Utilization of Two Diacritics for Arabic Text Steganography to Enhance Performance

Esraa Mohammad Ahmadoh
Collage of Computer & Information Systems, Umm Al-Qura University, Makkah, Saudi Arabia

Adnan Abdul-Aziz Gutub
Custodian of the Two Holy Mosques Institute of the Hajj & Omrah Research, Umm Al-Qura University, Saudi Arabia
Email: aagutub@uqu.edu.sa

***Abstract*—An improved Arabic text steganography method is proposed. It hides secret data into text cover media in two Diacritics chosen based on highest availability percentage among all Diacritics, which are eight in Arabic language. This utilization of Diacritics- or Harakat - for security purposes is benefiting from the natural existence of Diacritics as historical Characteristics of Arabic language, originated to just represent vowel sounds. The paper exploits the possibility of hiding data in two Diacritics, i.e. Fathah and Kasrah, adjusting the previously presented single (Fathah only) diacritic hiding scheme. This proposed two Diacritics stego-work is featuring higher capacity and security showing interesting promising results.**

***Index Terms*—Arabic text steganography, information security, diacritics steganography**

I. Introduction

As data technology usage is developing, communication is increasing and the amount of information exchanging is requiring transactions to be more secure. Steganography is becoming famous among security methods used to hide information for confidentiality, integrity, and availability. The word 'steganography' is originated from Greek coming from "Stegano" meaning hidden and "Graptos" meaning writing [1].

In steganography, the secure data will be embedded into another media object, i.e. stego-cover, so middle attacker and eavesdropper cannot harm it [1]. This steganography cover media can be image, sound or text as means to hold the hidden information [2]. The classification of different stego techniques are shown in Fig. 1, which is categorizing the types of the cover media according to all stego-cover possibilities [3]. The first category in the classification divides steganography according to the cover message type. The linguistic categorization exploits the computer-coding techniques to hide information [4]. Our study will focus on text under semagrams, as part of linguistic, i.e. to hide information through the use of signs and symbols.

Manuscript received March 3, 2015; revised July 7, 2015.

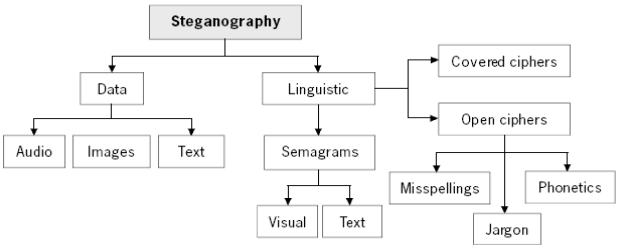

Figure 1. Classification of steganography [3]

There are three aspects in steganography information hiding where the systems contend with each other: capacity, security, and robustness. Capacity refers to the amount of information able to be hidden in the cover-medium, whereas security is important when a secret communication is kept to be confidential and undetectable by eavesdroppers. Lastly, robustness can be explained as the amount of modification the stego-medium can withstand before an adversary can destroy its hidden data [5].

In this paper, we propose an Arabic text steganography security approach that would use Diacritics for hiding. This utilization of Diacritics - or Harakat - for security purposes is a novel advantage from the natural existence of Diacritics as Characteristics of Arabic Language, which are created originally to just represent vowel sounds. Our method proposes to hide secret message in Arabic text using two Diacritics instead of the previously proposed [6] single diacritic technique. This hiding of secret data into two Diacritics, in the stego-cover text, is chosen based on the percentage study of highest number of Diacritics available among all found Diacritics. The paper exploits the possibility of hiding data in two Diacritics, i.e. Fathah and Kasrah, modifying the previously presented single (Fathah only) diacritic hiding scheme [6]. This study shows how this proposed two Diacritics stego-work is featuring higher capacity and security with acceptable robustness, showing interesting promising results.

In the following section, Section 2, we present some background information about Arabic language script. Then in Section 3, we review the related work from the literature about Arabic script text steganography focusing

doi: 10.18178/lnit.3.1.42-47

on utilizing Diacritics. Section 4, presents the proposed two Diacritics approach showing our organized algorithm and software program to increase capacity and security. Next, Section 5 provides some testing elaboration on the reasons and method behind choosing these two Diacritics, i.e. Fathah and Kasrah. This section also provides the comparison between our proposed work of utilizing two Diacritics with the previously presented single Diacritics stego approach in [6]. Afterwards, Section 6 summarizes the work in the conclusion giving ideas of possible future work.

