Novel steganography scheme using Arabic text features in Holy Quran

Huda Kadhim Tayyeh[1], Mohammed Salih Mahdi[2], Ahmed Sabah Ahmed AL-Jumaili[3]
[1]Department of Informatics Systems Management (ISM), College of Business Informatics,
University of Information Technology & Communications, Iraq
[2,3]Department of Business Information Technology (BIT), College of Business Informatics,
University of Information Technology & Communications, Iraq

Article Info

Article history:

Received Dec 8, 2018
Revised Feb 16, 2019
Accepted Mar 4, 2019

Keywords:

Arabic language
Diacritics
Information security
Quran
Steganography

ABSTRACT

With the rapid growth of the Internet and mobile devices, the need for hidden communications has significantly increased. Steganography is a technique introduced for establishing hidden communication, Most steganography techniques have been applied to audio, images, videos, and text. Many researchers used steganography in Arabic texts to take advantage of adding, editing or changing letters or diacritics, but lead to notable and suspicious text. In this paper, we propose two novel steganography algorithms for Arabic text using the Holy Quran as cover text. The fact that it is forbidden to add, edit or change any letter or diacritics in the Holy Quran provides the valuable feature of its robustness and difficulty as a cover in steganography. The algorithms hide secret messages elements within Arabic letters benefiting from the existence of sun letters (Arabic: ḥurūf shamsīyah) and moon letters (ḥurūf qamarīyah). Also, we consider the existence of some Arabic language characteristics represented as small vowel letters (Arabic Diacritics). Our experiments using the proposed two algorithms demonstrate high capacity for text files. The proposed algorithms are robust against attack since the changes in the cover text are imperceptible, so our contribution offers a more secure algorithm that provides good capacity.

Corresponding Author:

Huda Kadhim Tayyeh,
Department of Informatics Systems Management,
University of Information Technology & Communications,
Baghdad, Iraq.
Email: haljobori@uoitc.edu.iq

1. INTRODUCTION

An important issue today as well as for centuries is the hidden exchange and security of information,and the Internet has given this need special significance [1]. Different methods are used in data hiding such as watermarking, steganography, and cryptography [2]. A key controls the encryption of information in cryptography, so that no one can decrypt and access the information except the person who knows the key. Steganography is one of the best methods for secure communication [3]. The word steganography originates from the Greek language, which means hidden writing. ‘‘Stegano” means hidden and ‘‘graptos” means writing [4]. The goal in steganography is to conceal secret information under cover media, so unauthorised persons cannot discover the contained information. This cover media approach differentiates steganography from other methods for exchanging hidden information. After data the hiding, the text containing the secret information referred to as the stego-text, is sent from sender to receiver via the Internet.

The goal of the security is that no one can notice the secret information embedded into the stego-text easily by using a variety of detection techniques. Three criteria for designing steganography systems include robustness, perceptual transparency, and hiding capacity [5]. Robustness is the ability to protect the hidden information from damage when transmitted from the sender to receiver. Perceptual transparency means the ability of the attackers to notice the hidden information easily. By minimising the difference between the cover text and stego-text, high security can be achieved. The capacity represents the size of information bits that can be concealed by the cover text. Pictures [6], video clips [7], music, and sounds [8] are typical cover media, or carrier, for steganography methods.

The most challenging approach is text steganography due to the shortage of redundant information available in text files compared to other cover media types [9], [10]. The structure of the text files is usually just as how it is seen, whereas the structure for other carrier types is entirely different from how the media is observed. This makes the information hiding in non-text cover media easier and more difficult to be discovered compared to hiding information in text files. An advantage of text steganography is its simplicity in communication and occupies less memory resources [4]. So, different steganographic techniques are used for different languages depending on the structure [10].

The two steganography algorithms recommended in this paper are used grammar rule of the definite article al followed by sun letters (Arabic: ḥurūf shamsīyah) and moon letters (ḥurūf qamarīyah) along with the Arabic diacritics (Harakat) to hide data in Arabic text using Holy Quran as cover. The fact that Holy Quran consists of Arabic characters and Arabic diacritics (Harakat) provides the valuable feature of its robustness as a cover in steganography.The hiding information in cover media does not attract the human attention because the information is hidden without any perceptible change in the original word.

