ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

Proposed Approach for Steganography in Arabic Text Basedon B+_Tree,DNA Coding and Arabic Diacritics

Asst.Instructor. Estabraq Abdulredaa kadhim, Asst.Prof. Hala Bahjat AbdulWahab, Asst.Prof. Suhad Malallah Kadhem

Al-ESRA'A University College University of technology, Baghdad/Iraq

Manuscript Info

Manuscript History:

Received: 15 October 2014
Final Accepted: 22 November 2014
Published Online: December 2014

Key words:
Bovine mastitis, teat dip, Resistance, control

Corresponding Author

Estabraq Abdulredaa kadhim

Abstract

Steganography is the art of covering or hiding writing, the purpose of steganography is covert communication to hide a message from a third party. The aim of this paper is to introduce an efficient and strong approach for hiding secret message within diacritical Arabic Text by using B+_tree as a tool for compress the secret message and DNA nucleotides as method for coding and Arabic diacritics as a cover text for steganography. The proposed approach is implemented on many diacritical Arabic text and the results are tested according to the authorized measures that are used in this field. And comparison is performed between the results that are obtained in this work with results that are obtained from other works in this field. The proposed approach has achieved a high capacity ratio for steganography form (79% to 106%)and also provides good security (transparency) for steganography based on some of similarity measures from (0.791 to0.9024).

Introduction

One of the newest hot spots in security research is steganography in Arabic text. Arabic text has many appropriate features for data embedding. Arabic writing is very rich in diacritic marks, which the structure prefers in steganographic applications. Arabic diacritic marks represent efficient carriers to hide information into plaintext

B+ tree usually uses as a special dictionary for storing the secret massages (with their codes) in a manner that prevent redundancy of these massages or even sub massages in this dictionary (in order to provide efficient memory usage). So the proposed method includes two stages :(Store the secret message in this dictionary (if it is not found) and get its unique code (at send process)) and (retrieve the unique secret message when we have its code from this dictionary (at received process) [2].In this proposed method, we will use B+ tree as a tool for compression the secret message. And DNA nucleotides as method for coding, and Arabic diacritics as cover text for steganography

1. Steganography in Arabic Text

Text is one of the oldest media used in steganography; well before the electronic age, letters, books, and telegrams hid secret messages within their texts. The wealth of electronic textual information available as well as the difficulty of serious linguistic analysis makes this an interesting medium for steganographic information hiding [3].Soft-copy text is in many ways the most difficult place to hide data. This is due largely to the relative lack of redundant information in a text file as compared with a picture or a sound, while it is often possible to make imperceptible modifications to a picture, even an extra letter or period in text may be noticed by a casual reader **.**Steganography in text is an exercise in the discovery of modifications that are not noticed by reader[4].

Arabic text has many appropriate features for data embedding. Arabic language uses different symbols as diacritical marks, or simply diacritics which are also known as Harakat. The main reason to use these symbols is to distinguish between words that have same letters. It depends on Arabic Diacritics (Harakat), where diacritics are optional. Most of Arabic novels can be read without Diacritics, which depends on the language's grammar. Arabic diacritic marks represent efficient carriers to hide information within text. Also Arabic alphabets have letters, some of which are without dots (unpointed letters), others are with dots with different positions of dots (pointed letters) [5].

2. B+ Tree

B+ Tree is a variation of B-Trees a structure of nodes linked by pointers is anchored by a special node called the root, and bounded by leaves has a unique path to each leaf, and all paths are equal length stores keys only at leaves, and stores reference values in other, internal, nodes guides key search, via the reference values, from the root to the leaves. B+ tree is called an index to database, such that each record will be stored in the database, the reference number (and the key) of that record will be stored in the B+ tree. So when we want to reach a certain record, we need to know its key to get its reference number from the B+ tree. When we get the reference number of that record we can retrieve the required record directly. B+ tree is an arranged and balanced tree, and this is why it is so fast in retrieving the required data [2].

B+-trees distinguish internal and leaf nodes, keeping data only at the leaves, whereas ordinary B-trees would also store keys in the interior. B+-tree insertion, therefore, requires managing the interior node reference values in addition to simply finding a spot for the data, as in the simpler B-tree algorithm [6].