II. Background on Arabic Script

More than 420 million people worldwide speak Arabic, making it the sixth most spoken language [7]. In fact, the phrase "Arab" means "nomad," which makes sense considering Arabic originated from nomadic tribes in the desert districts of the (Middle East) Arabian Peninsula. Arabic language script evolved from Nabataean Aramaic script and has been used since the 4th century CE [8]. It includes 28 main letters, similar to the English alphabet, but written in opposite direction from right to left, in a cursive style very similar to Urdu and Farsi scripts (although their number of letters are a bit different). Arabic is considered preserved since the 7th century CE through the Prophet Muhammad's (peace be upon him) revelations recorded in the Holy Qur'an [7]. By the 8th Century CE, as many people converted to Islam, Arabic Language began spreading throughout the Middle East and North Africa. By religion, all Muslims are to use Arabic in reciting the Holy Quran in their obliged 5 times daily prayers.

Today, the Arab world is composed of all countries in the Middle East and North Africa, where Arabic is their native and official language, i.e. Arabic is found the native official language in 19 countries and co-official (second) language in 7 countries [7]. Arabic Classical language is preserved by the holy Quran, which is the sacred book of Muslims around the world [8].

◌َ (Fathah) فَتحة	◌ِ (Kasrah) كَسرَة	◌ُ (Damah) ضَمَّة	◌ْ (Sukun) سُكُون
◌ً تَنوين فَتحة (Tanwin Fathah)	◌ٍ تَنوين كَسرَة (Tanwin Kasrah)	◌ٌ تَنوين ضَمَّة (Tanwin Damah)	◌ّ (Shaddah) شدّة

Figure 2. Arabic text diacritics

Arabic language has some features that are unique to most other languages, including English [7]. Beside its direction from right to lift and cursive style, every Arabic character has different shape depending on its position within the word. In addition, it has many pointed (doted) letters, with one, two, or three dots on top or bottom of some characters [8]. Also, Arabic characters can have some extra additional shapes located top or bottom of the letter characters called "Diacritics"- known in Arabic as "harakat," which are eight shapes (Harakat) originally added for representing the vowel sounds only [9]. These eight Arabic text Diacritics (Harakat) are Fathah, Kasrah, Damah, Sukun, Tanwin Fathah, Tanwin Kasrah, Tanwin Damah, and Shaddah, as shown Fig. 2. These Diacritics are represented digitally inside the computer as separate (zero location) characters. Note that the use of Diacritics in Arabic is optional in practice and in modern standard Arabic writing; however, it is essential for the holy Quran and most religious and historical scripts [3].

III. Related Work

Several relative techniques found in the literature to serve Arabic text steganography [10]-[12]. For example, one of the first proposals found in this field was presented to depend on the points (dots) inherited in the Arabic, Urdu and Persian letters for hiding binary value [10]. The location of the point is slightly shifted up if the hidden bit value is one; otherwise, the location remains unchanged. This method didn't attract much attention although it can hide a large volume of information in Arabic text. It became high capacity in storing hidden bits due to the fact that Arabic language has 15 out of 28 letters having points, i.e. more than have the letters of Arabic language include dots (points). But the drawback is in the robustness such that the output font cannot be standard. Thus, the receiver will not be able to extract the secret message if the output font is not installed on his PC machine. Also, the hidden information can be lost in any retyping or OCR scanning process.

Other proposed Arabic stego-systems uses the redundant Arabic extension character "Kashida" for hiding secret bits [11]. The noted drawback of this idea is that Kashida character cannot be added at the beginning or ending of words. It can only be added between connected letters in the words. In the Kashida method presented in [12], the Kashida is linked to the pointed letters to hold secret bit one and the un-pointed letters to hold secret bit zero as shown in Fig. 3, as an example. Using this link to pointed characters enhances the features of security and robustness. But, have drawbacks in capacity of the cover medium if the size of secret bits in the secret object is large.

Secret bits	110010
Cover-text	من حسن اسلام المرء تركه مالا يعنيه
Steganographic text	من حسن اسلام المرء تركـه مـالا يـعنـيه 1 1 0 0 1 0

Figure 3. Hiding secret bits using Kashida character

Using Diacritics for steganography was interestingly presented in the paper "Arabic Text Steganography Using Multiple Diacritics" in 2008 [3]. The main idea of work [3] was to set multiple invisible instances of Arabic Diacritic marks over each other. That was possible because of the way in which Diacritic marks are displayed on screen and printed to paper. These multiple Diacritics are zero location characters and not visible on screen nor printed paper, giving interesting advantage in terms of the security and specifically security capacity [3]. Surprisingly, this method have been tested showing

inconsistence unexpected results. We thoroughly studied the technique of [3] finding that it didn't operate successfully. In fact, it was noted the new versions of Arabic Text Software Editors completely prevented multiple invisible instances of Arabic Diacritic marks over each other considering them as error.