2. RELATED WORK IN ARABIC TEXT STEGANOGRAPHY

Most text steganography methods are used for English texts, and only a few are applied to Arabic text [11]-[14]. The Arabic language is the sixth most spoken language with more than 420 million people speakers worldwide [15]. The Quran is the Holy book for more than one third the population of the world and is written in the classic Arabic language [16]. Some features of the Arabic language do not match to other languages, including English [15]. Writing in the Arabic language uses a cursive style with a right to left direction. Also, the shape of each Arabic character is different depending on its position in the word. The Arabic language is characterised by many dotted letters with some having one dot on top or bottom of a letter and others with two or three dots on top of a letter [16]. There exist additional marks positioned on the top or bottom of Arabic letters called “Diacritics” or Harakat, as it is known in Arabic. There are eight shapes of Diacritics representing only the vowel sounds [17] and are called Fathah, Kasrah, Damah, Sukun, TanwinFathah, TanwinKasrah, TanwinDamah, and Shaddah, as shown in Figure 1. The computer represents each Diacritic digitally as separate character. These Diacritics are fundamental for the Holy Quran and other religious and historical scripts, but non-compulsory in modern standard Arabic writing and practice [17]. The following summarises various approaches for Arabic text steganography.

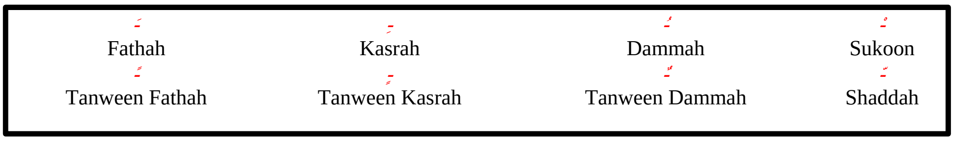


Figure 1. Arabic text diacritics

2.1. Kashida-based steganography

There is a possibility using Arabic letters to add an extension in words, and this feature is called “Kashida”, which does not affect the meaning of the words. So, words with an extension “Kashida” can be used to hide information and words without an extension will hide none [13], [17], [18]. Although, in this method, the message content will not be affected, but has the disadvantage that it cannot be added to the beginning or end of words and only in the middle of connected letters within a word. This restriction makes it more notable to the readers as it obviously changes the text while it also increases the size of the file.

2.2. Steganography by displacement of points

In this method, the information is embedded as binary values in the dots (points) of the letters of the language, such as in Arabic, Urdu, and Persian [11], [18]. When the point position is shifted up, then the value of the hidden bit is one. Otherwise, the dot position is unchanged, and the value is zero. With this approach, it is possible to hide a large amount of information in Arabic text without bringing attention to changes. Since the Arabic language includes 15 dotted letters out of 28, the capacity of hiding is high. However, a special font is required to accomplish this subtle variation, so the receiver will not be able to retrieve the hidden message if the same font is not available. In addition, if the message is re-typing or OCR scanning is performed, then the details of the hidden information are likely lost [5].

2.3. Unicode-based steganography

In accordance with Unicode standards, there are many forms of Arabic letters and are divided into two groups with one being the representative code and the other comprised of the possible shapes of the letters. With this method, it is possible to use various Unicode values for the same letter to hide bits of information [5], [19]. This method is not secure enough against the traditional intruders as some Unicode-based steganography techniques have a high capacity with less security [14] and vice versa.

2.4. Steganography using Arabic diacritics (Harakat)

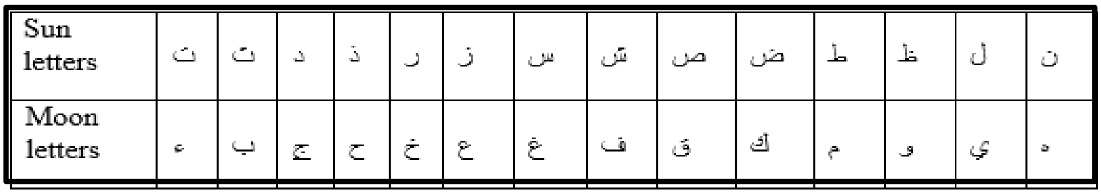
As previously defined, the diacritics are extension characters used optionally at the top and the bottom of Arabic letters. The diacritics symbols are used to differentiate between words composed of the same letters but pronounced differently. In Arabic text, it is found that “Fatha” covers almost half the used diacritics, while all other diacritics cover the other half. For this reason, “Fatha” is chosen to hide the binary value (1) and the other diacritics are chosen to hide (0) [12], [20]. This method’s key disadvantage is that it utilisesobvious changes and is easily recognisable by the reader.

2.5. Linguistic-based steganography

This technique is classified into the three types, including lexical-based steganography, translation-based steganography, and the noise-based approach. Linguistic steganography refers to the use of word synonyms to hide secret messages in ordinary language text [21]. The covering text is very natural and ordinary regarding the language and gives a reasonable accuracy for the selected synonym. It is important to ensure there is no repetition of the same cover text for hiding a message because this would bring it to the attention of readers. Also, this method offers a low capacity for hidden information [14], [21]. The message within translation-based steganography may be hidden in errors, or noise, in the text, which typically occurs during machine translation (MT). The confidential message is hidden by performing the substitution procedure on the translated text using translation differences from several MT systems [14]. In the noise-based approach, typographical and abbreviation errors are used to hide data in text, such as e-mails, blogs, and forums. However, this approach depends on mistakes made through human writing [21].