3. The Proposed Approach

Steganography Approach is consisting of preprocessing and embedding processes. Figure (1) illustrates the main structure of the proposed Steganography approach:

3.1. Compression (coding) the secret message Stage

Compression stage represents the first stage in the proposed approach that aims to convert the secret massage to small codes numbers by investigating from the compression feature that is available in B+ tree structure (indexing structure), that is illustrated in [2] in a manner that prevents redundancy of these messages or even sub messages in order to provide efficient memory usage. This stage is represented by store of the secret message and getting its unique code based on B+ tree indexing, in order to reach unique codes of the secret message. One of the most important parameters in this stage is represented by **Code-Counter (n).** This parameter refers to total number of secret messages that are store in dictionary. Algorithm (1) will illustrate the main steps for getting list of code of secret message that consist of one or more sentences. Algorithm (2) will illustrate the main steps for retrieving secret message from list of code

1.2. Preprocessing for Compress Codes

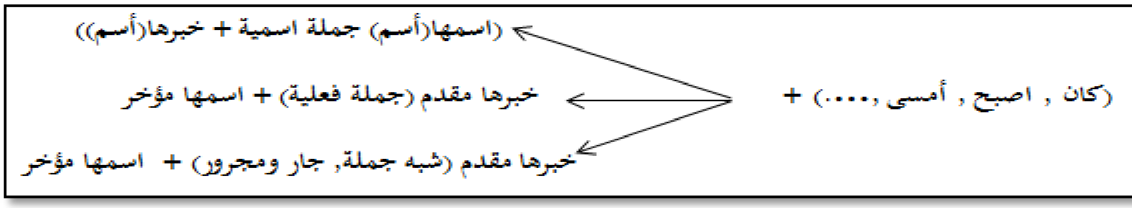
This stage is considered as a primary treatment for embedding stage. New method will be proposed for this purpose based on using DNA for coding requirement and then mapping to Arabic diacritics to be more suitable for embedding stage.

- **DNA Coding**: The preprocessing for encrypted code based on DNA coding is represent by converting the encrypted points into DNA bases that are four (A, T, C or G) instead of 0 or 1 to increase the probability space used for the diacritics predication in front of the eavesdropper. Table (1) illustrates the mapping from point number into DNA strand.
- **Arabic Diacritics:** Arabic diacritics consist of nine diacritics as shown in Table (2). The proposed approach selects most four popular Arabic diacritics that are used in Arabic language as Diacritic Bases and another one as Control diacritic**.**
    - **a. Diacritic Bases (DB):** After encrypted points passing through DNA coding process, each DNA strand will be converted into four Arabic diacritics by mapping each nucleotide into corresponding one Arabic diacritic as shown in Table (3). Table (4) illustrates the example for final mapping into secret diacritics. According to prime (251), (4^4) is able to encode numbers from 0 to 255. If prime number is increased then the length of DNA strand is also increased with same bases for example when prime = 271 DNA length = 4^5 .
    - **b. Control diacritic (CD) :** The role of this diacritic will be explained in embedding stage.

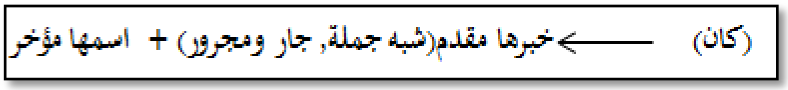
1.3. Preprocessing for Cover Text

Preprocessing for cover text is represented by extract of the Stego Key from cover text. Stego Key consists of two keys, primary key called Start-StegoKey (SSK) and Secondary key called End –StegoKey (ESK), (i.e. the secret diacritics will be embedded in the area for the cover text that is determined based on SSK and ESK in the cover text).

- **Start-Stego Key** (SSK): SSK represents a particular Arabic grammar rule such as (Kan Wa akawatiha) .This grammar rule always has a fixed forma in any Arabic sentences for example , some of (Kan Wa akawatiha) grammar rules are :



SSK in this proposed approach is represented following rule:-



Both sender and receiver will agree on the grammar rule and verify appearance in the selected cover text .In this work, the proposed steganography approach will depend on the first appearance of the grammar rule in cover text.

- **End –Stego Key (ESK):** ESK is represent by one of unused DNA strand, by investigating from the feature that is represented by DNA strand which is generated by (4^n) DNA strand and is greater than prime number. For this reason unused DNA strands will appear. Therefore ESK depends on the unused DNA strand, that is described in Table (1) and Table (4). i.e. there are five DNA strand number not used, one of them will be selected as ESK by agreement between the sender and receiver. Algorithm (3) illustrates the main steps of preprocessing for encrypt code and cover text.

