]	textes éléctroniques	2.8	299	✓	✗	✓	✗	remarquable
Méthode de ZKS[53]	non spécifié	non spécifié	non spécifié	non spécifié	non spécifié	non spécifié	non spécifié	remarquable
Méthode d'insertion de kashida aprés des lettres spécifique [54]	4 differents documents	0.0002	16.88	✓	✗	✓	✗	remarquable
Méthode d'insertion de kashida aprés des lettres spécifique [55]	4 sites differents	35,5	non spécifié	✓	✗	✓	✗	remarquable

Chapitre 4

Les méthodes proposées pour la dissimulation d'information dans le texte arabe

4.1 Introduction

La langue est un moyen de communication entre les humains, les sociétés utilisent la langue pour communiquer. La parole est une des méthodes les plus importantes de la communication, nous constatons que les linguistes se souciaient beaucoup des structures et des composants de la phrase ainsi que des règles de base qui peuvent être appliquées sur elle.

Dans ce chapitre, nous allons essayer de montrer les caractéristiques les plus importantes de la langue et certaines de ses bases qui peuvent être appliquées sur la phrase[58, 59, 60, 61, 62, 63], pour tenter d'en tirer les conclusions afin de dissimuler le message secret sans changer le sens(section 4.2).

Dans la section 4.3, nous proposerons plusieurs façons qui vont nous permettre d'exploiter plusieurs phrases dans le texte afin de cacher un bit ou plus dans une seule phrase, comme nous allons citer également les avantages de ces méthodes par rapport aux autres méthodes proposées en arabe et en d'autres langues 4.7, et nous allons également donner les resultats obtenus après l'aplication des méthodes sur des différents corpus 4.5 et terminer le chapitre en expliquant les limites des méthodes 4.8, puis par une conclusion.

4.2 Grammaire arabe

Les linguistes en langue arabe classent en trois sections principales les composantes principales de la phrase[58, 59, 64] :

1. **Nom** : il est de sens indépendant, il n'est pas associé à une durée spécifique, tels les noms des choses : table, chaise, relié généralement à des propositions et des déterminants, le nom est divisé dans la langue arabe en deux parties : nom d'emprunt et nom commun.
2. **Verbe :** il est indépendant du signifié du mot, et est associé à un temps donné, de sorte qu'il indique que l'événement se produit à un moment donné, et les verbes se divisent en langue arabe :
   (a) **Le verbe au passé :** ou l'action s'est déroulée dans un temps passé.
   (b) **Le verbe à l'impératif :** c'est donner un ordre ou un conseil.
   (c) **Le verbe au présent :**l'action se déroule au moment où l'on parle.
3. **Les conjonctions :** elles servent à relier deux phrases pour compléter le sens. Nous pouvons distinguer quatre types :
   (a) **Les conjonctions de prépositions :** servent à relier entres deux mots de natures différentes(verbes-Noms). Quelques conjonctions de prépositions sont montrées dans la figure 4.1

FIGURE 4.1 – Les conjonctions de prépositions

   (b) **Les conjonctions de coordinations :** reliant les noms a ses précédents ou deux actions qui se suivent. La figure 4.2 montre quelques conjonctions de coordinations utilisées dans la langue arabe.

FIGURE 4.2 – Les conjonctions de coordinations

(c) **Les conjonctions de négation :** ce sont des conjonctions oú l'action ne se réalise pas. Quelques conjonctions de négation sont montrées dans la figure 4.3

FIGURE 4.3 – Les conjonctions de négation

(d) **Les conjonctions de condition :** ce sont des prépositions qui mettent la condition de réalisation de la première préposition avec la seconde préposition. La figure 4.4 montre quelques conjonctions de condition dans la langue arabe

FIGURE 4.4 – Les conjonctions de condition

En fonction de la division du mot, la phrase peut être divisée en :

1. **Phrase simple :**la phrase simple se divise en :
   (a) **Phrase nominale/al jumlat-ul-ismiyya/** : cette phrase commence par un nom ou un prenom. Elle est composée de deux parties : la première partie appelée "mubtada'" est le sujet de la phrase. La deuxième partie, appelée "khabar" , apporte une information sur le sujet.

— Le mubtada' est le sujet de la phrase. Il peut représenter une personne ou une chose. Il est décrit ou précisé par El-khabar de la phrase. Par exemple, dans la phrase : Mohamed est debout. « Mohamed » est le mubtada‘ – sujet et « est debout » est le khabar – description ou précision sur le sujet. Le mubtada' peut être composé de :

— **Un nom défini**

Dans la phrase montrée dans la figure 4.5 "L'arabe est une langue " le mobtada' est le nom 'L'arabe' et il est exprimé par un nom défini.

FIGURE 4.5 – Mobtadaa'-nom défini-

— **Un pronom démonstratif**

Dans la phrase montrée dans la figure 4.6 "Ceci est un livre " le mobtada' est le nom 'Celui-ci' et il est exprimé par un pronom démonstratif.

FIGURE 4.6 – Mobtadaa'- pronom démonstratif-

— **Un nom propre**

Dans la phrase montrée dans la figure 4.7 "Ahmed est un savant "le mobtada'" est exprimé par un "Nom propre" qui est "Ahmed".

FIGURE 4.7 – Mobtada'-un nom propre-

— **Un pronom**
Dans la figure 4.8 le mobtada' dans la phrase "C'est une bonne méthode" est éxprimé par Un pronom "c'est".

FIGURE 4.8 – Mobtada'-un pronom-

— Le khabar quant à lui, n'a pas l'obligation d'être composé d'un nom ou d'un pronom mais lui aussi est toujours au cas nominatif -Marfu'- bien qu'il puisse être composé de :
— **Un seul nom**
Dans la phrase "La science est une lumière" montrée dans la figure 4.9, el-khabar est exprimé par un seul nom "lumière".

FIGURE 4.9 – Khabar-un seul nom-

— **Une phrase ou un complément introduit par une préposition**
Dans la phrase "L'homme dans la maison" montrée dans la figure 4.10, el-khabar est exprimé par un complément et une préposition "dans la maison".

FIGURE 4.10 – Khabar-complément et une préposition-

— **Phrase verbale**
Dans la figure 4.11, dans la phrase "L'homme entre à la maison", el-khabar est exprimé par une phrase varbale "entre à la maison".

FIGURE 4.11 – Khabar-phrase verbale-

— **Une phrase nominale**
Dans la phrase "Mohamed, sa mère est pieuse." montrée dans la figure 4.12, el-khabar est exprimé par une phrase nominale "sa mère est pieuse.".

FIGURE 4.12 – Khabar-une phrase nominale-

On peut y introduire ou débuter la phrase nominale par :
— **Kâna et ses soeurs**
La figure 4.13 montre un exemple de "Kâna et ses soeurs" les plus utilisées dans le texte arabe.

FIGURE 4.13 – Kâna et ses soeurs

— **Înâ et ses soeurs**
La figure 4.14 montre Înâ et quelques exemples de ses soeurs.

FIGURE 4.14 – Înâ et ses soeurs

— **Les verbes d'approches, les verbes de souhaits, les verbes d'actions**
Dans la figue 4.15, nous avons donné un exemple de deux verbes d'approches, deux verbes de souhaits et deux verbes d'actions.

FIGURE 4.15 – Les verbes d'approches, de souhaits et d'actions

(b) **La phrase verbale**
Elle se compose d'un verbe porte un sens, et un agent. Quand le verbe est transitif, nous aurons : (verbe + agent suffixé) + Complément d'objet .

(c) **la Semi-phrase**
Ce sont des phrases complétées par des adverbes ou des prépositions circonstancielles de lieux ou de temps ...etc.
— **Adverbes de temps**

Est un nom qui permet d'indiquer le temps dans lequel l'action se réalise. La figure 4.16 montre quelques adverbes de temps.

FIGURE 4.16 – Semi phrase -adverbes de temps-

— **Adverbes de lieux**
Indique l'emplacement, le lieu du déroulement d'une action. La figure 4.17 montre un exemple de quelques adverbes de lieux.