3. THE PROPOSED ALGORITHMS

The Arabic alphabet contains 28 letters with consonants divided into two groups, named **sun letters** and **moon letters** based on whether they assimilate the letter *lām* (ال) of a preceding definite article*al*- (الـ). Figure 2 lists the sun and moon letters.



Sun letters	ت	ث	د	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ل	ن
Moon letters	ء	ب	ج	ح	خ	ع	غ	ف	ق	ك	م	و	ي	ه

Figure 2. Sun and moon letters

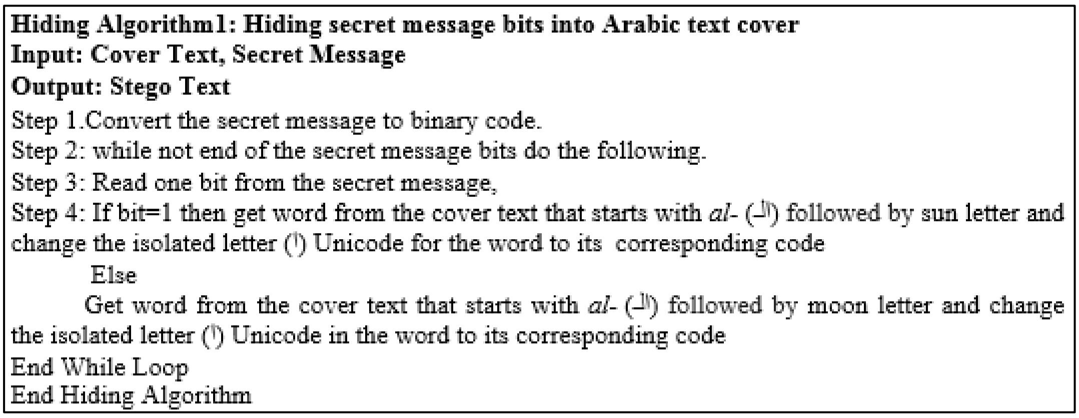
The proposed algorithms hide secret binary data into Arabic text using the grammar rule of the definite article *al* along with the Arabic diacritics (Harakat). In the first algorithm, the secret message is hidden in words beginning with *al*- (الـ) followed by a sun or moon letter. In the Unicode standard, the isolated letter (ا) has two codes because it is a representative letter. The first code is used only to save data in the digital media and the second is used for the correct shape for each letter. This feature is used in our algorithms to indicate the hiding location in each word that starts with the definite article *al*- (الـ). So, the secret message can be hidden in the cover text without any perceptible change in the original word.

3.1. Hiding process

The two proposed hiding algorithms are illustrated in the following sections.

3.1.1. Hiding process for proposed Algorithm 1

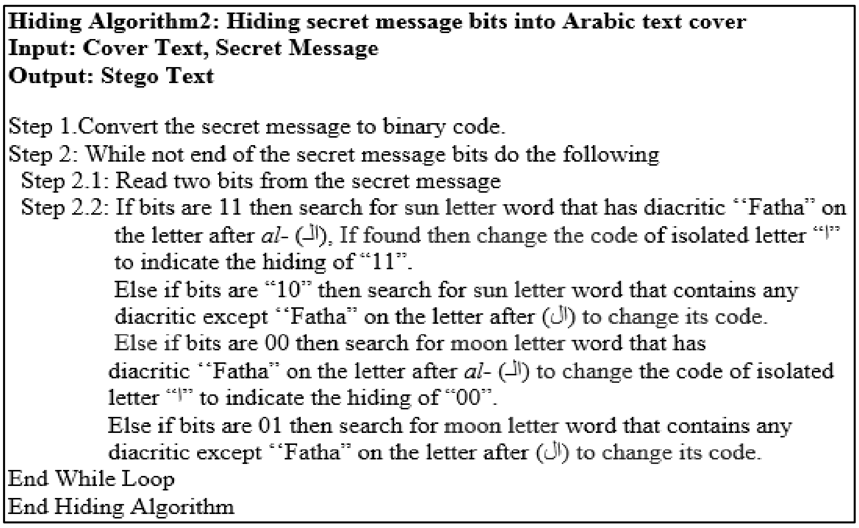
In this algorithm, one bit is hidden using the isolated letter (ا) in any word beginning with *al*- (الـ) followed by a sun or moon letter. The following algorithm 1 outline the hiding process



```
Hiding Algorithm1: Hiding secret message bits into Arabic text cover
Input: Cover Text, Secret Message
Output: Stego Text
Step 1.Convert the secret message to binary code.
Step 2: while not end of the secret message bits do the following.
Step 3: Read one bit from the secret message,
Step 4: If bit=1 then get word from the cover text that starts with al- (الـ) followed by sun letter and
change the isolated letter (ا) Unicode for the word to its  corresponding code
        Else
        Get word from the cover text that starts with al- (الـ) followed by moon letter and change
the isolated letter (ا) Unicode in the word to its corresponding code
End While Loop
End Hiding Algorithm
```

3.1.2. Hiding process for proposed Algorithm 2

With this algorithm, the embedding capacity is increased by hiding two bits in each word. So, a secret message is hidden in the cover text by using the isolated letter "ا" in sun or moon letter words and includes diacritics without any perceptible change in the original word.