1.4. Embedding Process

The proposed embedding method that will be produced aims to investigate from Arabic texts that usually appear with fully diacritics as cover texts such as sacred texts and poems. The following steps illustrate the embedding procedure:

- The first diacritic of the secret diacritics list is compared with the first diacritic in the cover diacritics list after SSK appears (grammar rule). For example, if the first secret diacritic is Fatha and the first diacritic in cover after Start- Stego Key is a 'Fatha', then diacritic is kept on the cover media and an index for both the secret diacritics and the cover media is incremented. If, however, the first diacritic in cover after grammar rule is not a 'Fatha' then it is removed from the cover media and the index for the cover media is incremented to explore the next diacritic. This process is repeated until the next 'Fatha' is found. Embedding process is continued until secret diacritic list is empty. Algorithms (4) illustrates the main steps of embedding stage

  Note: The extracting process for the secret diacritic is performed in the same manner but with reverse order. Algorithm (5) illustrates the main steps of Extracting stage.

  - **Control Diacritic** (CD) is an important tool in the proposed embedding process. It's represented by any Arabic diacritic (except Diacritic Bases). It separates between one strand of Diacritic and another. The main purpose for using the control diacritic through embedding process is in order to disperse the attention from the embedding area and to increase the transparency. Practically, if Control Diacritic is not used, some distortion may appear in entire cover text by there are intensive diacritics in some area and others not. Control diacritic that is relied upon in this work is Shadaa(ّ).

2. Implementation of the proposed approach

List of compressed code (B+_Tree codes) = {40, 234, 39, 167,8, 228, 126,174, 181, 231}
List of DNA Strands =["AGCT","GGCG","AGTA","CGTA", "ATCT", "GGTT", "TAGG", "CGGG","CATC","GGTA"]

List of Secret Arabic diacritics =

[◌ُ◌َ◌ْ◌ْ◌ِ◌َ◌ُ◌ِ◌ْ◌ْ◌ْ◌ِ◌ْ◌ْ◌ُ◌َ◌َ◌َ◌ْ◌ْ◌َ◌ِ◌َ◌ُ◌ُ◌َ◌ْ◌ِ◌ُ◌َ◌ْ◌ُ◌ْ◌ِ◌ْ◌ْ◌َ◌ِ◌ْ◌ُ]

3. Test the Experimental Results

This section will implement the most popular measurements for steganography in text in order to test experimental results of the proposed approach for each stage and perform comparison between the results that are obtained from the proposed approach that is produced in this thesis and the results that obtained from previous works in steganography in text.

There are mainly three aspects that should be taken into account when testing the results of the proposed method of steganography in text. They are security, capacity and robustness. The following sections illustrates the details of results of each aspect.

6. Measure of Capacity or Data Payload (byte/byte)

This section tests the steganography system with different cover text file sizes in terms of capacity.

1. Table (5) illustrates the results of capacity ratio for the proposed steganography approach that was tested according to the following equation :
   Capacity ratio = (amount of hidden bytes) / (size of the cover text in bytes) , and Table (6) clarify capacity ratio of other previous approaches . Tables (5) clarifies the robust of this proposed approach in term of capacity by providing high capacity ratio when compared with the capacity of other approaches of steganography in Arabic text that shown in Table (6). This proposed approach success in embedding large volume of information in small amount of text by using an efficient approach for text compression/ encoding that has a mighty effect on increasing the capacity.

4. Undetectability or Perceptual Transparency (Security)

This section shows the results of the proposed approach steganography in term of security. To maintain transparency and for better understanding and visual substantiation of test results, evaluation of the output of stego text will done the measurements of similarity. Measurements of similarity involve:

1. Computing Jaro-Winkler distance
2. Computing the Demo Levenshtein (Edit Distance)

7.1. The Jaro-Winkler Distance

This section implement the Jaro-Winkler Distance similarity measure by comparing the similarity between a cover text and the stego text, the Jaro-Winkler distance for measuring similarity between two strings has been computed. Table (7) illustrates the Jaro score for six covers and stego covers using the proposed approach and Table (8) illustrates the Jaro for another steganography approach within English text.

The similarity of each test in this proposed approach refers to that cover text and stego text is almost identical. Applying of Jaro approach was done on diacritics of cover and diacritics of stego because Arabic letters are absolutely don’t affected(i.e. Letters of the words remain the same after embedding process).