The related effective Diacritics approach was presented in [6]. We found this work promising for improving to propose our two Diacritics Arabic text steganography. The presented work in [6] considered studying the eight Diacritics (shown previously in Figure 2), assigning Diacritic Fathah, to hide the secret bit value equals "one," and the remaining seven Diacritics were assigned to hide "zero". The implementation of this Diacritic stego-method starts by sensing the first secret bits, if it was a "one," the system searches for the first Fathah Diacritic in the cover media to hide in it. However, if the Diacritic is not Fathah and the bit value is "one," the Diacritic is removed from the cover media and, in the same time, the index of the cover media is incremented only to read the next Diacritic. The same process is implemented to hide bit value "zero," except that "zero" will search for all the other seven Diacritics instead of the Fathah. An example of hiding E7=11100111 using this approach is shown in Fig. 4 [6]. Normally, the receiver has to have the original text so that the extracting algorithm can compare the Diacritics in the stego object with the original cover object to extract the secrets.

Cover Object	حَدَّثَنَا سُفْيَانُ عَنْ يَحْيَى
Secret Object	E7 (= 11100111)
Stego Object	حَدَّثَنَا سُفْيَانَ عَنْ يَحْيَى

Figure 4. Hiding (E7=11100111) using Diacritics approach [6]

The main advantage to choose this Diacritics method [6] for further study is found that it fulfilled all stego features adequately, i.e. ample capacity, good robustness and reasonable security. The idea is well clarified in [6] if additional elaboration is required, making it the choice of our improvement work in this paper. Fig. 5 below shows our example of using technique [6] to hide secret C9=11001001 in a historical poem phrase.

Before encryption
عَفَتِ الدِّيَارُ مَحَلُّهَا فَمُقَامُهَا بِمِنىً تَأَبَّدَ غَوْلُهَا فَرِجَامُهَا
After encryption
عَفَتِ الدِّيَارُ مَحَلُّهَا فَمُقَامُهَا بِمِنىً تَأَبَّدَ غَوْلُهَا فَرِجَامُهَا

Figure 5. Example of hiding (C9=11001001) using Diacritics approach [6]

IV. Proposed Two Diacritics Approach

Arabic text steganography is considered virgin field for this kind of security research [9], [12]. Our proposed two Diacritics algorithm is to hide secret binary data into Arabic text improving the previous single Diacritics scheme presented in [6]. The common assumption in these Diacritics stego technique is that the cover text is found full of Diacritics [3]. This assumption is normal and found common in Arabic scripts [7], [8]. Our method focuses on hiding in 'Fathah' and 'Kasrah' together providing increased capacity and security compared to the previous 'Fathah' only scheme. The two diacritics are adding capacity and we propose dealing with even and odd bits separately in the secret message to provide more security, i.e. making extraction of the secret message for eavesdropper not to be easy.

Our proposed two Diacritics stego algorithm can be summarized in 10steps as the following:

1. Insert the cover-text with Diacritics.
2. Insert the secret message (it is sequence of binary bits 0's and 1's).
3. Store it separated in two arrays as odd array and even array.
4. Check the value in odd array index with 'Fathah'.
   a. If it is '1' we keep the 'Fathah' as is.
   b. If it is '0' we remove the 'Fathah'.
5. Increment odd array index since it is holding secret bits.
6. Repeat going to step 4 until odd array ends.
7. Check the value in even array index with 'Kasrah'.
   a. If it is '1' we keep the 'Kasrah' as is.
   b. If it is '0' we remove it.
8. Increment even array index since it is holding secret bits.
9. Repeat going to step 7 until even array ends.
10. Finally, show the text before/after the stego process, i.e. before/after encryption.

