Arabic stego_system based on Arabic Language structure and pronunciation

Asst. Lecturer Ismael Abdulsattar*

Lecturer Naji M.Sahib**

Abstract

Arabic language is the one of many languages that uses Harakat (damma, kas-ra, fat-ha, sukoon, tanween) or short vowel marks for the correct pronunciation that by his turn will affect the meaning of the word. In this language each word has deferent meaning based on the Harakat that can change the pronunciation. Harakat can be used on a single alphabet of these languages. We can write without them but the only the Professional people can interpret the correct meaning of the sentence. Thus in this paper we code each one of these Harakat and then we use the sequence of this Harakat to obtain the correct sequence of the secret message and developing the mechanism that can select the word with the proper Harakat that's gives proper meaning based on huge database and provides mechanism also for high retrieval based on the requirement of the secret message.

Key words: *Arabic language, coding, Steganography, information retrieval***.**

* University of Al-Mustamsiriyah

** Diyala University

1. Introduction

Steganography is the art of covered or hidden writing, the purpose of Steganography is covert communication to hide a message from a third party this differs from cryptography the art of secret writing which is intended to make a message unreadable by a third party but does not hide the existence of the secret communication. Although Steganography is separate and distinct from cryptography, there are many analogies between the two, and some authors categorize Steganography as a form of cryptography since hidden communication is form of secret writing [1][2].

The Steganography process generally involves placing a hidden message in some transport medium, called the carrier, the secret message is embedded in the carrier to form Steganography medium. The use of a Steganography the use of Steganography key may be employed in hiding mechanism also as show in the equation (1)

$$steganography_{medium} = hidden_{message} + carrier + steganography_{key} \quad \text{... (1)}$$

As an increasing amount of data is stored on computers and transmitted over network, it is not surprising that Steganography has entered the digital age [3]. On computer and networks, Steganography applications allow for someone to hide any type of binary file in any other binary file. As well as there are several types of digital media that could use as cover and secret such as (Image, Video, Text, audio...Etc) and must care about requirement criteria that can fulfill user demand of security such as capacity and robustness and perceptual.

2. Affection of Harakat on the pronunciation and meaning

In Arabic there are three kinds of vowels which is three vowel letters, which are (Alef, Waw and Ya') and are used for long vowels, Hamza, and the vowel marks which are used for short vowels [4][5]. They are called "Tashkeel" in Arabic they are also known as Harakat just a different name by some references [6]. To make short vowels distinct from long ones when words are read, the Arabic script uses vowel marks to do so. This is implemented by writing the marks over or under a letter. Will show how to write the letter "Seen" with the four different vowel marks applied to it as show in the table(1)

Table (1)

The Mark's Name	Applied to the letter (Seen)	Pronounced as
Fat-ha	سَـ	**Sa**
Dhamma	سُـ	**Su**
Kas-ra	سِـ	**Si**
Sukoon	سْـ	**S**

Now will show how Vowel marks are very important in Arabic. The word's meaning may differ completely if a vowel mark applied to one letter is changed. Let have a look at these examples in the figure (1)