7.2. Damerau-Levenshtein Distance

In computer science theory, the Damerau-Levenshtein distance is a distance between two strings. In this measure, calculating the minimum number of operations is necessary to convert a string to another, a transaction is defined as the insertion, deletion, or substitution of a single character, or as a transposition of two characters. Table (9) illustrates the differences between cover and its stegotext , and the percentage of each test (i.e. a large percentage means the strings are very different; a small distance means the strings are very similar). Damerau-Levenshtein distance has been achieved in this proposed approach into novel scores when compared with another approaches that are shown in Table (10).

Algorithm (1): Compress the secret message and get list of unique code

Input: Secret message (sentences)
Output: List of compressed code, Code counter (n).

Process:
Begin
Step 1: Split the secret message into a set of sentences
Step2: For each sentence do the following
If(Sentence suppose as a new sentence),**then** do the following
Put the first word of the sentence as a key in b+ tree (Bt1),
Compute the length of the sentence
Give the sentence a new unique code
And use this code as a key for b+ tree (Bt2)
Increment Code counter (n)

If (Sentence has words that are already found in dbase but with no code),**then** do the following
1. Give it a new code
2. Store it in Bt2 as a key
3. Increment Code counter (n)

If(sentence has words that are already found in dbase except the last word) , t**hen** do the following
Store the last word in dbase and the reference of the previous word
Store at the previous word the new reference,
And give it a new code
Increment Code counter (n)

If (sentence has words that are already found in dbase)**,then** do the following
Store the remaining words in dbase and the references of the next and previous word
And give it a new code
Increment Code counter (n)

If(sentence is already found in dbase but with no code) **,then** do the following
Give it a new code
Store it in Bt2 as a key
Increment Code counter (n)

If (sentences, such that, some of their middle words are found in dbase) ,**then** do the following
Only the not found words will be store in dbase,
And we will store the first word in Bt1,
And store the new code in Bt2
Increment Code counter (n)
Step 3: End

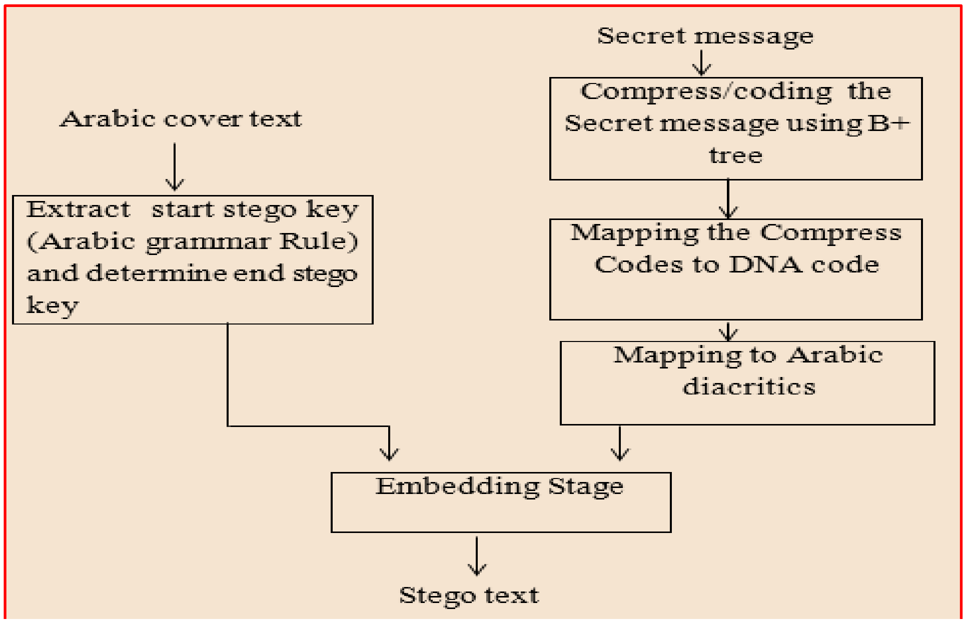


Figure (1): (Main stages of Steganography Approach)

Algorithm (2) :Retrieve the unique secret message from list of code

Input: List of compressed code
Output: Secret message

Process:
Begin
Step1: For each code in List of compressed code do the following
If(the code is found in Bt2) **Then** do the following

Retrieve the term of the last word of the message that the code refer to it
Search in the list of this term on this code
Get the length of the sentence that have the last word, and the reference of the previous word
Follow the reference of the previous word, and take its word and concatenate it with next word, then follow the reference of its previous word and so on, until we get the sentence

Step 2: End

Algorithm (3): Preprocessing for secret message and for Cover text

Input: Encrypted points, Diacritic Bases.
Output: List of Secret diacritics.