FIGURE 4.17 – Semi phrase -adverbes de lieux-

— **Les conjonctions de préposition et le complément d'objet indirect**
Est une phrase composée de "conjonction de Préposition" et "complément d'objet indirect"
Exemple :

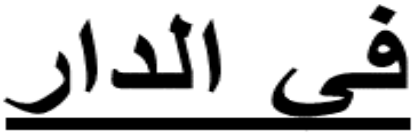


FIGURE 4.18 – Semi phrase -conjonctions de préposition-

Dans la figure 4.18, la phrase "Dans la maison" est composée d'une préposition "Dans" et un complément d'objet indirect "la maison"

2. **Phrase complexe**
Elle est composée de deux phrases qui peuvent êtres :

— **Phrase conditionnelle**
Ce type de phrase est composé d'un outil qui introduit la condition(Les conjonctions de condition), d'une proposition qui exprime la condition et d'une proposition "réponse".

4.3 Les méthodes proposées

Nous allons choisir la méthode selon le type de phrase :

1. **Phrase nominale**
Comme nous avons dit dans la partie précédente 4.2 la phrase nominale se compose de "mubtada'" et "khabar", mais dans certains cas, nous pouvons :
— **Supprimer "mubtada'"**
(a) **Méthode 1**
Dans cette méthode, nous utilisons la phrase avec "mubtada'" pour cacher le bit 1, et nous pouvons le supprimer si nous voulons cacher 0. Exemple :

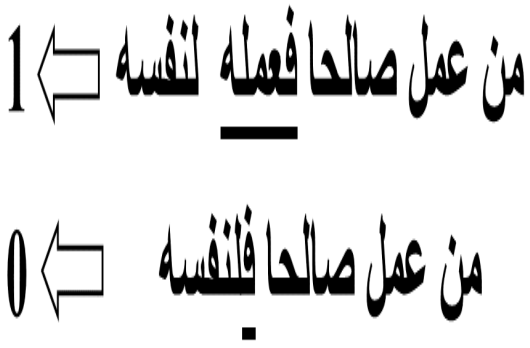


FIGURE 4.19 – Exemple de méthode1 -supprimer "mubtada'"-

Dans la phrase montrée dans la figure 4.19, dans la deuxième phrase nous avons supprimé "El-mobtada'" sans changer le sens de la phrase.

(b) **Méthode 2**
Nous pouvons supprimer "mubtada'" dans la réponse à la question si nous voulons cacher 0 et nous pouvons le laisser si nous voulons cacher 1. Exemple :

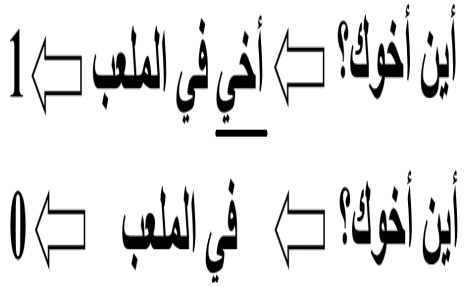


FIGURE 4.20 – Exemple de méthode2 -supprimer "mubtada'"-

Dans la deuxième phrase montrée dans la figure 4.20, le "mobtada'" dans la réponse à la question est supprimer, mais le sens de la phrase n'a pas changé.

(c) **Méthode 3**

Dans cette méthode, nous montrons "mubtada'" dans la phrases déclaratives seulement si nous voulons cacher 1, si non on le supprime. comme le montre la figure 4.21

FIGURE 4.21 – Exemple de méthode 3 -supprimé "mubtada'"-

(d) **Méthode 4**

Nous pouvons supprimer "mubtada'" après le pronom "ba'l" pour cacher 0, si non on le laisse pour cacher 1. Exemple :

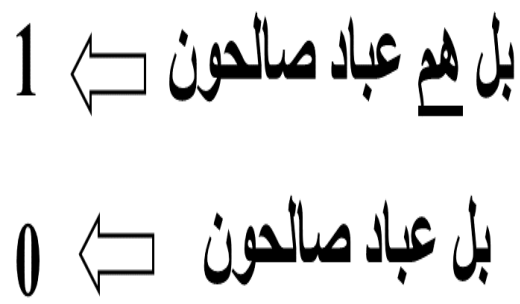


FIGURE 4.22 – Exemple de méthode 4-supprimé "mubtada'"-

Dans la figure 4.22, le "mobtada'" dans la deuxième phrase est supprimé après "ba'l" , mais le sens de la phrase n'a pas changé.

— **supprimer "El-khabar"**

(a) **Méthode1**

Nous pouvons, dans certains cas supprimer "El-khabar" si nous voulons cacher 0, si non, nous pouvons le laisser pour cacher 1. Comme le montre la figure 4.23

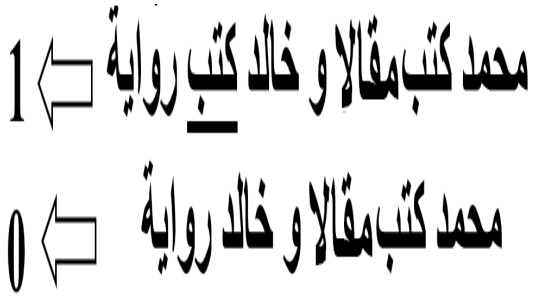


FIGURE 4.23 – Exemple de méthode 1-supprimer "El-khabar"-

(b) **Méthode2**

Nous pouvons supprimer "El-khabar" dans la réponse à la question comme le montre la deuxième phrase de la figure 4.24 pour cacher 0 et le laisser pour cacher 1. Exemple :

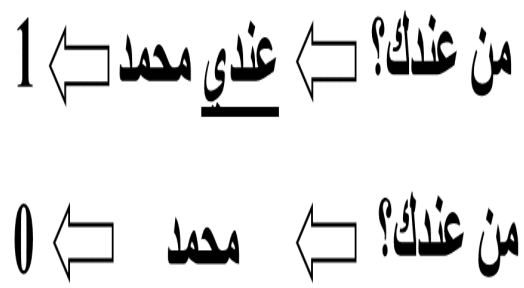


FIGURE 4.24 – Exemple de méthode 2-supprimer "El-khabar"-

(c) **Méthode3**

Nous pouvons supprimer "El-khabar" après "ida" si nous voulons cacher 1 comme le montre la deuxième phrase dans la figure 4.25, mais si nous voulons cacher 0, nous le laissons comme l'exemple de la première phrase dans la figure 4.25

FIGURE 4.25 – Exemple de méthode 3-supprimer "El-khabar"-

— **Supprimer "Mubtada'" et "Khabar"**

(a) **Méthode 1**

Nous pouvons supprimer "Mubtada'" et "Khabar" dans une phrase pour cacher 0, comme le montre l'exemple de la deuxième phrase , et les laisser pour cacher 1, comme le montre la première phrase dans la figure 4.26.

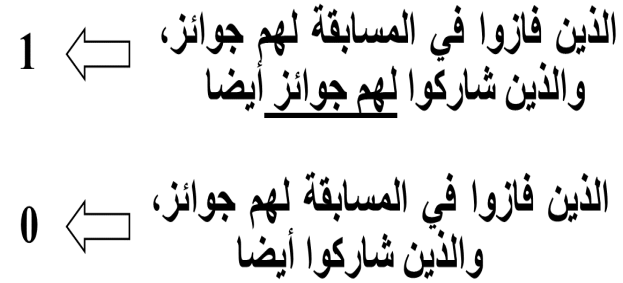


FIGURE 4.26 – Exemple de méthode 1-supprimer "Mubtada'" et "Khabar"-

(b) **Méthode 2**

Nous pouvons supprimer "El-Mubtada'" et "El-Khabar" dans la réponse à la question, comme le montre la deuxième phrase dans la figure 4.27, pour cacher 0, et les laisser pour cacher 1. Exemple :

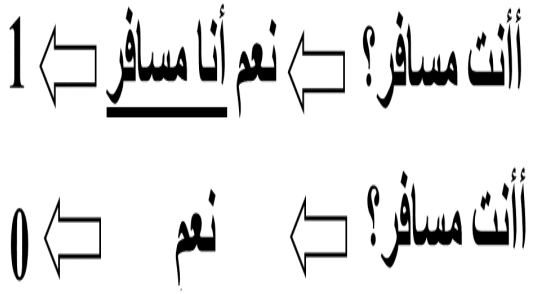


FIGURE 4.27 – Exemple de méthode 2-supprimer "Mubtada'" et "Khabar"-

— **la phrase nominale peut débuter aussi par :** Les verbes d'approches, les verbes de souhaits, les verbes d'actions, Kânâ et ses soeurs, Înâ et ses soeurs et les outils d'appel...etc.
Nous pouvons utiliser ces cararactéristiques aussi pour dissimuler l'information.

