A Hybrid Cryptography Method using Extended Letters in Arabic and Persian Language

Ke Wang*
School of Software Engineering, Jilin Technology College of Electronic Information, Jilin 132000, Jilin, China

***Abstract*—Cryptography is widely used in information security systems. In encryption, the goal is to hide information in such a way that only the sender and receiver are aware of the existence of communication and information. Encryption takes place in various media, such as image, sound and text. Today, the rapid growth of network technologies and digital tools has made digital delivery fast and easy. However, the distribution of digital data in public networks such as the Internet has various challenges due to copyright infringement, forgery, coding and fraud. Therefore, methods of protecting digital data, especially sensitive data, are very necessary. Accordingly, in this article, a combined solution is used based on the technique of stretching the letters and making minor changes in the letters that have closed spaces so that the bits related to the hidden text can be inserted into a Persian or Arabic language. For this purpose, a new solution has been designed, in which the cover text is similar to the normal text, with the difference that, in addition to the extended letters that are longer due to the status of the secret message, it also has some prepositions, which have spaces. They are empty and closed. Of course, this difference in the closed space between the original letters and the changed letters will be very slight, and as a result, there will not be much difference between them that the normal user can feel the change. Finally, the proposed solution has been evaluated with the help of the MATLAB program, and according to the rate parameter of encryption capacity, the results show that the proposed method has an average encryption capacity of more than 50% compared to other common solutions.**

Keywords—Cryptography; extended letters; Persian language; Arabic language

I. Introduction

Text-in-text encryption is used for the purpose that the host data does not have any visible signs of the presence of encrypted data in it, and thus it enables the secure transmission of information [1]. Therefore, the way to hide the data is important according to the changes that are made in the external details of the text. But what is of secondary importance is to protect the transmission against widespread attacks on the data sent and, of course, to make the original data unreachable in case the attackers discover the existence of a password in the host data. In fact, the advantage of using text as a cover media is that the text file consumes less storage space and less bandwidth is required for its transmission [2, 29, 30]. In order to encrypt a text, watermarking methods can be used in the text image, shifting the words in the middle space between them or shifting the background line to hide the data of the desired series in the background text, but this method which is considered in this research is the use of making changes in details. The majority of steganography research employs various types of cover media, such as images [3, 31], video clips [4, 32], and audio [5, 33]. Nevertheless, text steganography is not commonly favored due to the challenges associated with identifying redundant bits in text files [6, 7, 34]. Text documents typically exhibit a structure similar to what is visible, while other types of cover media have different structures, making it easier to hide information without noticeable alterations. However, text steganography offers advantages like lower memory usage and simpler communication.

The main goal of this project is to provide a new method for hiding information in Persian and Arabic texts. The choice of steganographic system often depends on the language and its structures. One technique cannot be universally applied to all languages. In the proposed method, it has been tried to avoid making changes in the appearance and format of the font as much as possible so that while increasing the capacity, the level of transparency and readability of the text does not change significantly compared to the state before hiding. So in this method, the technique of stretching Persian and Arabic letters is used in addition to creating changes in the letters that have closed empty space, which includes the letters: "ص،ض،ط،ظ،ع،غ،ف،ق،م،و،ه" covert writing of information in Persian and Arabic texts is discussed. Of course, for the letters "ع, غ" only when these letters are used in the middle of the word, they have a closed space. Therefore, in this method, only one position of the letters "ع, غ" (in the middle of the word) can be used to hide the information in the letters "ع, غ".

This paper is organized as follows. Section I provides an overview for text steganography. Section II discusses the related works. Section III presents our new text steganography technique that utilizing character extension. The evaluation and results are discussed in Sections IV and V, respectively and finally, conclusion and future work are outlined in Sections VI and VII.

A. Cryptography Basics

According to the type of application, several characteristics are considered to express the efficiency of cryptographic methods, which are different in different applications. The important things considered in the design of this method are as follows [3]:

- Transparency: In the encryption operation, it is necessary to hide the information in the desired text in such a way that the inserted information is not visually clear and understandable. This amount of similarity, before and after inserting information, is called transparency.

- Resistance: In a cryptography solution, resistance means against unwanted and unintentional changes that are created due to the presence of noise in the transmission channel or other changes that are intentionally made by an attacker, and it is done in the direction of changing the message so that it has the necessary resistance to a large extent.
- Capacity: The amount of information that can be included in the cover text is called capacity. The methods that are presented for the purpose of encryption should be designed in such a way that they are able to hide a significant amount of information in the cover text.

Considering that in order to encrypt texts, letters and characters need to be converted into a set of numbers 0 and 1 so that they are easy to manage [8, 9, 35]. Therefore, there must be a single standard to determine what characters each of these numbers and letters should display and how they should be stored. This standard is called Unicode.