Process:
Begin
Step 1: Convert the secret points number into DNA strand that showing in Table(1)
Step 2: Each Nucleotide in DNA strand is converted into corresponding Arabic Diacritic Marks as shown in Table (3), (4) and put them in **List of Secret diacritics**
Step 3: Look up for Start- Stego Key within cover text (i.e. Arabic grammar rule)
Step 4: Select End-Stego Key (unique DNA strand) from one of unused DNA nucleotide as in Table (4)
Step 5: Append End -Stego Key to tail of List of secret diacritics
Step6: Store diacritics of cover from Start-Stego Key in **list of Cover diacritics**
Step7: End

Algorithm (4) : Embedding Process

Input: : List of Secret diacritics , Diacritic Bases, Control Diacritic, Start stego key ,DNA strand length, list of Cover diacritics, **Initiate** i = 1 ; j=1 ; k=1 ;
Output: Stego Text

Process:
Begin
Step 1: While(list of secret diacritics <> empty)
Step 2: While (k <= DNA strand length)

If (list of secret diacritics [i] = list of Cover diacritics [j] **OR** list of Cover diacritics [j] ∉ **(**Diacritic Bases)) **Then**
keep list of Cover diacritics [j] ;
i++ ; j++ ; k++
Else
Remove list of Cover diacritics [j] ; j++

Step 3:end while
Step 4:while (list of Cover diacritics [j] is not reach to control diacritic)
keep list of Cover diacritics [j] ; j++
Step 6: End while
Step 7:k=1
Step 8: Go to Step1
Step9: End while
Step10 :END

Algorithm (5) : Extracting Stage

Input: : Stego text (List of stego diacritics) , Diacritic Bases , Control Diacritic, Start stego key(SSK) Appearance of SSK , End Stego key , DNA strand length, list of Cover diacritics, Initiate i = 1 ; j=1 ; k=1
Output: : List of Secret diacritics

Process:
Begin
Step1: Find Start Stego Key and Appearance of SSK

Step2: While (k <= DNA strand length)
If ((List of stego diacritics [j]) ∈(Diacritic Bases) **Then**
List of Secret diacritics [i]=list of stego diacritics [j]
j++; i++;
Step3 : End while
Step4 : If (Last diacritics in List of Secret diacritics= End Stego Key) **Then**
Go to Step 10
Else
Go to step 5
Step5 : While (list of stego diacritics [j] is not reach to control diacritic) **Do**
j++
Step 6 : End while
Step 7 :k=1
Step 8 : Go to Step2
Step 9: End while
Step10 :END

Secret Message=
"محافظة بغداد عاصمة جمهورية العراق، وتقع على نهر دجلة وتعتبر أكبر مدينة في العراق وثاني أكبر مدينة في الوطن العربي بعد القاهرة وثاني اكبر مدينة في آسيا الغربية بعد مدينة طهران. وتعتبر بغداد نقطة التقـــاء مهمــــــه لحركــــــة الطرق والجو والقطــــارات. وتعتبر بغداد المركز الاقتصـــادي والاداري والتجاريوالتعليمي ومقر الحكومه في العراق.مثلت بغداد واحده من اهم مراكز العلم على تنوعه في العالم وملتقى للعلماء والدارسين لعدة قرون من الزمن . لمدينة بغداد القديمة اسماء عدة كالمدينة المدورة ودار السلام . ويمر في المدينة نهر دجلة ، وينصفها الى جزئين : الكرخ (الجزء الغربي) والرصافة (الجزء الشرقي.)تزخر بغداد بالكثير من المعالم التاريخية والحضارية ، واهمها المدرسة المستنصرية ، والمساجد الاسلامية القديمة ، والقصور الاثرية ، و عدد من المقامات الدينية. تتركز اهم الانشطة في محافظة بغداد في تكرير النفط والصناعات الخفيفة وانتاج المواد الكيميائية والبلاستيكية والأجهزة الكهربائية. لمدينة بغداد أربع جامعات هي جامعة بغداد والجامعة المستنصرية والجامعة التكنولوجية وجامعة ومؤسسة المعاهد الفنية. يوجد في مدينة بغداد العديد من المتاحف ومن أهمها المتحف الوطني العراقي الذي يعرض الكنوز الأثرية لحضارة بلاد ما بين النهرين. والمتحف البغدادي للفنون الشعبية وفيها العديد من المساجد والأضرحة الدينية والكنائس والكاتدرائيات."