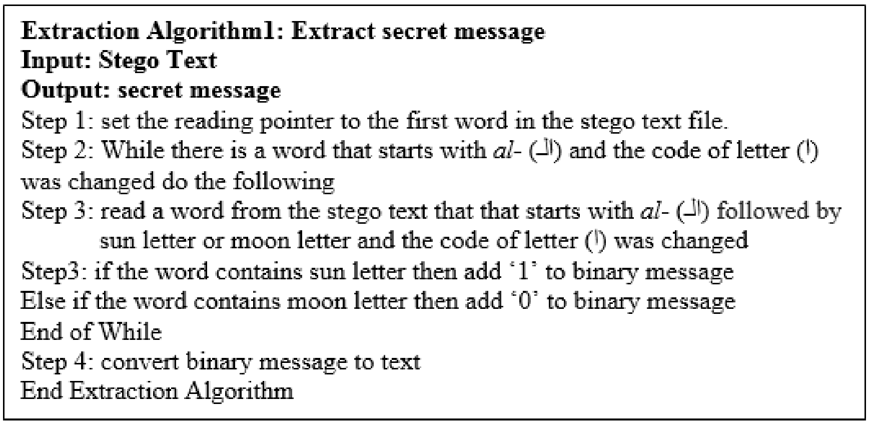
```
Hiding Algorithm2: Hiding secret message bits into Arabic text cover
Input: Cover Text, Secret Message
Output: Stego Text

Step 1.Convert the secret message to binary code.
Step 2: While not end of the secret message bits do the following
  Step 2.1: Read two bits from the secret message
  Step 2.2: If bits are 11 then search for sun letter word that has diacritic ‘‘Fatha” on
            the letter after al- (الـ), If found then change the code of isolated letter “ا”
            to indicate the hiding of “11”.
            Else if bits are “10” then search for sun letter word that contains any
            diacritic except ‘‘Fatha” on the letter after (ال) to change its code.
            Else if bits are 00 then search for moon letter word that has
            diacritic ‘‘Fatha” on the letter after al- (الـ) to change the code of isolated
            letter “ا” to indicate the hiding of “00”.
            Else if bits are 01 then search for moon letter word that contains any
            diacritic except ‘‘Fatha” on the letter after (ال) to change its code.
End While Loop
End Hiding Algorithm
```

3.2. Extraction process

A separate algorithm is utilised to extract the hidden message generated from Algorithm 1 or Algorithm 2.

3.2.1. Extraction from Algorithm 1

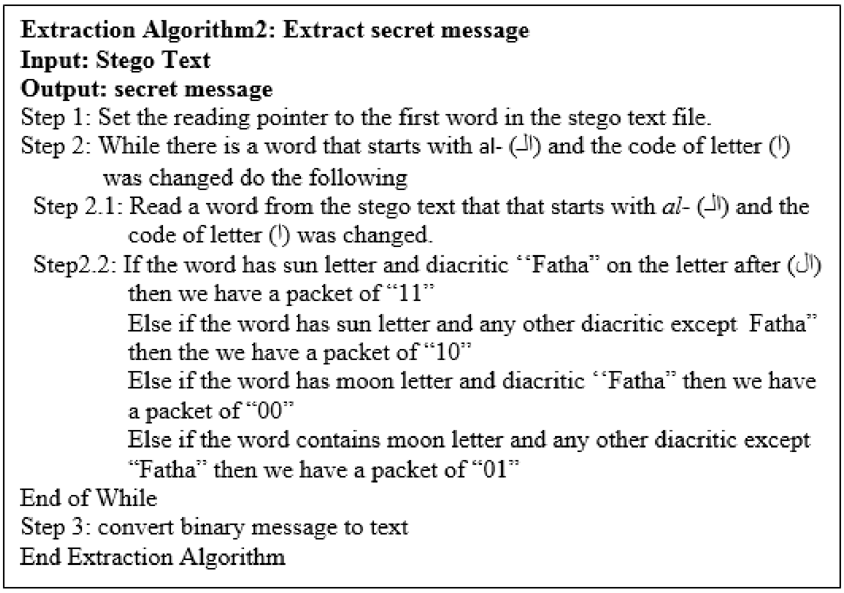
The following algorithm shows how to extract a hidden message from stego-text generated by Algorithm 1.



```
Extraction Algorithm1: Extract secret message
Input: Stego Text
Output: secret message
Step 1: set the reading pointer to the first word in the stego text file.
Step 2: While there is a word that starts with al- (الـ) and the code of letter (ا)
was changed do the following
Step 3: read a word from the stego text that that starts with al- (الـ) followed by
        sun letter or moon letter and the code of letter (ا) was changed
Step3: if the word contains sun letter then add '1' to binary message
Else if the word contains moon letter then add '0' to binary message
End of While
Step 4: convert binary message to text
End Extraction Algorithm
```