(a) **Méthode1**

Nous pouvons commencer par "khabar Kânâ" ou l'un de ses soeurs, comme le montre le deuxième exemple dans la figure, pour cacher 0 et le laisser comme il est, comme le montre l'exemple de la première phrase dans la figure 4.28, pour cacher 1. Exemple :

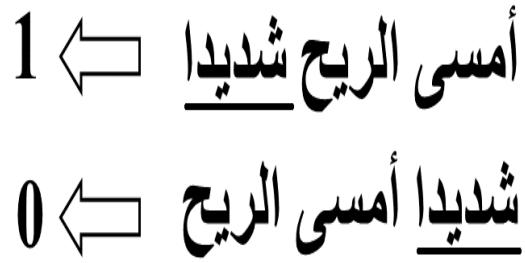


FIGURE 4.28 – Exemple de méthode 1 -kânâ" et ses soeurs-

(b) **Méthode2**

Après "Layssa", nous pouvons mettre le nom après la conjonction "Waw" avec le diacritique "Dhamma'" pour cacher 0, comme le montre le deuxième exemple dans la figure 4.29, et nous pouvons le mettre avec le diacritique "fatha" pour cacher 1, comme le montre la première phrase. Exemple :

FIGURE 4.29 – Exemple de méthode2 -kânâ" et ses soeurs-

(c) **Méthode3**

Après "El-Khabar" des verbes d'approches il existe toujours le pronom "an" et devant d'autres verbes aussi. Nous pouvons le laisser, comme le montre l'exemple de la première phrase dans la figure 4.30, pour cacher le bit 0, et supprimer "an" (transformer le verbe en "Masdaar"), comme le montre la deuxième phrase dans la figure 4.30, pour cacher le bit 1. Exemple :

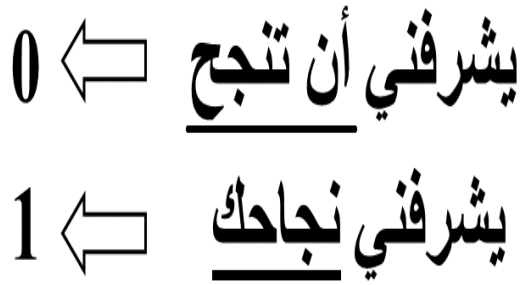


FIGURE 4.30 – Exemple de méthode3 -verbes d'approches-

(d) **Méthode4**

Nous pouvons supprimer " les outils d'appels " dans une phrase, comme le montre l'exemple 2 de la figure, pour cacher 0, et le laisser, comme le montre l'exemple 1 dans la figure 4.31, pour cacher 1. Exemple :

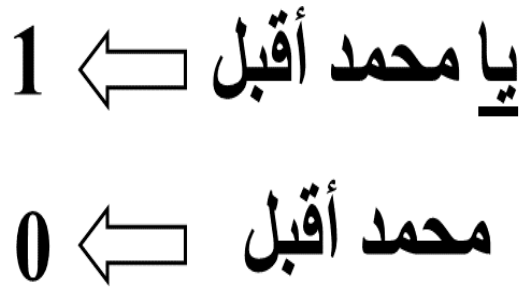


FIGURE 4.31 – Exemple de méthode4 -outils d'appel-

(e) **Méthode5**

Nous pouvons supprimer l'adverbe "fi" dans une phrase pour dissimuler le bit 1, comme le montre la première phrase, et l'ajouter pour cacher le bit 0, comme le montre le deuxième exemple de la figure 4.32. Exemple :

FIGURE 4.32 – Exemple de méthode5 -adverbe de temps-

(f) **Méthode6**

Nous pouvons aussi retarder les adverbes de lieu pour cacher le bit 0, comme le montre la deuxième phrase dans la figure 4.33, et les laisser tel qu'il est, comme le montre la première phrase, pour cacher 1. Exemple :

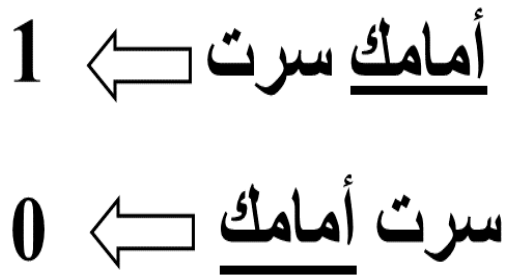


FIGURE 4.33 – Exemple de méthode6-adverbes de lieu-

— **La phrase complexe :** dans les phrases complexe, nous pouvons utiliser les deux propositions conditionnelles.

(a) **Méthode1**

Nous pouvons commencer par la présuposition conditionnelle "Réponse" pour cacher le bit 1, comme le montre la deuxième phrase dans la figure 4.34, si non, nous commençons par la phrase "Conditionnel" pour cacher le bit 0, comme le montre la première phrase dans la figure 4.34. Exemple :

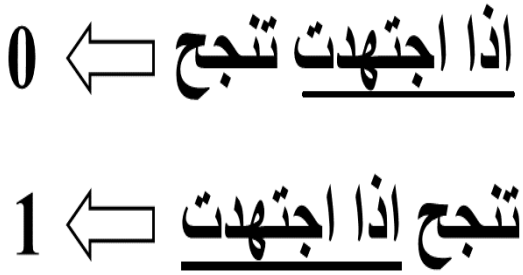


FIGURE 4.34 – Exemple de méthode6 -phrase conditionnelle-

— **Les conjonctions de cordinations**

(a) **Méthode1**

Nous changeons l'ordre des verbes ou des noms reliés seulement par les conjonction "Wa" et "Aw", comme le montre l'exemple de la deuxième phrase dans la figure 4.35 pour cacher 1 et nous laissons l'ordre tel qu'il est, comme le montre la première phrase, pour cacher 0. Exemple :

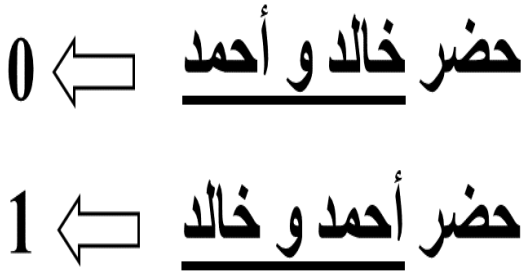


FIGURE 4.35 – Exemple de méthode1-les conjonctions-

(b) **Méthode2**

Nous pouvons remplacer un mot par son synonyme pour cacher le bit 1, et le laisser tel qu'il est pour cacher le bit 0. La deuxième phrase dans la figure 4.36, montre un exemple de synonyme du premier et de dernier mot de la première phrase. Exemple :

FIGURE 4.36 – Exemple méthode2-synonyme-

Algorithme

Pour appliquer les méthodes précédentes, nous allons suivre les étapes suivantes :

— Lire le texte de couverture.
— Traduire le message secret en binaire.
— Utiliser toutes les méthodes qui peuvent être appliquées sur chaque phrase.
— On reprend les même étapes sur chaque phrase jusqu'à la fin du message secret.

4.4 Robustesse, Capacité et Sécurité

Les méthodes proposées, sont plus **robustes** que les autres proposées dans la littérature pour la langue arabe et même quelques méthodes dans les autres langues parce que :

— Elles n'ont pas été détecté par les programmes d'OCR parce qu' aucun changement n'à été fait sur les caractéristiques des lettres ou de la langue, ainsi les programme d'OCR ne vérifie pas la sémantique.
— Elles respectent les caractéristiques de la langue, alors elles ne nécessitent pas une police spéciale.
— La fonction de copier-coller, ne pose aucun problème, ainsi que l'impression.

La sécurité dans les méthodes est aussi augmentée, parce que selon les expérimentations , plusieurs personnes n'ont pas remarqués la différence entre le texte originale et le texte modifié, en plus elles sont difficilement détectées parce que nous utilisons plusieurs méthodes dans un texte et même dans une seule phrase. Et ainsi dans ces méthodes, nous ne cachons pas le message secret directement nous cachons le message inverse de façon à rendre la détection du message secret plus difficile.