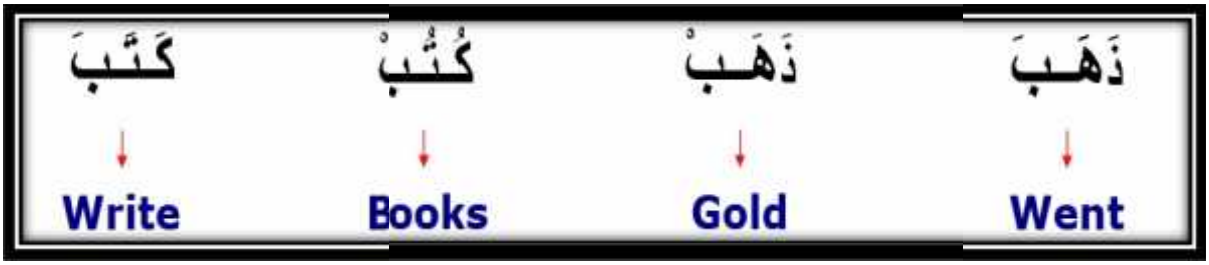

Figure (1) shows the affection of vowel marks on the meaning

3. Proposed Hiding Mechanism

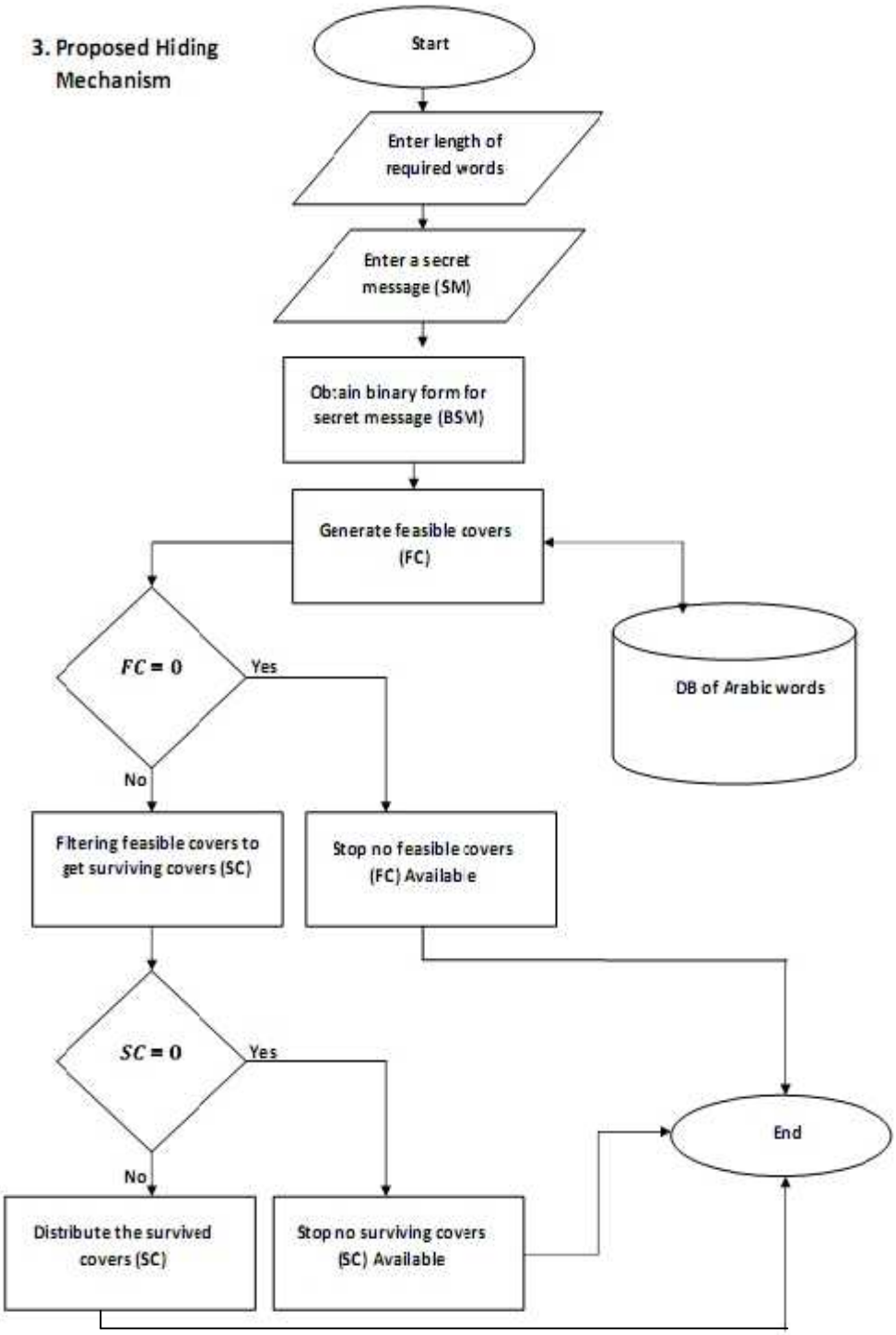

Table (2)

INDEX	IMAGE	CODE
1	ـُـ	0
2	ـَـ	1
3	ـِـ	01
4	ـٗـ	10
5	ـّـ	00
6	ـٌـ	11
7	ـّـ	100
8	ـّـِ	101
9	ـٌـ	110
10	ـًـ	111
11	ـًـ	1000
12	~	1001
13	ـّـًـ	1010

3.1 The proposed coding table

In Arabic language there are several Harakat that could give several varieties of binary bits which can form the secret message with high possibilities our proposed coding table as shown in the table (2) contain thirteen Harakat and each one of them assigned unique value with different length and that table build based on statistics study over appearance of Harakat.