Split List of Secret diacritics as four by four

◌ِ◌ُ◌ِ◌ُ	◌ُ◌َ◌ْ◌ْ	◌ِ◌َ◌ُ◌ِ	◌ْ◌ْ◌ْ◌ِ	◌ْ◌ْ◌ُ◌َ	◌َ◌َ◌ْ◌ْ	◌َ◌ِ◌َ◌ُ	◌ُ◌َ◌ْ◌ِ	◌ُ◌َ◌ْ◌ُ	◌ْ◌ِ◌ْ◌ْ	◌َ◌ِ◌ْ◌ُ

Stego Cover=

بسم الله الرحمن الرَّحيم الحمدُ لله الَّذي لا يفره الْمَنْعُ وَالْجُمُودُ وَلا يُكْديهِ الإعْطَاءُ وَالْجُودُ إذ كلّ
معطٍ منتقصٌ سواهُ وكلّ مانعٍ مذمومٌ مَا خَلاهُ وَهُوَ الْمَنَّانُ بفوائد النّعم وعوائد الْمَزِيدِ وَالْقِسَمِ
عِيَالُهُ الْخَلائِقُ بِجُودِهِ ضَمِنَ أَرْزَاقَهُمْ وَقَدَّرَ أقواتَهُمْ وَنَهَجَ سَبِيلَ الرَّاغبين إليه والطّالبين ما لديه
وليس بما سُئل بأجْوَدَ مِنْهُ بِمَا لَمْ يُسْأَلْ الأَوَّلُ الَّذي لَيْسَ لَهُ قَبْلٌ فَيَكُونَ شَيءٌ قَبْلَهُ وَالآخِرُ الَّذي ليس
لَهُ بعْدٌ فَيَكُونَ شَيءٌ بَعْدَهُ والرَّادِعُ أناسيّ الأَبْصَارِ عَنْ أَنْ تَنَالَهُ أَوْ تُدْرِكَهُ مَا اخْتَلَفَ عَلَيْهِ دَهْرٌ
فَتَخْتَلِفَ مِنْهُ الْحَالُ وَلا كَانَ في مَكَانٍ فيجُوز عليْهِ الانتقَالُ وَلَوْ وَهَبَ مَا تَنَفَّسَتْ عنه معادِنُ الْجِبَالِ
وَضَحِكَتْ عَنْهُ أصْدَافُ الْبِحَارِ مِنْ فِلِزِّ اللُّجَيْنِ وَسبائك العقيان ونُثَارَةِ الدُّرِّ وحصيد الْمَرجَان
لبعض عبيده ما أثّر ذلك في جُودِهِ وَلا أنْفَدَ سَعَةَ مَا عِنْدَهُ وَلَكَانَ عِنْدَهُ مِنْ ذَخَائِرِ الإنْعَامِ مَا لا
تُخْطِرُ لِكَثْرَتِهِ عَلى بَالٍ وَلا تُنْفِدُهُ مَطَالِبُ الأنَامِ ، لأنَّهُ الجَواد الَّذي لا يغيضه سؤال السَّائِلِينَ وَلا
يُبْخِله إلْحاح الْمُلحّين وَإنّما أمرُه إذا أراد شيْئاً أنْ يقُول لهُ كُنْ فَيَكُونُ فَمَا ظَنّكم بِمنْ هو هكذا ولا
هكذا غيْره سبْحَانَهُ وَ بِحَمْدِهِ
أيُّها السَّائِلُ اعقل عنّي ما سألتني عَنْهُ وَلا تَسْأَلَنَّ أحداً عنْه بعْدي فَإنّي أكفيك مؤُونَة الطّلب وشدّة
التّعمُّقِ في المذهب وكيف يُوصف الّذي سألتني عَنهُ وهو الّذي عجزت الْمَلائِكَةُ عَلى قُرْبِهِمْ مِنْ
كُرْسِيّ كَرامته وطُول ولهِهِمْ إلَيْهِ وَتَعْظِيمِ جَلالِ عِزَّته وقُرْبهم من غَيْبِ مَلَكُوتِهِ أنْ يَعْلَمُوا مِنْ
عِلْمِهِ إلاّ ما علَّمهُمْ وَهُوَ مِنْ مَلَكُوتِ الْقُدْسِ بِحَيْثُ هُمْ مِنْ مَعْرِفَتِهِ عَلى مَا فَطَرَهُمْ عَلَيْهِ فَقَالُوا :
سُبْحَانَك لا عِلْمَ لَنَا إلاّ ما علمتنا إنّك أنت العليمُ الْحَكِيم.