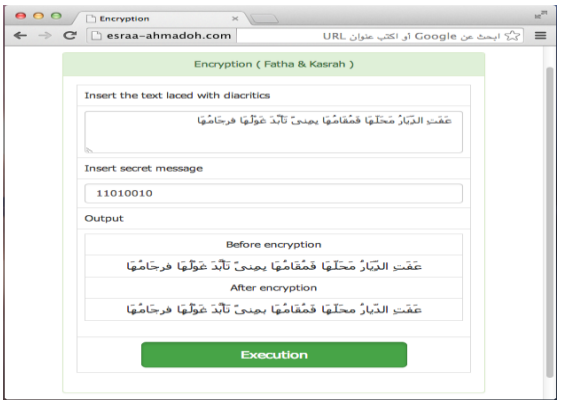

Figure 6. Interface of the proposed two Diacritics stego approach implementation

The proposed algorithm starts by splitting the secret message into two arrays of binary values as odd and even lists. The overview can be looked as the odd array list to be hidden in the 'Fathah' diacritics and the even list to be stored in the 'Kasrah' Diacritics. This splitting is performed for extra security added within our proposed method compared to the single Diacritic approach in [6]. Our program reads the first odd bit of the secret message and then compares it with the first 'Fathah' in the cover text. If, for example, the first odd bit to be hidden was a 'one', this first 'Fathah' will remain; otherwise, the 'Fathah' will be removed. This process will repeat itself until all secret bits in the odd array are considered. Similarly, the program reads the even bits array and

applies its' effect on the 'Kasrah' disappearing or existence. The algorithm is programmed by php language using UTF-8 Encoding due to its convenience fully supporting Arabic text [13]. Fig. 6 shows the interface of our implementation of the algorithm.

V. Experimentations and Comparisons

As previously described, Arabic language uses Diacritics in holy Quran and religious and historical scripts [3]. To run the algorithm appropriately, we use historical Arabic poetry [14], which is Arabic stego cover text [15] that is full of Diacritics [16].

We start our study by choosing one of the most famous historical poems in Arabic poetry Known as part of the Mu'allaqat focusing on what is written by the poet (Labīd ibn Rabī'ah). Our experimentation is selecting the secret bits to be hidden and then storing them in the cover text using the single Diacritics scheme as well as the two Diacritics approach to observe the differences, as shown in Fig. 7.