3.2.2. Extraction from Algorithm 2

The following algorithm shows how to extract a hidden message from stego-text generated by Algorithm 2.



```
Extraction Algorithm2: Extract secret message
Input: Stego Text
Output: secret message
Step 1: Set the reading pointer to the first word in the stego text file.
Step 2: While there is a word that starts with al- (الـ) and the code of letter (ا)
        was changed do the following
  Step 2.1: Read a word from the stego text that that starts with al- (الـ) and the
            code of letter (ا) was changed.
  Step2.2: If the word has sun letter and diacritic ''Fatha" on the letter after (ال)
            then we have a packet of "11"
            Else if the word has sun letter and any other diacritic except  Fatha"
            then the we have a packet of "10"
            Else if the word has moon letter and diacritic ''Fatha" then we have
            a packet of "00"
            Else if the word contains moon letter and any other diacritic except
            "Fatha" then we have a packet of "01"
End of While
Step 3: convert binary message to text
End Extraction Algorithm
```

4. EXPERIMENTAL RESULTS

First, this section further explains the two proposed algorithms through examples. Then, the performance of the algorithms is examined based on their embedding ratio factor. In the proposed algorithms, the secret message is hidden in Arabic texts using the Holy Quran surahs as cover. The diacritics in the Holy Quran surahs are compulsory resulting in large cover file size.

4.1. Experiment 1

For the first experiment, we used the cover media of Surat Al-Fatiha (in plain text) to hide the secret code '001110' within this Arabic text following Algorithm 1, which generates:

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ
(1) الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ (2) الرَّحْمَٰنِ الرَّحِيمِ (3) مَالِكِ يَوْمِ الدِّينِ (4) إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ (5) اهْدِنَا الصِّرَاطَ الْمُسْتَقِيمَ (6) صِرَاطَ
الَّذِينَ أَنْعَمْتَ عَلَيْهِمْ غَيْرِ الْمَغْضُوبِ عَلَيْهِمْ وَلَا الضَّالِّينَ (7)

According to the hiding process of Algorithm 1, we search for the moon letter words in the cover text to hide bit 0. To hide 1, we search for a sun letter word in the cover text and change the code of isolated letter "ا" to mark the hiding of bit 1. Figure 3 demonstrates how to hide '001110' in Arabic text (Surat Al-Fatiha).

Cover text	الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ (2) الرَّحْمَٰنِ الرَّحِيمِ (3) مَالِكِ يَوْمِ الدِّينِ (4) إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ (5) اهْدِنَا (1) الصِّرَاطَ الْمُسْتَقِيمَ (6) صِرَاطَ الَّذِينَ أَنْعَمْتَ عَلَيْهِمْ غَيْرِ الْمَغْضُوبِ عَلَيْهِمْ وَلَا الضَّالِّينَ (7)								
Sun/moon letter words	الضَّالِّ ين	الْمَغْضُوبِ	الْمُسْتَقِيمَ	الصِّرَاطَ	الدِّينِ	الرَّحِيمِ	الرَّحْمَٰنِ	الْعَالَمِينَ	الْحَمْدُ
The changed words	□	□	■	□	■	■	■	■	■
Hidden bits			0		1	1	1	0	0
Stego-text	الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ (2) الرَّحْمَٰنِ الرَّحِيمِ (3) مَالِكِ يَوْمِ الدِّينِ (4) إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ (5) اهْدِنَا (1) الصِّرَاطَ الْمُسْتَقِيمَ (6) صِرَاطَ الَّذِينَ أَنْعَمْتَ عَلَيْهِمْ غَيْرِ الْمَغْضُوبِ عَلَيْهِمْ وَلَا الضَّالِّينَ (7)								

Figure 3. Thehiding process for the secret message ‘001110’ in the Arabic text where means the Unicode of letter “ا” is changed and means the word is not used

To extract the hidden message from the stego-text produced from the previous example, we perform the following:The first word in the stego-text is identified that starts with *al-* (الـ) followed by a sun or moon letter in which the code of the letter (ا)is changed (i.e., the word “**الْحَمْدُ**” has a moon letter). So, this defines a bit 0, which initiates the extracting string with a 0. This process is repeated until the entire secret message is extracted. Figure 4 demonstrates how to extract a secret message from stego-text.