La capacité dans nos méthodes est faible par rapport aux autres qui utilisent les lettres, mais elles offre une capacité meilleur que les autres méthodes qui utilisent la sémantique dans les autres langues.

4.5 Résultats et expérimentations

Dans cette section, nous discuterons les capacités obtenues par les méthodes proposées en étudiant 6 textes. Les textes sont de différents types : une petite histoire, un article sportif, scientifique, médicale, sociale et un article politique. La longueur des textes varie entre 23 et 82 phrases.

Pour obtenir la capacité, nous avons évalué chaque phrase dans le texte et le nombre des bits que nous pouvons cacher dans cette phrase.

Dans chaque texte, nous avons tout d'abord marqué toutes les possibilités de transformations, puis selon le type de la phrase nous sélectionnons les règles qui peuvent être appliquées , et appliquer les modifications correspondantes sans changement du sens.

En moyenne, dans chaque texte, nous avons pu cacher plus que la moitié du nombre des phrases, ou parfois plus du nombre des phrase trouvées dans le texte.

Nous avons aussi demandé a 20 personnes d'âge différents (17-50), de l'éducation et du travail (éducation supérieure incomplète et l'enseignement supérieur en informatique et technologie) de lire les deux textes et de dire s'il y a une différence entre les deux, et s'il y a une différence, ce qu'est exactement dans le texte a été changés.

Aucune limite de temps pour réaliser l'expérience. Ils pouvaient lire les textes tout le temps sont ils ont besoin pour prendre une décision.

La plupart des lecteurs n'ont rien remarqué lorsque nous avons utilisés les autres méthodes, sauf si nous avons utilisés la méthode des synonymes plusieurs

fois dans le texte. Ci-dessous le tableau des résultats :

TABLE 4.1 – Résultats de l'expérimentation

corpus	type	Nombre de phrase	Nombre des bits cachés	Rapport
hawrah1980. blogspot.com	histoire	49	40	0.82
http://mawdoo3. com	Article sur la pollution	53	37	0.70
http://www. kaheel7.com/ ar/index.php/ 2010-02-02-22-17-58/ 1929--	Article médicale	47	50	1.06
https://ar. wikipedia.org/ wiki/	Article scientifique	30	30	1
http://www. alriyadh.com/ 732667	Article du sport	23	43	1.87
http://www. palestineremembered. com/GeoPoints/ al_Samu__1095/ Article_16114. html	Article politique	82	55	0.67

4.6 Buts des méthodes

le but de ces méthodes est beaucoup plus, pour protéger les droits d' auteurs et les droits des propriétés intelectuelles, ou même de cacher des petits messages.

4.7 Avantages des méthodes

Par rapport aux méthodes proposées dans la langue arabe

Ces méthodes sont les premièrs méthodes proposées dans la langue arabe qui permet de cacher le message secret en utilisant la semantique, aussi, elles

sont plus robustes que les autres qui utilisent les caractéristiques des lettres ou les diacritiques, parce qu'elles ne peuvent pas être détectées par le programme de**OCR**, et elles peuvent être appliquées au texte sous n'importe quelle forme, et elles ne posent pas de problème avec le processus de copier et coller et elle n'ont pas besoin d'une police spéciale, ainsi, elles respectent les caractéristiques de la langue. Elles sont également plus secrètes, et n'attirent pas l'attention.

Par rapport aux autres méthodes proposées dans la sémantique dans les autres langues

Dans les autres langues, autres méthodes ont proposé et surtout dans la langue anglaise, mais généralement elles utilisent un seul type de phrase et peuvent cacher au maximum un seul bit dans une phrase, alors la capacité de ces méthodes dépend de l'existence de ce type de phrase dans le texte. Dans nos méthodes, nous avons proposé plusieurs techniques, pour augmenter le nombre des phrases utilisé afin d'augmenter la quantité des bits cachés, où, dans une phrase, nous pouvons cacher un bit ou plus.

4.8 Limites

Malgré les avantages de ces méthodes, ses applications sont limitées, par exemple, elles ne peuvent pas être appliquées dans les textes qui n'acceptent pas le changement, comme la poésie et les citations, ainsi que la capacité est limitée par rapport aux certaines méthodes qui utilisent chaque lettre de la phrase pour cacher l'information.

4.9 Conclusion

Dans ce chapitre, nous avons présenté les premières méthodes proposées pour cacher les messages secrets dans le texte arabe en utilisant la sémantique qui est plus robuste et plus sécurisée.

Conclusion générale et perspectives

Dans ce mémoire de fin d'études, nous avons présenté les résultats de nos recherches dans le domaine de la dissimulation d'information. Nous avons commencé par une introduction oú nous avons parlé de nos objectifs, les motivations 1.

Après nous avons retracé l'historique de stéganographie et expliqué les différentes méthodes de dissimulation et les comparer avec la cryptographie, ainsi les différents types, supports, techniques et enfin la sécurité dans la stéganographie 2.

Ensuite, nous avons expliqué les différents types de stéganographie texte avec quelques exemples et nous avons donné un aperçu détaillé sur les œuvres publiées dans la dissimulation dans les textes arabes, et les avons classé selon les techniques utilisées, et nous avons discuté leurs capacités, robustesses et sécurités 3.

Nous avons également expliqué nos propres méthodes qui utilisent les règles de grammaires arabe pour cacher les bits secrets et nous avons testé nos méthodes sur six textes différents 4.

Nous restons convaincus que nos méthodes proposées malgré leurs robustesses et sécurités très élevées, n'offrent pas une grandes capacités comme quelques autres méthodes proposées dans la syntaxique, alors parmi les perspectives envisagées, nous proposons de s'intéresser a augmenter la capacité en utilisant plus des règles qui permettent de cacher plus d'informations dans chaque phrase du texte.

Annexes

Texte originale

قصة ملخص الحياة :

طلب ولد من أبيه أن يلخص له خبرته في الحياة والحكمة التي أخذها منها، فقال له : هل تقدر على الاستماع ؟

فقال : نعم .

فقال يابني : إيّاك أن تتكلم في الناس والأشياء إلا بعد أن تتأكد من صحة المصدر، وإذا جاءك أحد بنبأ فتبين قبل أن تتهور! وإيّاك والشائعة، لا تُصدق كل ما يقال ولا نصف ما تبصر، وإذا ابتلاك الله بعدو قاومه بالإحسان إليه، ادفع بالتي هي أحسن فستنقلب العداوة حباً.

إذا أردت أن تكتشف صديقاً سافر معه ! ففي السفر ينكشف الإنسان ويذوب المظهر وينكشف المخبر، ولماذا سمي السفر سفرا ؟ إلا لأنه عن الأخلاق والطبائع يُسفر.

وإذا هاجمك الناس وأنت على حق فافرح ! لأنهم يقولون لك أنت ناجح ومؤثر، ولا يُرمى إلا الشجر المثمر .

بني عندما تنتقد أحداً فبعين النحل تعوّد أن تبصر ولا تنظر للناس بعين ذباب فتقع على ما هو مستقذر ! نم باكراً يا بني فالبركة في الرزق صباحاً، وأخاف أن يفوتك رزق الرحمن لأنك تسهر، وحينما يثق بك أحد، فإياك ثم إياك أن تغدر .

سأذهب بك لعرين الأسد، وسأعلمك أن الأسد لم يصبح ملكاً للغابة لأنه يزأر ! ولكن لأنه عزيز النفس، لا يقع على فريسة غيره مهما كان جائعاً يتضور، فلا تسرق جهد غيرك فتتجوّر !

سأذهب بك للحرباء، حتى تشاهد بنفسك حيلتها ! فهي تلون جلدها بلون المكان، لتعلم أن في البشر مثلها نسخ تتكرر .

تعود يا بني أن تشكر، اشكر الله فيكفي أنك مسلم ويكفي أنك تمشي وتسمع وتبصر، اشكر الله واشكر الناس، فالله يزيد الشاكرين، والناس تحب الشخص الذي عندما تبذل له يقدر .

أعظم فضيلة في الحياة هي الصدق، واعلم ان الكذب وإن نجا هو أرذل رذيلة .

بني وقّر لنفسك بديلاً لأي شيء، استعد لأي أمر، حتى لا تتوسل لنذل يذل ويحقر .

واستفد من كل الفرص، لأن الفرص التي تأتي الآن قد لا تتكرر.