Secret Bits Hidden	Single Diacritics previous scheme [6]	Two Diacritics proposed approach
9 Bits 100010110	Before encryption عَفَتِ الدِّيَارُ مَحَلُّهَا فَمُقَامُهَا بِمِنىً تَأَبَّدَ غَوْلُهَا فَرِجَامُهَا After encryption عَفتِ الدِّيارُ مَحَلُّها فَمُقَامُها بِمِنىً تَأَبَّدَ غَوْلُهَا فَرِجَامُهَا	Before encryption عَفَتِ الدِّيَارُ مَحَلُّهَا فَمُقَامُهَا بِمِنىً تَأَبَّدَ غَوْلُهَا فَرِجَامُهَا After encryption عفتِ الدّيارُ مَحَلُّها فَمُقَامُها بِمِنىً تَأَبَّدَ غَوْلُهَا فَرِجَامُهَا
	Before encryption عَفَتِ الدِّيَارُ مَحَلُّهَا فَمُقَامُهَا بِمِنىً تَأَبَّدَ غَوْلُهَا فَرِجَامُهَا After encryption عَفتِ الدِّيارُ مَحَلُّها فَمُقَامُها بِمِنىً تَأَبَّدَ غَوْلُهَا فَرِجَامُهَا	Before encryption فَمَدَافِعُ الرَّيَّانِ عُرِّيَ رَسْمُهَا خَلِقاً كَمَا ضَمِنَ الوُحِيَّ سِلامُهَا After encryption فمدافِعُ الرَّيَّان عُرِّيَ رَسْمُهَا خَلِقاً كَمَا ضَمِنَ الوُحِيَّ سِلامُهَا
20 Bits 10110010101010010010	Before encryption عَفَتِ الدِّيَارُ مَحَلُّهَا فَمُقَامُهَا بِمِنىً تَأَبَّدَ غَوْلُهَا فَرِجَامُهَا After encryption عَفتِ الدِّيَارُ مَحلُّها فَمُقامُها بِمِنىً تأبّدَ غَوْلُها فَرِجامُها	Before encryption عَفَتِ الدِّيَارُ مَحَلُّهَا فَمُقَامُهَا بِمِنىً تَأَبَّدَ غَوْلُهَا فَرِجَامُهَا After encryption عفتِ الدِّيارُ محلُّها فمُقامُها بِمِنىً تَأَبَّدَ غَوْلُهَا فَرِجَامُهَا
	Before encryption عَفَتِ الدِّيَارُ مَحَلُّهَا فَمُقَامُهَا بِمِنىً تَأَبَّدَ غَوْلُهَا فَرِجَامُهَا After encryption عَفتِ الدِّيَارُ مَحلُّها فَمُقامُها بِمِنىً تأبّدَ غَوْلُها فَرِجامُها	Before encryption فَمَدَافِعُ الرَّيَّانِ عُرِّيَ رَسْمُهَا خَلِقاً كَمَا ضَمِنَ الوُحِيَّ سِلامُهَا After encryption فمَدافِعُ الرَّيّانِ عُرِّيَ رسْمُهَا خلِقاً كمَا ضَمِنَ الوُحِيَّ سِلامُهَا
30 Bits 101100101111001010101001010010	Before encryption عَفَتِ الدِّيَارُ مَحَلُّهَا فَمُقَامُهَا بِمِنىً تَأَبَّدَ غَوْلُهَا فَرِجَامُهَا After encryption عَفتِ الدِّيَارُ مَحلُّها فَمُقامُها بِمِنىً تَأَبَّد غَوْلُهَا فرِجَامُها	Before encryption عَفَتِ الدِّيَارُ مَحَلُّهَا فَمُقَامُهَا بِمِنىً تَأَبَّدَ غَوْلُهَا فَرِجَامُهَا After encryption عفتِ الدِّيارُ محَلُّهَا فمُقامُها بمِنىً تَأَبَّدَ غَوْلُها فَرِجَامُهَا
	Before encryption دِمَنٌ تَجَرَّمَ بَعْدَ عَهْدِ أَنِيسِهَا حِجَجٌ خَلَونَ حَلالُهَا وَحَرامُهَا After encryption دِمَنٌ تجَرَّم بعْدَ عهْدِ أَنِيسِهَا حِجَجٌ خَلون حَلالُها وَحرامُهَا	Before encryption رُزِقَتْ مَرَابِيْعَ النُّجُومِ وَصَابَهَا وَدَقُّ الرَّوَاعِدِ جَوْدُهَا فَرِهَامُهَا After encryption رُزِقتْ مَرابِيْع النُّجُوم وَصَابها ودقُّ الرّوَاعِدِ جَوْدُها فرِهَامُهَا