Converting data so that the system can read and use it is called encoding. In fact, encoding is the process of converting data into a format required for a number of information processing needs, including:

- Data encryption
- Formulation of the program and its implementation
- Data transfer, storage and compression/decompression
- Application data processing, such as file conversion

For the coding standard mentioned above, the American Standards Association introduced a 7-bit coding method called ASCII in 1960. At that time, the ASCII character set, including 128 characters (7 bits), was defined, mostly for Latin languages. ASCII encoding is important because many communication tools and protocols only accept ASCII characters. In fact, this is the accepted minimum standard for the text. Some Unicode encodings, due to the universal acceptance of ASCII, convert their code points into a series of ASCII characters so that they can be moved without any problems. In the 1980s, it was decided to use a full byte (i.e., 8 bits) for encoding in the ASCII character set instead of 7 bits [10, 36]. Therefore, the number of characters reaches 256. Based on this, the characters after 127 to 255 were also considered reserved codes, and other languages, including Persian and Arabic, were generally included in this range. But in this range between different languages, there was no single standard, and each language showed its alphabet code. In other words, code 200 in one language returns a different letter in another language. As a result, there was a need for a single standard to consider unique codes for each character while being compatible with all languages. Based on this, the Unicode unit standard has been presented. Unicode is a set of characters with unique numbers, which is called a code point. Each code point represents a single character. Accordingly, the Unicode standard specifies three encoding methods, allowing a character to be encoded within one or more bytes (i.e., in 8, 16, or 32 bits). These coding methods are as follows [11, 37]:

- UTF-16
- UTF-8
- UTF-32

The difference between these coding methods is in the way letters, numbers and symbols are presented between the languages of different countries. So that the way characters are presented in one country is different from another country [12, 38]. Table I shows the Unicode standard for Persian and Arabic letters, along with the code of each letter in this standard. In cryptography, this table is used to convert each letter to its corresponding value of 0 or 1.

TABLE I. UNICODE STANDARD FOR A NUMBER OF PERSIAN LETTERS, ALONG WITH THE CODE OF EACH LETTER

Form	letter name	Code	Form	letter name	Code
ت	letterت	062A	ء	letterحمزه	0621
ث	letterث	062B	آ	letterآ	0622
ج	letterجیم	062C	ا	letterالف	0627
چ	letterچ	0686	أ	letter الف withء	0623
د	letterدال	062F	ب	letterب	0628
ذ	letterذال	0630	پ	letterپ	067E

II. RELATED WORK

Numerous research studies have explored concealing one text within another, employing diverse linguistic techniques. These studies classify linguistic steganography into two primary categories: syntax-based and semantic-based approaches [13][14] [39].

For instance, [15] and [16] proposed a method based on synonyms, aimed at data concealment. The method consists of two phases: in phase one converting the hidden message into binary codes. In phase two, with the help of a synonyms file, both sender and recipient must possess the same word list for encryption and decryption. If the sender inserts a Zero, no word replacement is necessary. However, if another value is inserted, the synonym file serves as the basis for word replacement. This process continues until the end of the secret message, and the receiver can decrypt the message using an inverse strategy.

In [17] proposed method that takes three input sources (secret message, natural language, and the key) and produces one output known as the Stego-text. The system identifies each word and its corresponding group, generating lexical replacement groups and variant forms of similar words. To ensure proper embedding of the correct word in the carrier file while considering the context, a lexical analyzer for the Chinese language was utilized in the proposed system. The steganography algorithm, based on three inputs representing the source natural language text, the information to hide, and the key, generates one output displaying the stego-text with embedded data. The steps of the steganography algorithm include:

- Embedding data: Encrypting the information into a binary bit sequence using the key (embedded data).
- Text preprocessing: Segmentation of sentences using the Chinese lexical analyzer ICTCLAS, where English characters, Chinese and English punctuation, blank spaces, and new line characters are considered inter-sentence symbols.

In [18], a novel method was proposed to hide information not only in specific pointed letters but in any letters. The researchers utilized pointed letters with an extension to represent the secret bit 'one,' and un-pointed letters with an extension to hold the secret bit 'zero.' It was emphasized that letter extensions do not affect the writing content and are considered redundant characters for formatting purposes in Arabic electronic typing. However, not all letters can be extended due to their positions in words and the nature of Arabic writing. Extensions can only be added between connected letters, not at the end of words or before the beginning letter. Thus, letters without extensions or intentionally without extensions are considered not to hold any secret bits.