لا تتشكى ولا تتذمر ! أريدك متفائلاً مقبلاً على الحياة، اهرب من اليائسين والمتشائمين، وإياك أن تجلس مع رجل يتطير !

لا تتشمت ولا تفرح بمصيبة غيرك، وإياك أن تسخر من شكل أحد، فالمرء لم يَخلق نفسه ! ففي سخريتك أنت في الحقيقة تسخر من صنع الذي أبدع وخلق وصور، اللهم ابني لأبي بيتاً في الجنة و اجعل ملتقانا هناك .

Stégo-texte

قصة ملخص الحياة :

طلب ولد من أبيه تلخيص خبرته في الحياة والحكمة التي أخذها منها، فقال له : هل تقدر على الاستماع ؟

فقال : نعم أقدر على الاستماع.

فقال بني :

إيّاك والتكلم في الأشياء و الناس إلا بعد **التأكد** من صحة المصدر، و تبين إذا جاءك أحد بنبأ قبل التهور ! وإيّاك والشائعة، لا تُصدق نصف ما تبصر ولا كل ما يقال، و قاوم بالاحسان إذا ابتلاك الله بعدو، ادفع بالتي هي أحسن فستنقلب العداوة حباً.

سافر مع الصديق إذا أردت أن تكتشه! ففي السفر يذوب المظهر و ينكشف المخبر و ينكشف الإنسان، ولماذا سمي السفر سفرا؟ سمي السفر سفرا إلا لأنه عن الطبائع و الأخلاق يُسفر .

وافرح إذا هاجمك الناس وأنت على حق ! لأنهم يقولون لك أنت مؤثر و ناجح، ولا يُرمى إلا الشجر المثمر.

يا بني عندما تنتقد أحداً فبعين النحل تعوّد الابصار ولا تنظر للناس بعين ذباب فتقع على ما هو مستقذر !

يا بني نم باكراً فالبركة في الرزق صباحاً، وأخاف افاتتك رزق الرحمن لأنك تسهر، و**عندما** يثق بك أحد، فإياك ثم إياك الغدر

سأذهب بك لعرين الأسد، وسأعلمك أن الأسد لم يصبح ملكاً للغابة لأنه يزأر! **بل** لأن الاسد عزيز النفس، لا يقع على فريسة غيره مهما جائعاً كان يتضور، فلا تسرق جهد غيرك فتتجوّر !

سأذهب بك للحرباء، حتى تشاهد بنفسك حيلتها! فهي تلون جلدها بلون المكان، لتعلم أن في البشر مثلها نسخ تتكرر .

تعود بني الشكر، اشكر الله فيكفي أنك مسلم و يكفي أنك تسمع وتبصر و تمشي ، اشكر الله واشكر الناس، فالله يزيد الشاكرين، والناس تحب الشخص الذي يقدر عندما تبذل له .

الصدق أعظم فضيلة في الحياة، واعلم ان الكذب هو أرذل رذيلة وإن نجا .

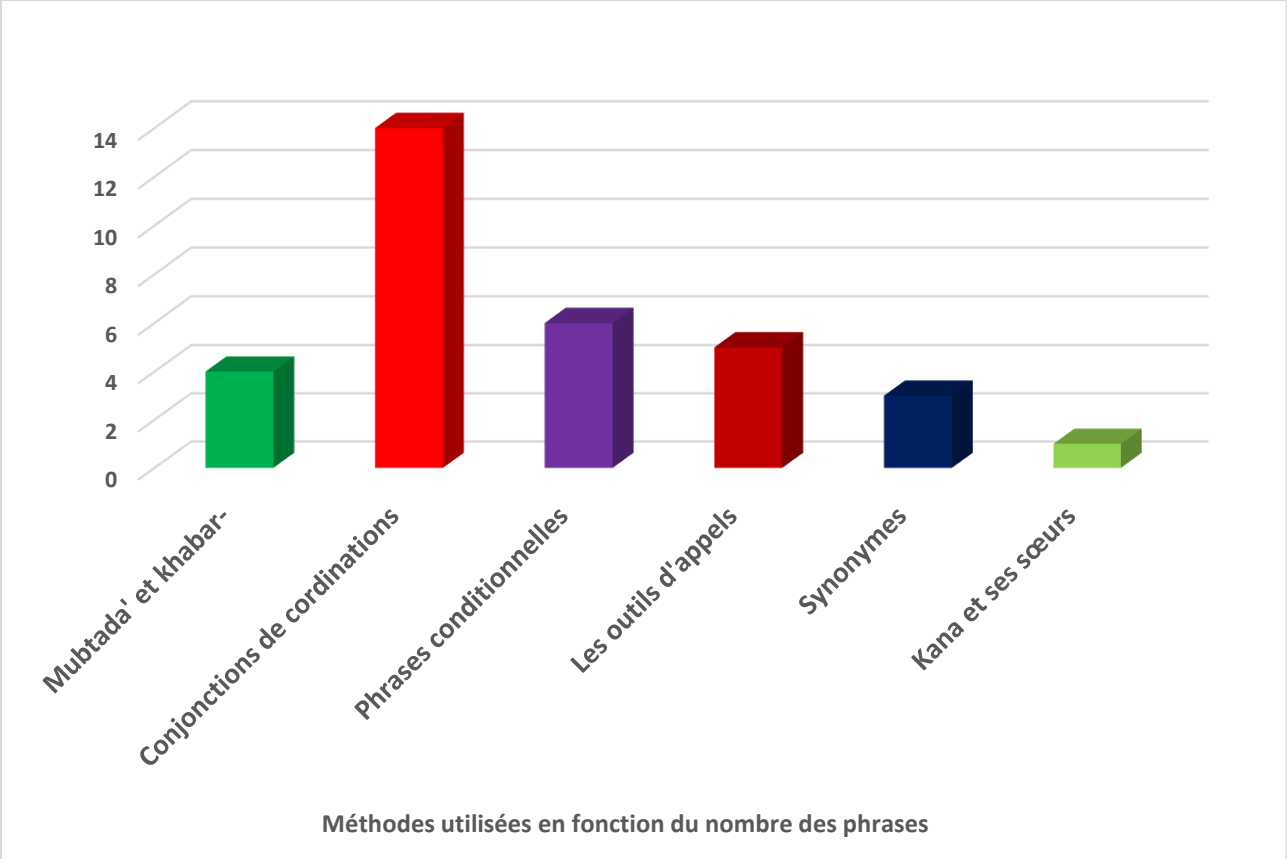
يا بني وفّر لنفسك بديلاً لأي شيء، استعد لأي أمر، حتى لا تتوسل لنذل يحقر ويذل .

واستفد من كل الفرص، لأن الفرص التي تأتي الآن قد لا تتكرر.

لا تتذمر و لا تتشكى ! أريدك متفائلاً مقبلاً على الحياة، اهرب من المتشائمين و اليائسين، وإياك أن تجلس مع رجل يتشاءم !

لا تفرح و لا تتشمت بمصيبة غيرك، وإياك و السخرية من شكل أحد، فالمرء لم يَخلق نفسه ! ففي سخريتك أنت في الحقيقة تسخر من صنع الذي خلق و صور و أبدع, اللهم ابني لأبي بيتآ في الجنة و اجعل ملتقانا هناك .

Le graphe montre le nombre des phrases utilisées dans chaque méthode dans ce texte.



Texte originale

مقال عن التلوث

أصبحنا اليوم نعاني من التلوث الشديد للبيئة، وهو من أخطر انواع التلوث ويرجع السبب الرئيسي للتلوث هو الأنسان والمتضرر منه أيضا الانسان والكائنات الحية الأخرى.

التلوث هو

احداث تغير في البيئة التي تحيط بالكائنات الحية بفعل الإنسان وأنشطته اليومية مما يؤدي إلى ظهور بعض الموارد التي لا تتلاءم مع المكان الذي يعيش فيه الكائن الحي ويؤدي إلى اختلاله

من السبب وراء تلوث البيئة؟

الإنسان هو السبب الرئيسي والأساسي في أحداث عملية التلوث في البيئة وظهور جميع الملوثات بأنواعها المختلفة فالتوسع الصناعي والتقدم التكنولوجي وسوء استخدام الموارد كلها مرتبطة بالإنسان فالإنسان يتكاثر ويصنع ويستخدم هذه المواد.