50 Bits 10010011100101110011010010101001010010011001	وَالعِيْسُ سَاكِنَةٌ عَلَى أَطْلالِهَا عُوداً تَأَجَّلُ بِالفَضَاءِ بِهَامُهَا وَجَلا السُّيُولُ عَنِ الطُّلُولِ كَأَنَّهَا زُبَرٌ تُجِدُّ مُتُونَهَا أَقْلامُهَا أَوْ رَجْعُ وَاشِمَةٍ أُسِفَّ نُؤُورُهَا كَفِفاً تَعَرَّضَ فَوْقَهُنَّ وِشَامُهَا فَوَقَفْتُ أَسْأَلُهَا وَكَيفَ سُؤَالُنَا صُمَّ الخَوَالِدَ مَا يَبِيْنُ كَلامُهَا After encryption وَالعِيْسُ سَاكِنَةٌ عَلَى أَطْلالِهَا عُوداً تَأَجَّلُ بِالفَضَاءِ بِهَامُهَا وَجَلا السُّيُولُ عَنِ الطُّلُولِ كَأَنَّهَا زُبَرٌ تُجِدُّ مُتُونَهَا أَقْلامُهَا أَوْ رَجْعُ وَاشِمَةٍ أُسِفَّ نُؤُورُهَا كَفِفاً تَعَرَّضَ فَوْقَهُنَّ وِشَامُهَا فَوَقَفْتُ أَسْأَلُهَا وَكَيفَ سُؤَالُنَا صُمَّ الخَوَالِدَ مَا يَبِيْنُ كَلامُهَا	وَالعِيْسُ سَاكِنَةٌ عَلَى أَطْلالِهَا عُوداً تَأَجَّلُ بِالفَضَاءِ بِهَامُهَا وَجَلا السُّيُولُ عَنِ الطُّلُولِ كَأَنَّهَا زُبَرٌ تُجِدُّ مُتُونَهَا أَقْلامُهَا أَوْ رَجْعُ وَاشِمَةٍ أُسِفَّ نُؤُورُهَا كَفِفاً تَعَرَّضَ فَوْقَهُنَّ وِشَامُهَا فَوَقَفْتُ أَسْأَلُهَا وَكَيفَ سُؤَالُنَا صُمَّ الخَوَالِدَ مَا يَبِيْنُ كَلامُهَا After encryption وَالعِيْسُ سَاكِنَةٌ عَلَى أَطْلالِهَا عُوداً تَأَجَّلُ بِالفَضَاءِ بِهَامُهَا وَجَلا السُّيُولُ عَنِ الطُّلُولِ كَأَنَّهَا زُبَرٌ تُجِدُّ مُتُونَهَا أَقْلامُهَا أَوْ رَجْعُ وَاشِمَةٍ أُسِفَّ نُؤُورُهَا كَفِفاً تَعَرَّضَ فَوْقَهُنَّ وِشَامُهَا فَوَقَفْتُ أَسْأَلُهَا وَكَيفَ سُؤَالُنَا صُمَّ الخَوَالِدَ مَا يَبِيْنُ كَلامُهَا
	عَرِيتْ وَكَانَ بِهَا الجَمِيْعُ فَأَنكَرُوا مِنْهَا وُعُورِ نُؤْيِهَا وَتُمَامُهَا شَاقَتْكَ طُعْنُ الحَيِّ حِيْنَ تَحَمَّلُوا فَتَكَشَّفُوا قُطُناً تَصِرُّ جَيَامُهَا مِنْ كُلِّ مَحْفُوفٍ يُظِلُّ عَصِيَّهُ زَوْجٌ عَلَيْهِ كِلَّةٌ وَقِرَامُهَا After encryption عَرِيتْ وَكَانَ بِهَا الجَمِيْعُ فَأَنكَرُوا مِنْهَا وُعُورِ نُؤْيِهَا وَتُمَامُهَا شَاقَتْكَ طُعْنُ الحَيِّ حِيْنَ تَحَمَّلُوا فَتَكَشَّفُوا قُطُناً تَصِرُّ جَيَامُهَا مِنْ كُلِّ مَحْفُوفٍ يُظِلُّ عَصِيَّهُ زَوْجٌ عليْهِ كِلَّةٌ وَقِرَامُهَا	Before encryption رَحَلاً كَأَنَّ بِعَاجَ نُوْصِحَ فَوْقَهَا وَظِبَاءَ وَخْرَةَ عُطَّفاً أَرَامُهَا خُوِرَتْ وَرَابِلَهَا السُّرَاةُ كَأَنَّهَا أَجْزَاءُ بِيشَةَ أَثْلُهَا وَرِضَامُهَا بَلْ مَا تَذَكَّرُ مِنْ نَوَارٍ وقَدْ نَأَتْ وَتَقَطَّعَتْ أَسْبَابُهَا وَرِمَامُهَا After encryption رَحلاً كَأنَّ بِعاجَ نُوْصِحَ فَوْقَهَا وَظباء وَخْرةَ عُطَّفاً أرامُها خُوِرَتْ وَرَابِلَهَا السُّرَاةُ كَأَنَّهَا أَجْزَاءُ بِيشَةَ أَثْلُهَا وَرِضَامُهَا بَلْ مَا تَذَكَّرُ مِنْ نَوَارٍ وقَدْ نَأَتْ وَتَقَطَّعَتْ أَسْبَابُهَا وَرِمَامُهَا