Table(1): Mapping numbers to DNA strand

Number	DNA Nucleotides
0	ATAC
1	ATAG
2	ATAA
4	ATTT
5	ATTC
6	ATTG
7	ATTA
8	ATCT
.	.
.	.
249	GACC
250	GACG
251	GACA "unused"
252	GAGT"unused"
253	GAGC"unused"
254	GAGG"unused"
255	CACA "unused"

Table (2): Arabic diacritics

Damaah	◌ُ
Fatha	◌َ
Kasrah	◌ِ
Sokoon	◌ْ
Tanween al- fatih	◌ً
Tanween al- kaser	◌ٍ
Tanween al- dam	◌ٌ
Shadda	◌ّ
Madda	~

Table(3): Mapping nucleotide into Arabic diacritic

A= Damaah (◌ُ)	T=Fatha (◌َ)	C= Kasrah(◌ِ)	G=Sokoon (◌ْ)

Table(4):Mapping number into Arabic diacritics	
DNA strand	Arabic diacritics
ATAT	◌ُ◌َ◌ُ◌َ
ATAC	◌ُ◌َ◌ُ◌ِ
ATAG	◌ُ◌َ◌ُ◌ْ
ATAA	◌ُ◌َ◌ُ◌ُ
ATTT	◌ُ◌َ◌َ◌َ
ATTC	◌ُ◌َ◌َ◌ِ
ATTG	◌ُ◌َ◌َ◌ْ
ATTA	◌ُ◌َ◌َ◌ُ
ATCT . .	◌ُ◌َ◌ِ◌َ . .
GACC	◌ْ◌ُ◌ِ◌ِ
GACG	◌ْ◌ُ◌ِ◌ْ
GACA "unused"	◌ْ◌ُ◌ِ◌ُ
GAGT "unused"	◌ْ◌ُ◌ْ◌َ
GAGC"unused"	◌ْ◌ُ◌ْ◌ِ
GAGG"unused"	◌ْ◌ُ◌ْ◌ْ
CACA "unused"	◌ِ◌ُ◌ِ◌ُ

Table (5): Capacity ratio of proposed Approach					
Diacritics Arabic Cover	Book Address	Secret Message Size (byte)	Real used Size of Cover (byte)	Hiding CapacityRatio,(byte/ byte) %	Average Capacity %
Al- Ashbaah	Nahij Al Balaghaa book	1900	1929	98	79
Abu- Hamzaa Al thamali	Mafateeh Al Jinaan boo[l])	1900	2229	85	
AL- Tawheed	Nahij Al Balaghaa book	1900	3549	54	
Sefat Al mutaqeen	Nahij Al Balaghaa book	902	787	115	91
Al- Ashbaah	Nahij Al Balaghaa book	902	1033	87	
Komail	Mafateeh Al Jinaan book[l])	902	1262	71	

Jawshan – Al-Kabeer	Mafateeh Al Jinaan book[])	500	482	104	106.4
Sefat – Al – Mutaqeen	Mafateeh Al Jinaan book[])	500	491	102	
Al-Eftitaah	Mafateeh Al Jinaan book[])	500	442	113	

Table(6) : Capacity ratio of other approaches

Approach	Average of Capacity (byte/byte) %
Shirali-Shaherza [7]	74.32
Gutub and Fattani [8]	33.68
Diacritic approach [9]	8.019
Kashidaa Approach [10]	10.25
Rehab and Nidaa Approach [5]	6.40

Table (7): Jaro similarity score for proposed approach

Test number	Jaro similarity score
Tes1	0.7913
Tes2	0.8188
Tes3	0.8214
Tes4	0.9024
Tes5	0.8231
Tes6	0.8412