مستويات التلوث

1-التلوث غير الخطير

وهو التلوث المتجول الذي يستطيع الإنسان أن يتعايش معه بدون أن يتعرض للضرر أو المخاطر كما انه لا يخل بالتوازن البيئي وفي الحركة التوافقية بين عناصر هذا التوازن.

2-التلوث المدمر

وهو التلوث الذي يحدث فيه انهيار للبيئة والإنسان معا، ويقضي على كافة أشكال التوازن البيئي وهو متصل بالتطور التكنولوجي الذي يضن الإنسان انه يبدع فيه يوماً بعد يوم ويحتاج اصلاح هذا الخطأ سنوات طويلة ونفقات باهظة.

انواع التلوث

- تلوث الهواء
- التلوث بالنفايات
- التلوث البصري
- التلوث السمعي
- تلوث الماء
- تلوث التربة

•التلوث الغذائي

أ ضرار التلوث

•ظهور المشاكل البيئية المختلفة ومن ضمنها الانفجار السكاني
•المطر الحمضي
•اختلال التنوع البيولوجي وانقراض بعض مظاهر الحياة النباتية والحيوانية
•أكل طبقة الأوزون
• ظاهرة الاحتباس الحراري
•ظاهرة التصحر وفقر التربة الزراعية
•تعرض المجال الجوي للمطارات للتلوث الجوي مما يؤدي الى خفض مجال الرؤية
•الانقلابات الحرارية وعدم استقرار المناخ.
إلحاق أضرار بالآثار فالتركيزات العالية من اكاسيد الرصاص والكبريت تعمل على تآكل ألوان الآثار على مدار آلاف السنين وذلك لقدرة تلك الاكاسيد على التفاعل مع مكونات تلك الألوان
• .حدوث الحرائق عن طريق الاشتعال الذاتي للغازات السامة القابلة للاشتعال
نسب متزايدة من الاكاسيد الضارة والمعادن الثقيلة العالقة بالهواء وخاصة الرصاص الذي ساهم بها قطاع صهر المعادن وتوليد الكهرباء ومصانع البلاستك والكيماويات
•عدم سهولة تنقية مياه الصرف الصحي
•بقاء المكونات الصناعية بالتربة الزراعية لفترة طويلة من الزمن
•تقليص مساحات الأراضي الزراعية لمقابلة الغزو الصناعي
•تزايد نشر الرطوبة الجوية بالهواء بكثرة المسطحات المائية لصرف المخلفات الصناعية
زيادة التدفق الحراري الاتية من المناطق الصناعية والمحملة بالملوثات المختلفة من العوالق والتربة والدخان

-: علاج التلوث

• .الوعي الذاتي لدى الشخص لان التلوث ينذر بفنائه •
• وقف تراخيص مزاولة النشاط الصناعي الذي يدمر البيئة
•تهجير الصناعات الملوثة للبيئة بعيدا عن اماكن المدن
•تطور اساليب مكافحة تلوث الهواء
• تطوير وسائل التخلص من القمامة والنفايات وخاصة عمليات حرق النفايات في الهواء الطلق
• القيام بعمليات التشجير على نطاق واسع للتخلص من ملوثات الهواء وامتصاصها.
الكشف الدوري للسيارات ومراقبة عوادمها
•اللجوء الى الغاز الطبيعي كإحدى مصادر الطاقة البديلة عن مصادر الطاقة الحرارية
• معالجة التلوث النفطي بإضافة مذيبات كيمياوية لترسيبه في قاع لمياه
اقامة المحميات البحرية التي تشتمل على كائنات بحرية نادرة مهددة بالانقراض.
.• اللجوء الى استخدام المبيدات العضوية والموارد الطبيعية والابتعاد عن المبيدات الكيمياوية

Stégo-texte

مقال عن التلوث

أصبحنا اليوم نعاني من التلوث **الكبير** للبيئة، وهو من أخطر **أشكال** التلوث ويرجع السبب الرئيسي للتلوث هو الأنسان والمتضرر منه أيضا **الكائنات الحية الأخرى والانسان**.

التلوث

هو احداث تغير في البيئة التي تحيط بالكائنات الحية بفعل الإنسان وأنشطته اليومية مما يؤدي إلى ظهور بعض الموارد التي لا تتلاءم مع المكان الذي يعيش فيه الكائن الحي ويؤدي إلى اختلاله

من السبب وراء تلوث البيئة؟

الإنسان السبب الأساسي والرئيسي في ظهور **جميع الملوثات بأنواعها المختلفة و احداث عملية التلوث في البيئة، فالتقدم التكنولوجي والتوسع الصناعي** وسوء استخدام الموارد كلها مرتبطة بالإنسان **فالإنسان يصنع ويستخدم هذه المواد و يتكاثر**.

مستويات التلوث

1-التلوث غير الخطير

وهو التلوث المتجول الذي يستطيع الإنسان **التعايش** معه بدون **التعرض للمخاطر أو للضرر** كما انه لا**يفسد** التوازن البيئي وفي الحركة التوافقية بين عناصر هذا التوازن.

2-التلوث المدمر

وهو التلوث الذي يحدث فيه انهيار **للإنسان و البيئة معا**، ويقضي على كافة أشكال التوازن البيئي وهو **يتعلق** بالتطور التكنولوجي الذي يضن الإنسان انه يبدع فيه يوماً بعد يوم ويحتاج اصلاح هذا الخطأ **نفقات باهظة و سنوات طويلة**.

انواع التلوث

- تلوث الهواء
- التلوث بالنفايات
- التلوث البصري
- التلوث السمعي
- تلوث الماء
- تلوث التربة