Figure 7. Several sizes of secret bits hidden in cover texts using the different stego techniques

This Arabic Mu'allaqat poetry [14] written by the poet (Labīd ibn Rabī'ah) was studied in terms of the number of Diacritics it contains. Table I summarizes the number of Diacritics and its percentage within this poem (Labīd ibn Rabī'ah) that proofed our choice of Fathah and Kasrah. The program was then tested for the seven

famous poems in Arabic Mu'allaqat poetry, i.e. similar other Arabic poet texts with Diacritics, showing comparable results, as listed in Table II.

All the seven famous poems in Arabic Mu'allaqat poetry have been studied in terms of Capacity assuming both techniques, i.e. single Diacritics of [6] and our proposed two Diacritics schemes as shown in Fig. 8.

TABLE I. NUMBER OF DIACRITICS AND ITS PERCENTAGE IN THE POEM OF (LABĪD IBN RABĪ'AH)

Diacritics	Fathah	Kasrah	Damah	Sukun	Tanwin Fahah	Tanwin Kasrah	Tanwin Damah	Shaddah
Number of Diacritics	1355	446	406	355	36	62	27	175
Percentage within all Diacritics	47%	16%	14%	12%	1%	2%	1%	6%

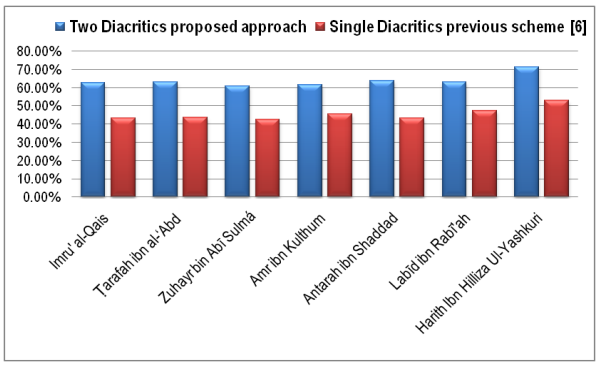

Figure 8. Comparing capacity percentages within the poems

It was expected that whatever poems giving good capacity percentage in single Diacritics scheme would give similar good results in two Diacritics method, as the found the best for what the poet (Harith Ibn HillizaUl-Yashkuri) [14]. Strangely, some poems did not workout consistent, i.e. the poem of (Antarah ibn Shaddad) gave higher capacity using our two diacritics when compared to the poem of (Amr ibn Kulthum) opposite to its capacity results using the single Diacritics method. This result can be observed when comparing several other poems, which is proofing the complete data dependency effect in showing the results. In our experiments, we choose poem of (Harith Ibn HillizaUl-Yashkuri) to give the highest capacity, which turned out providing the acceptable security too, as shown in Table III.

TABLE II. NUMBER OF DIACRITICS IN THE SEVEN FAMOUS POEMS IN ARABIC MU'ALLAQAT POETRY

Number of Diacritics	Imru' al-Qais	Ṭarafahibn al-'Abd	Zuhayr bin AbīSulmá	AmribnKulthum	AntarahibnShaddad	LabīdibnRabī'ah	HarithIbnHillizaUl-Yashkuri
Fathah	1209	1528	953	1451	936	1355	1041
Kasrah	548	687	412	510	444	446	358
Damah	294	354	257	352	249	406	311
Sukun	452	579	413	570	287	355	19
Shaddah	175	220	126	187	155	175	129
TanwinFathah	26	33	29	47	25	36	21
TanwinKasrah	86	78	49	52	48	62	40
TanwinDamah	8	35	9	15	19	27	39

VI. CONCLUSION

This paper presents an improved Arabic text steganography approach using two Diacritics ('Fathah' and 'Kasrah') to hide information as a new version improving the single (only 'Fathah') Diacritic method presented before. This utilization of Diacritics - or Harakat - for security purposes is found very useful for the text written originally with Diacritics such as religious Arabic or historical books.

The new two Diacritics method program was tested and compared with the single Diacritics technique run over the historical Arabic the seven famous poems well known as Arabic Mu'allaqat poetry. The results showed interesting higher general performance, but with some unexpected strange capacity observations proofing the effect of data dependency. Our proposed method showed quality of higher capacity and security with acceptable robustness. All results gave interesting promising directions opening the door for more studies to consider different other Diacritics for hiding information as well as testing more on different other historical data sets.

TABLE III. COMPARING DIFFERENT FEATURES OF THE STEGO TECHNIQUES

	Capacity	Security	Robustness
Two Diacritics proposed approach	71.45%	71.45%	28.55%
Single Diacritics previous scheme [6]	53.16%	53.16%	46.48%

ACKNOWLEDGMENT

The authors wish to thank the graduate program information security track offered by the College of Computer and Information Systems at Umm Al-Qura University, Makkah, Saudi Arabia, for supporting this research work.

REFERENCES

[1] A. Odeh and K. Elleithy,"Steganography in arabic text using zero width and kashidha letters," *International Journal of Computer Science & Information Technology*, vol. 4, no. 3, June 2012.