Table(8): Jaro similarity score for other approaches

Approach	Jaro similarity score
Approach in [11]	0.8771
Approach in [12]	0.443
Approach in [33]	0.95

Table(9): Damerau-Levenshteindistance of this proposed approach

Test Address	NO. Differences	Percentage of Differences between Cover Text and Stego Text
Al-ashbaah	127	18.24%
Al-Eftitaah	93	16.93%
Abu hamza Al thamali	103	17.39%
Al-Tawheed	125	12.13%
Jawshan al kabeer	82	18.51%
Komail	184	23.98%

Table (10) : Damerau-Levenshtein distance of other previous approaches

Previous Approaches	NO. Differences
Khan's method [12]	282
Khan's method [12]	326

8. Conclusion

1. Use Arabic grammar rule as start stego key within cover text succeed to make stego key be more strong and secure against attackers because Arabic language have more than thousand grammar rule with their special cases , so it's so difficult to expected which grammar was used as stego key .

3. Using DNA strand as end stego key provide difficulty for attackers about the expected one DNA strand from 255 strands when prime=251.
4. Existence of control diacritics within cover text was excluded the suspense of attackers about the real embedding area and provide some permutation on the cover.
5. Because of there is specific secret area within cover, so cover will have over-crowding and few-crowding of diacritics area. This case may raise suspicion , So, we solved this by applying the same scenario of embedding process on areas before start stego key and after end stego key but without using secret diacritics only using random set of diacritic bases.
6. According to the results from popular measurements that used in steganography in text the proposed approach was gave an efficient results when comparison with pervious works in this field.
7. The proposed approach was provided high capacity ratio when compared with other text steganography techniques.
8. The cover media can be reused more than once if needed. However, hiding capacity will decrease drastically every time a new message is embedded into the cover text

Refrences

[1] Dr. HalaBahjat Abdul Wahab **"A New Steganography Approach depending on Isolation Curve in Digital Images", University of Technology .**

[2] Suhad M. Kadhem**," Using B+ Tree To Represent Secret Messages For Steganography Purpose"**,Eng.& Tech Journal, Vol (28),No(15),Iraq , March 2010.

[3] Krista Bennett**," Linguistic Steganography: Survey, Analysis, And Robustness Concerns For Hiding Informationin Text",** Center for Education and Research in Information Assurance and Security, Purdue University, West Lafayette,2004

[4] Auday Jamal Fawzi**,"Data Hiding in Arabic Text**",**,**PhD thesis , University of Technology , Iraq ,January,2007 .

[5] Rehab F. Hassan Dr. Nidaa F. Hassan, "**Data hiding in Arabic text based on Letters, Diacritics and Extension",**First Information Technology Conference, Iraq, April, 2009.

[6]Jan jannink, **"Implementing Deletion in B+ Trees"** ,Stanford University,Vol.24,No. 1,March 1995.

[7] M. Hassan Shirali-Shahreza, Mohammad Shirali- Shahreza, **"A New Approach to Persian/Arabic Text Steganography,"**5th IEEE/ACIS International Conferenceon Computer and Information Science (ICIS-COMSAR06), July 2006.

[8] Adnan Gutub , Manal Fattani, **"A Novel Arabic Text Steganography Method Using Letter Points and Extensions",** WASET International Conference on Computer, Information and Systems Science and Engineering (ICCISSE), Vienna, Austria, May ,2007.

[9] F. Al-Azawi, MoayadFadhil, **"Arabic Text Steganography using Kashidaa Extensions with Huffman code** ", Asian network for Science Information,2010.

[10] Adnan Abdul-Aziz Gutub, Wael Al-Alwani, and Abdulelah Bin, Mahfoodh **"Improved Method of Arabic Text Steganography Using the Extension Kashida Character",**Bahria University Journal of Information & Communication Technology, December, 2010

[11]Souvik Bhattacharyya , Indradip Banerjee and GautamSanyal," A Novel Approach of Secure Text Based Steganography Model using Word Mapping Method(WMM)",International Journal of Computer and Information Engineering ,2010.

[12] Khan Farhan Rafat, M. Sher**"On the Limits of Perfect Security for Steganographic System"**, Department of Computer Science, International Islamic University Islamabad, 44000, Pakistan,Oct ,2013.

[13]Monika Agarwal **,"Text Steganographic Approach : A comparison",**
International Journal of Network Security & Its Applications (IJNSA),
Vol (5), No(1), Jabalpur, India, January 2013