Proposed Coding table provide verity of code length index one and two have length one from tree till six have length two seven to ten have length three and eleven to thirteen have length four. And the interference among these values will give multiple options to select best one based on the sequence that can give proper meaning.

4. Result

Will show in the result section some of the obtained result using our proposed algorithm in several stage first in short sentences as show in the figure below

Secret message	1110110000100111111	1100101011011001O1
Word cover length	{3,4,5}	{4,5}
Stego_object	ذهب محمد مسرعا	ينزل المطر من السماء
Retrieve Format	ذَهَبَ مُحَمَّدٌ مُسْرِعًا	يَنْزِلُ الْمَطَرُ مِنَ السَّمَاءِ

Second test done over multi line cover to hid long secret message As show in the following figure

Secret message	1001111001011100101000011001111010011011111100111
Word cover length	{4,5}
Stego_object	و اللسان آفتان عظيمتان ، إن خلص من إحداهما لم يخلص من الأخرى آفة الم ، و آفة السكوت ، و قد يكون كل منهما أعظم إثما من الأخرى في وقتها
Retrieve Format	وَلِسانِ آفَتَانِ عَظِيمَتَانِ ، إِنْ خَلَصَ مِنْ إِحداهما لَمْ يَخْلُصْ مِنَ الأُخْرَى آفَةُ الْكَلامِ ، وَ آفَةُ السكوتِ ، وَ قَدْ يَكُونُ كُلُّ مِنْهُمَا أَعْظَمُ إِثْمًا مِنَ الأُخْرَى فِي وَقْتِهَا

It can be seen that the length of the cover words will have great effect over hiding mechanism that increase the number of possible cover with various length .

5. Conclusion

In this approach the distortion will be zero the size of the selected cover will remain as it is without any change and selecting the length of the cover word from one hand will consider as constrain over hiding capacity but from other hand will increase security level reduce the suspicion level one of the drawbacks of this algorithm it is not easy to form the words to give meaningful statement.

References

[1] Cox I. J., et al. (2008) "**Digital Watermarking and Steganography**", 2nd Edition, Morgan Kaufmann Publishers, Elsevier Inc.

[2] Shirali-Shahreza M. H. and Shirali-Shahreza M., (2006) "**A New Approach to Persian/Arabic Text Steganography"**, 5th IEEE/ACIS International Conference on Computer and Information Science, 310- 315.

[3] Gutab A. and Fatani M. (2007) "A **Novel Arabic Text Steganography Method Using Letter Points and Extensions**", World Academy of Science, Engineering and Technology, 27, 28-31.

[4] Bennett K. (2004) "**Linguistic Steganography: Survey, Analysis and Robustness Concerns for Hiding Information in Tex**t", Purdue University, CERIAS Tech.

[5] Gopalan K., (2003) "**Audio Steganography using bit modification**", Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2, 421-424.

[6] Doerr G. and Dugelay J. L., (2003) "**Ä Guide Tour of Video Watermarking**", Signal Processing: Image Communication, 18, 263-282.

نظام اخفاء عربي بالاعتماد على قواعد وتلفظ اللغة العربية

م .ناجي مطر سحيب* م .م.اسماعيل عبد الستار جبار**

المستخلص

اللغة العربية هي واحدة من اللغات التي تستخدم الحركات (الضمه, الفتحه, الكسره ,السكون ,التنوين) او مايسطلح على تسميتها (short vowel marks) للحصول على نطق صحيح الذي بدوره يؤثر على معنى الكلمة حيث كل كلمة في هذه اللغة تعطي عدة معاني بالاعتماد على الحركات التي تغير النطق. الحركات يمكن ان تستخدم على كل حرف من الكلمة يمكن الكتابة بدونها ولاكن فقط المحترفين يمكنهم تميز المعنى الصحيح للجملة. هكذا في هذه الورقة البحثية تم ترميز كل واحدة من حركات اللغة العربية بحيث تستخدم سلسلة تتابع هذه الحركات للحصول على تتابع صحيح من الرسالة السرية وتطوير ميكانيكية الحصول على الكلمات المناسبة التي تحوي على الحركات المطلوبة والحصول على المعنى المناسب بالاعتماد على قاعدة بيانات كبيرة.مع تزويدها بميكانيكية عالية الاسترجاع بالاعتماد على الرسالة السرية.

الكلمات المفتاحية: اللغة العربية, الترميز, الاخفاء, الاسترجاع المعلومات.

* جامعة ديالى
** الجامعة المستنصرية