Stego-text	الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ (2) الرَّحْمَٰنِ الرَّحِيمِ (3) مَالِكِ يَوْمِ الدِّينِ (4) إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ (5) اهْدِنَا (1) الصِّرَاطَ الْمُسْتَقِيمَ (6) صِرَاطَ الَّذِينَ أَنْعَمْتَ عَلَيْهِمْ غَيْرِ الْمَغْضُوبِ عَلَيْهِمْ وَلَا الضَّالِّينَ (7)								
Sun/moon letter words	الضَّالِّينَ	الْمَغْضُوبِ	الْمُسْتَقِيمَ	الصِّرَاطَ	الدِّينِ	الرَّحِيمِ	الرَّحْمَٰنِ	الْعَالَمِينَ	الْحَمْدُ
The changed words	□	□	■	□	■	■	■	■	■
Hidden bits			0		1	1	1	0	0
Secret message	001110								

Figure 4. The extraction process of a secret message from stego-text where means the Unicode of letter “ا” was changed and means the word was not used

4.2. Experiment 2

The same Arabic text cover media (Surat Al-Fatiha) is used to hide the secret code ‘001110’ following Algorithm 2 resulting in:

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ
(1)الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ (2) الرَّحْمَٰنِ الرَّحِيمِ (3) مَالِكِ يَوْمِ الدِّينِ (4) إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ (5) اهْدِنَا الصِّرَاطَ الْمُسْتَقِيمَ (6) صِرَاطَ
الَّذِينَ أَنْعَمْتَ عَلَيْهِمْ غَيْرِ الْمَغْضُوبِ عَلَيْهِمْ وَلَا الضَّالِّينَ (7)

According to this hiding process, we search the cover for the first moon letter word containing the diacritic ‘‘Fatha” and change the code on the letter after (ال).In this case, the word “**الْحَمْدُ**” satisfies the two conditions, so we can change the code of the isolated letter “ا” to mark the hiding of bit 00.To hide 11, we search for the next sun letter word containing the diacritic ‘‘Fatha” on the letter after (ال) to change its code. In this case, the word “**الرَّحْمَٰن**” has the sun letter “ر” and diacritic ‘‘Fatha,” so we change the code of the isolated letter “ا.” The last two bits 10 are hidden in the word “**الدِّين**” since it contains the sun letter “د” and diacritic “Kasrah.” Figure 5 demonstrates how to hide ‘001110’ in Arabic text (Surat Al-Fatiha) using Algorithm 2.

Cover text	الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ (2) الرَّحْمَٰنِ الرَّحِيمِ (3) مَالِكِ يَوْمِ الدِّينِ (4) إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ (5) اهْدِنَا (1) (الصِّرَاطَ الْمُسْتَقِيمَ (6) صِرَاطَ الَّذِينَ أَنْعَمْتَ عَلَيْهِمْ غَيْرِ الْمَغْضُوبِ عَلَيْهِمْ وَلَا الضَّالِّينَ (7								
Sun/moon letter words	الضَّالِّين	الْمَغْضُوبِ	الْمُسْتَقِيمَ	الصِّرَاطَ	الدِّينِ	الرَّحِيمِ	الرَّحْمَٰنِ	الْعَالَمِينَ	الْحَمْدُ
The changed words	□	□	□	□	■	□	■	□	■
Hidden bits					10		11		00
Stego-text	الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ (2) الرَّحْمَٰنِ الرَّحِيمِ (3) مَالِكِ يَوْمِ الدِّينِ (4) إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ (5) اهْدِنَا (1) (الصِّرَاطَ الْمُسْتَقِيمَ (6) صِرَاطَ الَّذِينَ أَنْعَمْتَ عَلَيْهِمْ غَيْرِ الْمَغْضُوبِ عَلَيْهِمْ وَلَا الضَّالِّينَ (7								

Figure 5. The hiding process for the secret message ‘001110’ in Arabic text where means that the Unicode of letter “ا” is changed and means the word was not used

The following is performed to extract the hidden message from the stego-text produced from this example:We identify the first word in the stego-text starting with *al-* (الـ) followed by a sun or moon letter and the code of the letter (ا)was changed. If found, check the diacritic on the letter. Since the word “**الْحَمْدُ**” has a moon letter, the code of letter (ا)was changed, and the diacritic is ‘‘Fatha,” we extract two bits 00. This process is repeated until the entire secret message is extracted. Figure 6 demonstrates how to extract a secret message from stego-text utilising this approach.