In [19], authors introduced a new approach for hiding information by manipulating white spaces between words and paragraphs. The method offered greater capacity to conceal more bits of data within a cover-text. While the previous embedding scheme used the space between words, it required a significant amount of space to encode a few bits. However, combining inter-word and inter-paragraph spacing allowed effective utilization of most white spaces in a text document, increasing the data-hiding capacity.

In [20], the authors built upon a previous Arabic text steganography method using letter points and extensions to address the low-capacity aspect. They proposed a technique to hide information within suitable positions inside words, not limited to pointed letters only. These positions were carefully determined to preserve the beauty of Arabic text if justified, ensuring that the message could be hidden without affecting the cover text. Extensions were inserted at the specified positions to represent the secret bit 'one,' while leaving the positions empty represented the secret bit 'zero.'

Another study [21] introduced the Line Shifting method, which involved shifting lines vertically to pass data via the carrier. However, this method had weaknesses, including the potential detection of line shifting when using character recognition programs and the risk of hidden data destruction when retyping the carrier file.

In [22], researchers adopted a syntactic method based on punctuation to convey the secret message. They added punctuation in suitable locations to hide the data without impacting the carrier message's meaning or affecting the embedded data within it. The study highlighted the importance of finding optimal trade-offs between bit rates, robustness, and perceivability through experimental definitions.

In [23], the researchers used emoticon-based steganography to facilitate secret chatting. They classified emoticons semantically and controlled the symbol order using a secret key to pass the secret message among parties. Different groups of emoticons were created to hide data, and the order of passing a symbol represented specific bit values. This method proved to be robust and beneficial for chat systems, with some systems allowing users to generate their customized symbols.

III. Proposed Method

In this research, a method of hiding in the text is presented so that by using it, confidential texts can be hidden inside the text. For this purpose, a combined solution is used based on the technique of stretching the letters and making minor changes in the letters with a closed, empty space so that the bits related to the hidden text can be hidden inside a normal hidden text. He made a drawing. For this purpose, a new solution has been designed, in which the cover text is similar to the normal text, with the difference that, in addition to the letters that are longer due to the status of the secret message, it also has some prepositions, which have spaces. They are empty and closed. Of course, this difference in the closed space between the original letters and the changed letters will be very slight, and as a result, there will not be much difference between them that the normal user can feel the change [24].

As shown in Fig. 1, the proposed encryption solution will have two inputs. One of its entries will be related to the cover text used for encryption, and the other entry will be the secret message that must be hidden inside the cover text. The proposed solution in the form of a new font acts as a converter and embedder, which can be used to change the cover text in such a way that the secret text can be hidden inside it. This solution will have the following features:

- High capacity to encrypt confidential text.
- High security due to very minor changes.
- High power in encryption.
- High transparency due to minor changes in font.

In fact, through the use of this combined solution, the capacity of confidential messages that must be encrypted can be increased because many Persian and Arabic letters are stretchable and have closed spaces. On the other hand, due to the fact that the changes in the letters are minor, normal users will not be able to recognize the desired changes, and as a result, security is established. Also, this solution is able to be resistant to the factors that can change the data and, as a result, damage the comprehensiveness because the encryption operation is carried out bit by bit according to the state of the letter. The original data is encrypted. On the receiver's side, there is a need to perform an operation in the direction of the photo, and for this purpose, a special tool is provided to detect the changed letters and, as a result, the hidden secret message. As a result, in this case, according to the status of each letter and its length, the main message can be extracted from the cover text, and there is no need for a private key for data exchange. Because in encryption algorithms, the encryption key must also be transmitted, and if for any reason this key is intercepted or discovered, all the encrypted data can be accessed, and as a result, the security and integrity of the data will be compromised [25].

A. How the Method Works

In Arabic and Persian languages, each character can have a different state according to its position. In fact, in this category of languages, a number of letters have the ability to change their shape according to the type of connection that they are at the beginning, end, or middle of a word, which is an important ability for encryption. For example, the letter "ع" at the beginning of the word is written as "عـ " in the middle as "معا " and at the end of the word as "تع " while in English words, each letter is written separately and does not have this feature. Fig. 2 shows a section of different states of Persian letters.