•التلوث الغذائي

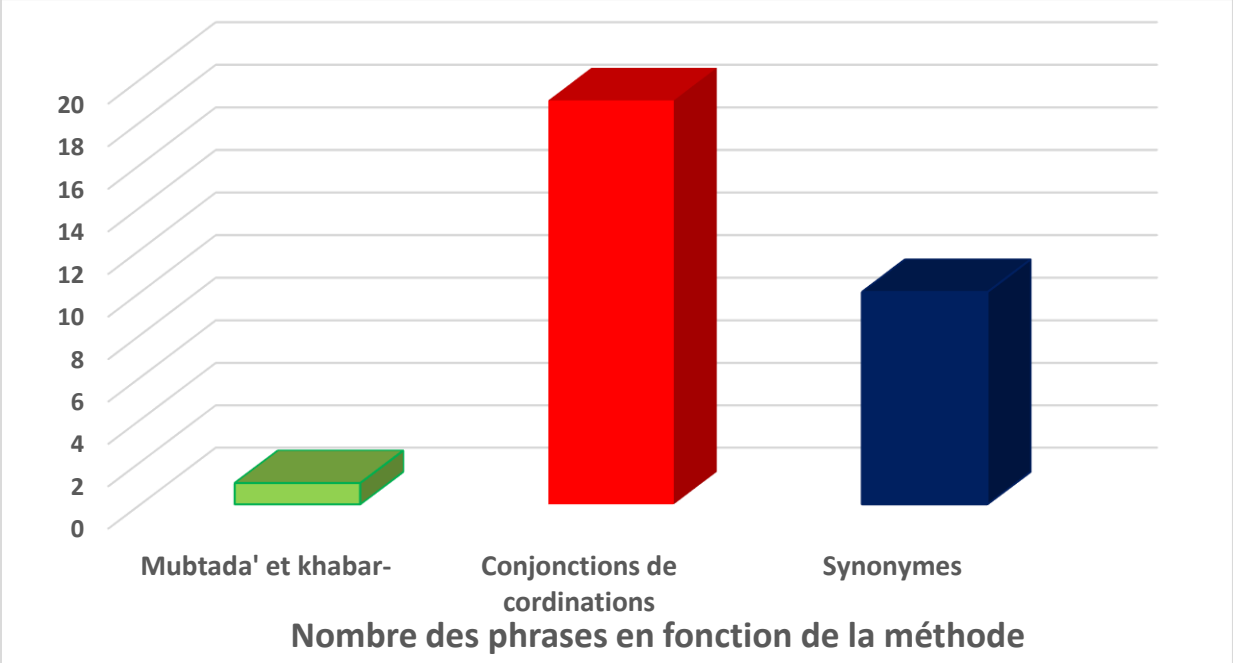
مخاطرالتلوث

•ظهور المشاكل البيئية **المتنوعة** ومن **بينها** الانفجار السكاني
•المطر الحمضي
•**انقراض بعض مظاهر الحياة الحيوانية و النباتية واختلال التنوع البيولوجي**
•تأكل طبقة الأوزون
• ظاهرة الاحتباس الحراري
•**ظاهرة فقر التربة الزراعية والتصحر**
•تعرض المجال الجوي للمطارات للتلوث الجوي مما يؤدي الى خفض مجال الرؤية
•**عدم استقرار المناخ و الانقلابات الحرارية**.
إلحاق أضرار بالآثار فالتركيزات العالية من اكاسيد **الكبريت و الرصاص** تعمل على تآكل ألوان الآثار على مدار آلاف السنين وذلك لقدرة تلك الاكاسيد على التفاعل مع مكونات تلك الألوان
• .حدوث الحرائق عن طريق الاشتعال الذاتي للغازات السامة القابلة للاشتعال
نسب متزايدة من الاكاسيد الضارة والمعادن الثقيلة العالقة بالهواء وخاصة الرصاص الذي ساهم بها **قطاع توليد الكهرباء و صهر المعادن والكيماويات ومصانع البلاستك**
•**صعوبة** تنقية مياه الصرف الصحي
•بقاء المكونات الصناعية بالتربة الزراعية لفترة طويلة من الزمن
•تقليص مساحات الأراضي الزراعية لمقابلة الغزو الصناعي
•تزايد نشر الرطوبة الجوية بالهواء بكثرة المسطحات المائية لصرف المخلفات الصناعية
زيادة التدفق الحراري الاتية من المناطق الصناعية والمحملة بالملوثات المختلفة من التربة و العوالق و الدخان

-: علاج التلوث

• .الوعي الذاتي لدى الشخص لان التلوث ينذر بفنائه
• وقف تراخيص مزاولة النشاط الصناعي الذي يدمر البيئة
•تهجير الصناعات الملوثة للبيئة بعيدا عن اماكن المدن
•تطور اساليب مكافحة تلوث الهواء
تطوير وسائل التخلص من **النفايات و القمامة** وخاصة عمليات حرق النفايات في الهواء الطلق
•
• القيام بعمليات التشجير على نطاق واسع للتخلص من ملوثات الهواء وامتصاصها.
الكشف الدوري للسيارات ومراقبة عوادمها
•**استعمال** الغاز الطبيعي كإحدى مصادر الطاقة البديلة عن مصادر الطاقة الحرارية
• معالجة التلوث النفطي بإضافة مذيبات كيمياوية لترسيبه في قاع لمياه
اقامة المحميات البحرية التي تشتمل على كائنات بحرية نادرة مهددة بالانقراض.
•. اللجوء الى استخدام **الموارد الطبيعية و المبيدات العضوية** والابتعاد عن المبيدات الكيمياوية

Le graphe ci-dessous montre le nombre des phrases utilisées dans chaque méthode pour cacher l'information dans le texte.



Bibliographie

[1] Jdn. URL http://www.journaldunet.com/encyclopedie/definition/389/32/20/watermarking.shtml. [Accés le 03 mai 2016]. 1

[2] Yoan Miche. "developing fast machine learning techniques with applications to steganalysis problems", institut national polytechnique de grenoble, - inpg, 2010. 5

[3] A. Belgoraf A. M'hamed A.Ali-Pacha, N.Hadj-Said. *"Stéganographie : Sécurité par Dissimulation", Vol16 n° 01.* 2006, Université des Sciences et de la Technologie d'Oran –USTO, Institut National des Télécommunications Evry- Paris. 5, 6, 8, 11, 18, 19, 22

[4] M. G. Kuhn F.A.Petitcolas, R.J. Anderson. "information hiding-a survey". Proceedings of the IEEE (USA), 1999. 6

[5] M.L.Bensaad. "steganography and digital watermarking". Thèse de Doctorat, Université amar telidji, Laghouat - Algeria, 2013-2014. 6, 10, 12

[6] Initiation au marquage d'image, . URL http://www.picsi.org/parcours_impression_17.html. [Accés le 2 février 2016]. 6

[7] Jean-François Audenard. Le tatouage des données numériques. URL http://www.orange-business.com/fr/blogs/securite/webtech/le-tatouage-de-donnees-numeriques. Publié le 11 Août 2015, [Accés le 5 mai 2016]. 6

[8] A.Bloom J.Fridrich J.Cox, L.Miller and T.Kalker. "digital watermarking and steganography ". Morgan Kaufmann Publ ishers , USA, 2008. 7

[9] Philippe Nguyen. *"La sécurité à l'ère numérique"*. URL http://www.cairn.info/revue-les-cahiers-du-numerique-2003-3-page-135.html. [Accés le 03 mai 2016]. 7

[10] Bouderbala Mohamed. Implimentation d'un algorithme de tatouage vidéo robuste dans le domaine comprésé. Master's thesis. URL http://bu.umc.edu.dz/theses/electronique/BOU5254.pdf. Univérsité Mentori constantine. 7

[11] Gregory Kipper. *"Investigator's Guide to Steganography", 2003, Publications,1st edition.* 9

[12] La stéganographie au cours des siècles. URL http://www.bibmath.net/crypto/index.php?action=affiche&quoi=stegano/histstegano. [Accés le 02 février 2016]. 9, 10, 11

[13] Muhammad alBukhari. *"Sahih al-Bukhari", Number 3007.* 10

[14] Bennoui Issam. "l'utilisation de la stéganographie pour la sécurité des données". Thèse du Master, Université Hadj-Lakhder Batna, 2012. 11, 15

[15] G Labouret. *"Introduction à la cryptographie".* 24-12-1999, Version Internationale 6.5.1. 12

[16] A. Chedad. *"Steganoflage : A new image steganography algorithm. Thèse de doctorat".* PhD thesis. University of Ulster, UK, 2009. 13

[17] Saint Marcel Frédéric. "stéganographie vs tatouage". Technical report. URL http://lig-membres.imag.fr/donsez/ujf/easrr0203/tatouagestegano/tatouagestegano.pdf. 13

[18] C.Fontaine F.Raynal, F.A.P.petitcolas. *"Stéganographie :visite de l'univers numérique pour dissimuler des informations".* 24 fevrier 2002. 15

[19] Les langages secrets dans l'antiquité gréco-romaine., . URL http://bcs.fltr.ucl.ac.be/FE/08/stegano.htm. [Accés le 15 février 2016]. 15

[20] Bénoni Martin. "stéganographie : techniques". Octobre 2007. 17, 20, 21

[21] Jacques Fontanille. "marquage d'image et stéganographie". URL http://www.picsi.org/fiche_64.html. [Accès le 24 février 2016]. 17, 18

[22] Cheng Yao. *"Stéganographie".* URL http://perso.telecom-paristech.fr/~tupin/CONF/Steganographie.pdf. 17

[23] "la stéganographie : comment cacher un message secret ". URL http://www.funinformatique.com/la-steganographie-comment-cacher-un-message-secret/. [Accès le 17 février 2016]. 18

[24] T.Raggo C.H.Michael. *"Data Hiding : Exposing Concealed Data in Multimedia, Operating Systems, Mobile Devices and Network Protocols"*. 23 November 2012. 19

[25] Les langues parlées dans le monde. URL `http://www.statistiques-mondiales.com/langues.htm`. [Accés le 15 avril 2016]. 23

[26] Ethnologue : languages of the world, seventeenth edition. dallas, texas : sil international. URL `http://www.ethnologue.com/statistics/size`. [Accés le 7 mars 2016]. 23

[27] Yasser M Alginahi. "a survey on arabic character segmentation.". . International Journal on Document Analysis and Recognition (IJDAR) ,Vol.16, No.2 ,pp.105-126., 2013. 23

[28] Venu Govindaraju Lorigo, Liana M. "offline arabic handwriting recognition : a survey.". Pattern Analysis and Machine Intelligence, IEEE Transactions ,Vol.28, No.5, pp. 712-724, 2006. 23

[29] Sameh M. Awaideh Abdul-Rahman M. Elshafei Adnan Gutub Aabed, Mohammed. "arabic diacritics based steganography". IEEE International Conference on Signal Processing and Communications, pp.756-759, Dubai, UAE, 24-27, Nov 2007. 24, 39, 42, 48, 49, 53