[2] S. S. Mohammad, "A new persian/arabic text steganography using 'La' word," *Advances in Computer and Information Sciences and Engineering*, pp. 339-342, 2008.
[3] A. Gutub, Y. Elarian, S. Awaideh, and A. Alvi, "Arabic text steganography using multiple diacritics," in *Proc. WoSPA 2008 - 5th IEEE International Workshop on Signal Processing and Its Applications*, University of Sharjah, Sharjah, UAE, 18-20 March 2008.
[4] D. Vitaliev, "Digital security and privacy for human rights defenders," *The International Foundation for Human Right Defenders*, pp. 77-81, Feb. 2007.
[5] S. Mersal, S. Alhazmi, R. Alamoudi, and N. Almuzaini, "Arabic text steganography in smartphone," *International Journal of Computer and Information Technology*, vol. 3, no. 2, pp. 441-445, March 2014.
[6] A. Mohammed, S. Awaideh, A. R. Elshafei, and A. Gutub, "Arabic diacritics based steganography," in *Proc. IEEE International Conference on Signal Processing and Communications*, Dubai, UAE, 24-27 November 2007, pp. 756-759.
[7] S. Ridout. (January 2015). Complete List of Arabic Speaking countries 2014. [Online]. Available: http://istizada.com/complete-list-of-arabic-speaking-countries-2014/
[8] A. G. Chejne, *The Arabic Language: Its Role in History*, Minneapolis: University of Minnesota Press, 1969.
[9] F. Al-Haidari, A. Gutub, K. Al-Kahsah, and J. Hamodi, "Improving security and capacity for Arabic text steganography using 'Kashida' extensions," in *Proc. AICCSA 2009 - The 7th ACS/IEEE International Conference on Computer Systems and Applications*, Rabat, Morocco, 10-13 May 2009, pp. 396-399.
[10] S. S. H. Mohammad and S. S. Mohammad, "A new approach to Persian/Arabic text steganography," in *Proc. IEEE/ACIS International Conference on Computer and Information Science*, 2006, pp. 310-315.
[11] A. Gutub and A. Al-Nazer, "High capacity steganography tool for arabic text using 'Kashida'," *The ISC Int'l Journal of Information Security*, vol. 2, no. 2, pp. 109-120, July 2010.
[12] A. Gutub, F. Al-Haidari, K. Al-Kahsah, and J. Hamodi, "e-Text watermarking: Utilizing 'Kashida' extensions in arabic language electronic writing," *Journal of Emerging Technologies in Web Intelligence*, vol. 2, no. 1, pp. 48-55, February 2010.
[13] K. Al-Sham'aa. (Jan. 2015). Ar-PHP: PHP Talks Arabic – be ready. [Online]. Available: http://www.ar-php.org/faq-php-arabic.html
[14] Poet. Harith Ibn Hilliza Ul-Yashkuri. (Jan. 2015). [Online]. Available: http://oldarab.jo1jo.com/7lzah.htm
[15] International Language Institute (ILI). (6 February 2014). Facts about the Arabic Language. [Online]. Available: http://www.arabicegypt.com/news/facts-about-the-arabic-language
[16] M. G. Vennice, T. Rao, M. Swapna, and J. S. kiran, "Hiding the text information using steganography," *International Journal of Engineering Research and Applications*, vol. 2, no. 1, pp. 126-131, Jan.-Feb. 2012.

Esraa Mohammad Ahmadoh is currently a graduate student, pursuing Master of Sciences (MS) degree in Computer Sciences & Engineering, at Umm Al Qura University (UQU) under the umbrella of Ministry of Higher Education. Her MS program at UQU is specialized in the information security track offered by the College of Computer and Information Systems offered at UQU-Makkah Campus, Saudi Arabia.

Prof. Adnan Abdul-Aziz Gutub is currently working as the Vice Dean of the Custodian of the Two Holy Mosques Institute of the Hajj & Omrah Research, within Umm Al Qura University (UQU), Makkah -Saudi Arabia.
Adnan is ranked as Professor in Computer Engineering specialized in Information and Computer Security within UQU. His experience was gained from his previous long-time work in Computer Engineering at King Fahd University of Petroleum and Minerals (KFUPM) in Dhahran, Saudi Arabia. He received his Ph.D. degree (2002) in Electrical & Computer Engineering from Oregon State University, USA. He had his BS in Electrical Engineering and MS in Computer Engineering both from KFUPM, Saudi Arabia.
Adnan's research interests involved optimizing, modeling, simulating, and synthesizing VLSI hardware for crypto and security computer arithmetic operations. He worked on designing efficient integrated circuits for the Montgomery inverse computation in different finite fields. He has some work in modeling architectures for RSA and elliptic curve crypto operations. His current interest in computer security also involved steganography such as image based steganography and Arabic text steganography.
In summer 2013, Adnan has been awarded 3-month visiting scholar grant in collaboration with Purdue University, West Lafayette, Indiana, USA. He had been involved in research of current studies related to Arabic Text Steganography in Data Security as well as Elliptic Curve Crypto Processor Designs. Previously, Adnan have been twice awarded the UK visiting internship for 2 months of summer 2005 and summer 2008, both sponsored by the British Council in Saudi Arabia. The 2005 summer research visit was at Brunel University to collaborate with the Bio-Inspired Intelligent System (BIIS) research group in a project to speed-up a scalable modular inversion hardware architecture. The 2008 visit was at University of Southampton with the Pervasive Systems Centre (PSC) for research related to text steganography and data security.
Administratively, Adnan Gutub filled many executive and managerial academic positions at KFUPM as well as UQU. At KFUPM - Dhahran, he had the experience of chairing the Computer Engineering department (COE) for five years until moving to Makkah in 2010. Then, at UQU - Makkah, Adnan Chaired the Information Systems Department at the College of Computer & Information Systems followed by his leadership of the Center of Research Excellence in Hajj and Omrah (HajjCoRE) serving as HajjCoRE director for around 3-years until the end of 2013. Then, he was assigned his current position as the Vice Dean of HRI, i.e. the Custodian of the Two Holy Mosques Institute of the Hajj & Omrah Research.