Stego-text	الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ (2) الرَّحْمَٰنِ الرَّحِيمِ (3) مَالِكِ يَوْمِ الدِّينِ (4) إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ (5) اهْدِنَا (1) (الصِّرَاطَ الْمُسْتَقِيمَ (6) صِرَاطَ الَّذِينَ أَنْعَمْتَ عَلَيْهِمْ غَيْرِ الْمَغْضُوبِ عَلَيْهِمْ وَلَا الضَّالِّينَ (7								
Sun/moon letter words	الضَّالِّين	الْمَغْضُو بِ	الْمُسْتَقِيمَ	الصِّرَاطَ	الدِّينِ	الرَّحِيمِ	الرَّحْمَٰنِ	الْعَالَمِينَ	الْحَمْدُ
The changed words	□	□	□	□	■	□	■	□	■
Hidden bits					10		11		00
Secret message	001110								

Figure 6. The extraction process for a secret message from stego-text where means that the Unicode of letter “ا” was changed and means the word was not used

4.3. Results and Analysis

The goals of a good steganographic scheme are high embedding payload and high imperceptibility. Tomeasure the performance of the proposed algorithms, seven Arabic text files (Holy Quran surahs) were selected for computing imperceptibility and payload. The file size of the Holy Quran surah is large because the Arabic textsurah includes diacritics and many special characters. These characters are compulsory, and it is not acceptable to add, change or delete any character. So, most steganography methods, such as shifting points, Kashida-based, and linguistic-based steganographyare not applied to the Holy Quran Arabic text. The proposed algorithms effectively counter visual attack because they do not raise any doubt from apparent changes in the text. This is not the case for other format-based algorithms that modify the text to hide secret information. The hiding capacity of the algorithms is calculated for evaluation using the formula:

$$\text{hiding capacity} = \text{bits of secret message}/\text{bits of stego-text} \tag{1}$$

Table 1 shows the computed results for the hiding capacity of the proposed algorithms, which suggest the capacity for the methods is adequate.

Table 1. The Computed Embedding Ratios

Filename	Cover size in kb	Algorithm 1 Capacity (bit)	Capacity ratio (b/kB)	Algorithm 2 capacity	Capacity ratio (b/kB)
AlFatihah	2.5	11	4.4	22	8.8
AlBaqarah	437.6	853	1.94	1670	3.8
Al'Imran	251.5	546	2.17	1092	4.3
'AnNisa	269.5	505	1.87	1010	3.74
AlMa'idah	201.7	424	2.1	848	4.2
AlAn'am	215.7	343	1.59	686	3.18
Yusuf	242.3	367	1.51	734	3.02
		Total Average Capacity = 2.23		Total Average Capacity = 4.43	

Because the proposed approach is a hybrid between diacritics, grammar rules, and Unicode approaches therefore, it is difficult to compare it with similar approaches. The second proposed algorithm is compared to two diacritics approaches. Table 2 shows the average capacity of the two approaches using the data set published in [12]. For harakat approach [22], the average capacityis 3.27 where it is 6.4 for diacritics-based approach [12]. The results show that the average capacity of proposed approach is more than harakat approach [22] and less than diacritics-based approach [12]. According to imperceptibility, all the diacritics approaches have low imperceptibility due to the change of cover files. Notice that the two algorithms presented here have high imperceptibility and are designed for religious documents as cover.

Table 2. Comparison between Diacritics Methods

Approach	Average Capacity %	Imperceptibility	Evaluation
Harakat approach	3.27	Low	- The approach attracts the attention of the reader because the diacritics are inserted to the cover. - It is not suitable for religious documents as cover.
High Capacity Diacritics-based Method For Information Hiding in Arabic Text	6.4	Low	- Some Diacritics are deleted from the cover to hide 0. - The approach raises the attention of the reader. - It is not suitable for religious documents as cover.

5. CONCLUSION AND FUTURE WORK

This paper presents a novel steganography scheme useful for Arabic language electronic writing. The proposed algorithms are new because they are the first to use Holy Quran surah's as cover media along with combining Arabic grammar, diacritics, and Unicode rules to hide secret information. Therefore, this method is robust with a very low possibility of deciphering.

The experimental results of the algorithms demonstrate the following:

1. The information is hidden with minimum changes in the cover text, so the perceptual transparency is satisfied.
2. The proposed algorithms are robust against traditional attack since the secret message is hidden in the cover text using minimum changes and in different positions.
3. The hiding process uses diacritics without adding, shifting or deleting them.
4. The proposed methods do not need the cover file to extract the message.
5. The proposed methods do not change the cover file size and do not require the availability of a specific font.
6. The capacity ratios for the proposed algorithms are not very high due to the type of the cover.

REFERENCES

[1] E. A. Abbood, *et al.*, "Text in Image Hiding using Developed LSB and Random Method," *International Journal of Electrical and Computer Engineering (IJECE)*, vol/issue: 8(4), pp. 2091-2097, 2018.

[2] R. Din, *et al.*, "A Comparative Review on Data Hiding Schemes," *International Journal of Electrical and Computer Engineering (IJECE)*, vol/issue: 11(2), pp. 768-774, 2018.