[30] Écriture arabe. URL `http://www.cairn.info/revue-document-numerique-2002-3-page-155.htm`. [Accés le 4 mars 2016]. 24

[31] Steven Low Nicholas F. Maxemchuk Lawrence O.Gorman Brassil, Jack T. "electronic marking and identification techniques to discourage document copying". Selected Areas in Communications, IEEE Journal, vol.13, pp.1495-1504, 1995. 26, 27

[32] N. Morimoto W. Bender, D. Gruhl and A. Lu. "techniques for data hiding". IBM Systems Journal, Vol. 35, No 4, pp. 313-336,1996. 26, 28

[33] S. Shahreza M. Shahreza. "steganography in tex documents". Proceedings of Intelligent System and Knowledge Engineering, ISKE 2008. 3rd International Conference, Nov. 2008. 26, 28, 29

[34] Shirali-Shahreza. "text steganography by changing words spelling". . ICACT 2008. 10th International Conference 17-20 Feb. 2008. 26, 29

[35] Wang.Xiaofeng. "digital watermarking research based on text". Third International Conference on Information Science and Technology, pp.433 - 436, 2013. 29

[36] Mercan Topkara Mikhail J. Atallah Topkara, Umut. "the hiding virtues of ambiguity : quantifiably resilient watermarking of natural language text through synonym substitutions". Proceedings of the 8th workshop on Multimedia and security, pp. 164-174. ACM, 2006. 30

[37] Ibrahim Kamel. "hiding information in the placement of maneuverable words". Innovations in Information Technology (IIT), International Conference, 2012. 30

[38] Lamiaa A. Elrefaei Reem Ahmed Alotaibi. "arabic text watermarking : A review". International Journal of Artificial Intelligence and Applications (IJAIA) Vol. 6, No. 4, July 2015. 31, 38

[39] Mohammad Shirali-Shahreza Shirali-Shahreza, M. Hassan. "a new approach to persian/arabic text steganography". . In Computer and Information Science, 2006 and 2006 1st IEEE/ACIS International Workshop on Component-Based Software Engineering, Software Architecture and Reuse. ICIS-COMSAR. 5th IEEE/ACIS International Conference ,pp. 310-315. IEEE, 2006. 32, 43, 47, 48, 50, 52

[40] Mohammad Shirali-Shahreza. "a new persian/arabic text steganography using “la” word”. . In Advances in Computer and Information Sciences and Engineering, pp. 339-342. Springer Netherlands,2008. 33, 47, 48, 50, 52

[41] M. Hassan Shirali-Shahreza Shirali-Shahreza, Mohammad. "an improved version of persian/arabic text steganography using" la" word.". . Telecommunication Technologies 2008 and 2008 2nd Malaysia Conference on Photonics. NCTT-MCP 2008. 6th National Conference on. IEEE, pp.372-376, 2008. 33, 47, 48, 50, 52

[42] Mohammad Shirali-Shahreza. "pseudo-space persian/arabic text steganography". . In Computers and Communications, 2008. ISCC 2008. Symposium on IEEE, pp. 864-868, 2008. 34, 35, 47, 48, 50, 52

[43] M. Hassan Shirali-Shahreza Shirali-Shahreza, Mohammad. "arabic/persian text steganography utilizing similar letters with different codes". . The Arabian Journal For Science And Engineering, vol.35, pp.213-222, 2010. 35, 47, 48, 50, 52

[44] R. Mahmod N. A. Roslan and N. Udzir. "sharp-edges method in arabic text steganography". Journal of Theoretical and Applied Information Technology, Vol. 33, No. 1, November 2011, pp. 32 – 41. 36, 48, 51

[45] Esam Ali Khan. "using arabic poetry system for steganography". Asian Journal of Computer Science And Information Technology 4 : 6 (2014) 55 - 61, 2014. 36, 37, 47, 48, 50, 52

[46] NUR IZURA UDZIR ZURIATI AHMAD ZURKARNAIN NUUR ALIFAH ROSLAN, RAMLAN MAHMOD. "primitive structural method for high capacity text steganography". Journal of Theoretical and Applied Information Technology Vol. 67 No, 20 th September 2014year. 37

[47] Yousef Elarian Sameh Awaideh Aleem Alvi Gutub, Adnan. "arabic text steganography using multiple diacritics". . In 5th IEEE International Workshop on Signal Processing its Applications-WoSPA, 2008. 40, 43, 44, 48, 49, 53

[48] Mohamed Bachir Yagoubi Bensaad, Mohamed Lahcen. "high capacity diacritics-based method for information hiding in arabic text". Innovations in Information Technology (IIT), International Conference on. IEEE.,2013. 40, 41, 42, 47, 48, 49, 53

[49] Mujtaba S. Memon Shah, Asadullah. "a novel text steganography technique to arabic language using reverse fatha". Pakistan Journal of Engineering Technology et Science (PJETS), vol.1, pp.106-113, 2011. 40, 53

[50] Esraa Mohammad Ahmadoh. "utilization of two diacritics for arabic text steganography to enhance performance". Lecture Notes on Information Theory Vol. 3, No. 1, June 2015. 42, 48, 49, 53

[51] Adnan Gutub Khalid Al-Kahsah Jameel Hamodi Al-Haidari, Fahd. "improving security and capacity for arabic text steganography using 'kashida'extensions". Computer Systems and Applications. AICCSA 2009. International Conference on IEEE/ACS ,pp. 396-399, 2009. 44, 49, 54

[52] Fahd Al-Haidari-Khalid M. Al-Kahsah Jamil Hamodi Gutub, Adnan Abdul-Aziz. "e-text watermarking : Utilizing 'kashida' extensions in arabic language electronic writing". . Journal of Emerging Technologies in Web Intelligence, vol.2, pp. 48-55., 2010. 44, 49, 54

[53] Khaled Elleithy Odeh, Ammar. "steganography in arabic text using zero width and kashidha letters". . International Journal of Computer Science and Information Technology (IJCSIT),vol. 4,pp. 1-11, 2012. 45, 54

[54] Muhammad N. Kabir-Omar Tayan Alginahi, Yasser M. "an enhanced kashida-based watermarking approach for arabic text-documents". . In Electronics, Computer and Computation (ICECCO), 2013 International Conference on IEEE , pp. 301-304, 2013. 45, 49, 54

[55] Ammar Odeh. "steganography in arabic text using kashida variation algorithm (kva)". . Systems, Applications and Technology Conference (LISAT), 2013 IEEE Long Island , 2013. 46, 54

[56] Razan Alamoudi-Noura Almuzaini Samira Mersal, Safiah Alhazmi. "arabic text steganography in smartphone". International Journal of Computer and Information Technology (ISSN : 2279 – 0764) Volume 03 – Issue 02, March 2014. 48, 51, 52

[57] Manal Fattani Gutub, Adnan. "a novel arabic text steganography method using letter points and extensions". . n WASET International Conference on Computer, Information and Systems Science and Engineering (ICCISSE), Vienna, Austria ,pp. 28-31, 2007. 49, 54

[58] Mohamed Ibn Abed-Allah Ibn Malek Mohamed Abde-Alaziz Abed-Allah. Alfiat Ibn malek, 1991. 55, 56

[59] Salim et khayrani. *Chareh el-Ajromia*. 2009. 55, 56

[60] Dhawahiria. Manhadj said al afghani fi chareh nadaria nahwia. Thése de Doctorat, Université Abou Elkacem Saad Allah d'Alger 2, 2009. 55

[61] Faysal Ibn Abdelaziz Ali Moubarak. *lobab al-iraab fi taysiir nahou li tolab*. 55

[62] hadif elmobtada wa khabar jawazan, . URL `http://www.alukah.net/literature_language/0/95575/`. [Accés le 9 avril 2016]. 55

[63] hadif elmobtada awa khabar jawazan, . URL `https://www.youtube.com/watch?v=RkN-tw5koPE`. [Accés le 11 avril 2016]. 55

[64] Apprendre les langues arabe et française. URL `http://arabeclassique.forumactif.com/t2204-la-phrase-nominale-arabe-le-mubtada-et-le-khabar`. [Accés le 7 avril 2016]. 56