[3] R. S. Sabri, *et al.*, "Analysis Review on Performance Metrics for Extraction Schemes in Text Steganography," *International Journal of Electrical and Computer Engineering (IJECE)*, vol/issue: 11(2), pp. 761-767, 2018.

[4] N. Provos and P. Honeyman, "Hide and Seek: An Introduction to Steganography," *IEEE Security & Privacy*, pp. 32-44, 2003.

[5] Shahreza M. S. and Shahreza M. H., "An improved version of Persian/Arabic text steganography using ‘‘La" word," *Proceedings of IEEE 6th national conference on telecommunication technologies*, pp. 372–6, 2008.

[6] R. Chandramouli and N. Memon, ‘Analysis of LSB based image steganography techniques," *Proceedings of the International Conferenceon Image Processing*, vol. 3, pp. 1019–1022, 2001.

[7] G. Doërrand and J. L. Dugelay, “A Guide Tour of Video Watermarking,” *Signal Processing: Image Communication*, vol/issue: 18(4), pp. 263-282, 2003.

[8] K. Gopalan, “Audio steganography using bit modification,” *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '03),* vol. 2, pp. 421-424, 2003.

[9] M. H. S. Shahreza and M. S. Shahreza, “A New Approach to Persian/Arabic Text Steganography,” *5th IEEE/ACIS International Conference on Computer and Information Science (ICISCOMSAR 06)*, pp. 310- 315, 2006.

[10] W. Bender, *et al.*, “Techniques for data hiding,” *IBM Systems Journal*, vol/issue: 35(4), pp. 313-336, 1996.

[11] O. Ammar, *et al.*, “Steganography by multipoint Arabic letters,” *Systems, applications and technology conference (LISAT)*, pp. 1–7, 2012.

[12] B. M. Lahcen and Y. M. Bachir, “High capacity diacritics-based method for information hiding in Arabic text,” *International conference on innovations in information technology,* pp. 433-436, 2011.

[13] A. F. Al-Azawi and M. A. Fadhil, “Arabic Text Steganography using Kashida Extensions with Huffman Code,” *Journal of Applied Sciences,* vol. 10, pp. 436-439, 2010.

[14] A. A. Mohamed, “An improved algorithm for information hiding based on features of Arabic text: A Unicode approach,” *Egyptian Informatics Journal,* vol/issue: 15(2), pp. 79–87, 2014.

[15] ISTIZADA, “Complete List of Arabic Speaking countries 2017,” Available: *http://istizada.com/complete-list-of-arabic-speaking-countries*, 2018.

[16] A. G. Chejne, “The Arabic Language: its Role in History,” University of Minnesota Press, Minneapolis, 1969.

[17] F. Al-Haidari, *et al.*, “Improving security and capacity for Arabic text steganography using 'Kashida' extensions,” *Proc. AICCSA 2009 - The 7th ACS/IEEE International Conference on Computer Systems and Applications*, Rabat, Morocco, pp. 396-399, 2009.

[18] Al-Nazer A. and G. Adnan, “Exploit Kashida adding to Arabic e-text for high capacity steganography,” *Proceedings of the third international conference on network and system security NSS ’09*, IEEE, pp. 447–51, 2009.

[19] Shirali S. M. and S. S. S. Persian, “Arabic Unicode text steganography,” *The fourth international conference on information assurance and security*, IEEE, pp. 62–6, 2008.

[20] Aabed M. A., *et al.*, “Arabic diacritics based steganography,” *Proceedings of the international conference on signal processing and communications,* pp.756–9, 2007.

[21] Listega D. A., “List-based steganography methodology,” *International Journal of Information Security*, vol/issue: 8(4), pp. 247–261, 2009.

[22] M. Aabed, *et al*., “Arabic diacritics based steganography,” *Signal Processing and Communications, 2007. ICSPC 2007. IEEE International Conference*, pp. 756-759, 2007.

BIOGRAPHIES OF AUTHORS

Dr. Huda Kadhim Tayyeh PhD in computer Science and Information Systems from University of Technology, Baghdad, Iraq. Head ofInformatics Systems Management (ISM) Department, College of Business Informatics/University of Information Technology & Communications. Instructor in Informatics Systems Management (ISM) Department, College of Business Informatics (BIC), University of Information Technology & Communications (UOITC).

Mohammed Salih Mahdi MSc in a security of cloud computing in 2012 from University of Technology, Baghdad, Iraq. Instructor in Business Information Technology (BIT) Department, College of Business Informatics (BIC), University of Information Technology & Communications (UOITC).

Dr. Ahmed Sabah Ahmed AL-Jumaili PhD in computer Science and Information Systems from University of Technology, Baghdad, Iraq. Head of Quality Assurance and University Performance at Business Informatics College (BIC), University of Information Technology & Communications (UOITC). Instructor in Business Information Technology (BIT) Department, College of Business Informatics (BIC), University of Information Technology & Communications (UOITC).