According to the Unicode standard, each letter can have a unique code according to its position. On the other hand, in some cases, in order to make a letter more beautiful, it can also be written with a stroke. For example, the word "خانه", which consists of four letters "خ", "ا", "ن" and "ه" can be written as "خـانه" without changing its meaning or the number of letters. In fact, stretching the letters is one of the advantages of Persian and Arabic languages; this ability does not have the slightest effect on the readability or changing the meaning of the text [26]. The only problem is that not all letters have this ability, and it is limited to the position in which that letter is placed; that is, in general, the ability to stretch can be applied to the connecting letters of a word, and this feature cannot be applied to a number of letters that are at the end or beginning of a word [12]. Accordingly, in this research, a combined solution has been used, which solves this limitation to a large extent. In the following, the method of using the ability to stretch the letter is shown with an example. At first, let's assume that we intend to perform the encryption process with the help of dragging on the text of Table II. First, we select the secret bits that should be hidden, which in this example is equal to the value (110010). The encryption operation is performed from the least valuable bits to the most valuable ones. Also, considering that the texts are Persian and Arabic and start from right to left; as a result, encryption is done from right to left. Accordingly, the first secret bit is equal to "0", and considering its value is zero, the letter to be hidden in it is extended once, but if it were one, it would be extended twice. Now, in this example, encryption should be done by stretching the letter "م" once. The value of the next secret bit is equal to "1", and the second letter of this word is "ن" considering that this letter is at the end of the word, so it cannot be extended in this position. The next point from which the operation can be continued is at the beginning of the word Turkey, and considering that the secret bit is equal to "1", as a result, the letter "ت" is doubled and changed to "تـــ ". Find this operation is carried out confidentially until the end of data encryption.

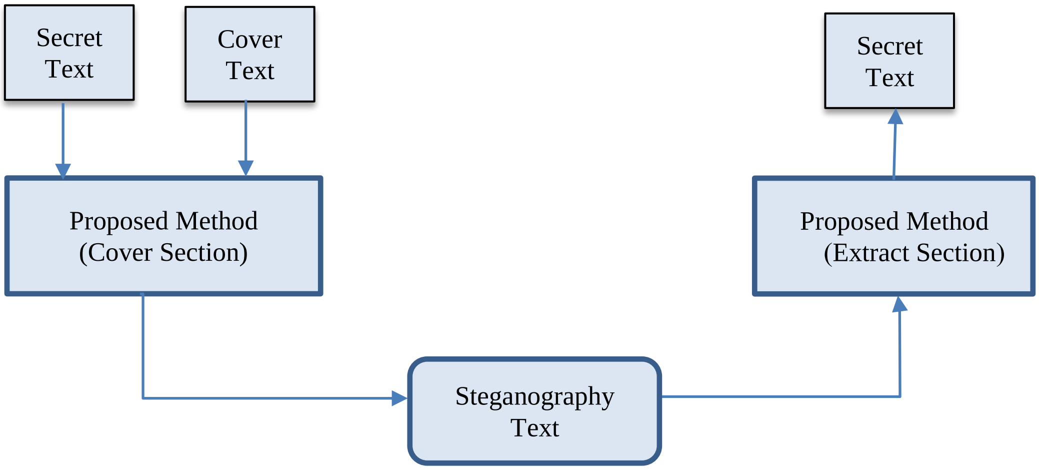
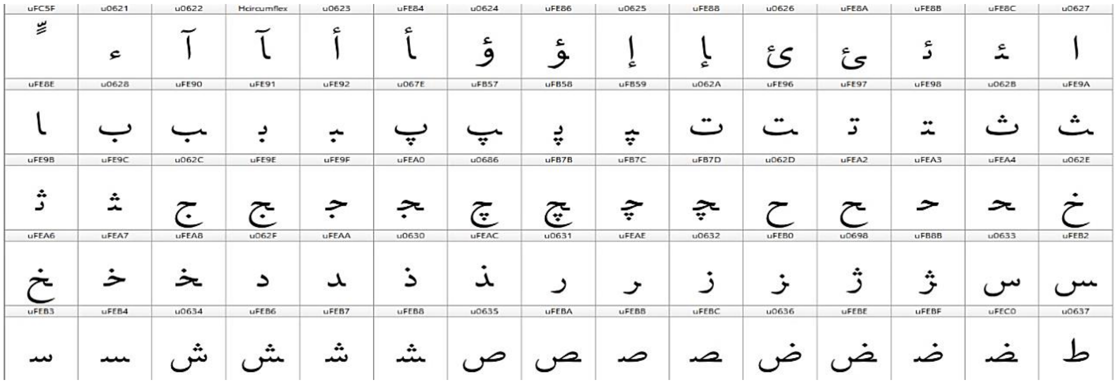

Secret Text
Cover Text
Proposed Method (Cover Section)
Steganography Text
Proposed Method (Extract Section)
Secret Text

Fig. 1. The general suggested method.



uFC5F	u0621	u0622	Hcircumflex	u0623	uFE84	u0624	uFE86	u0625	uFE88	u0626	uFE8A	uFE8B	uFE8C	u0627
ﱟ	ء	آ	ﺂ	أ	ﺄ	ؤ	ﺆ	إ	ﺈ	ئ	ﺊ	ﺋ	ﺌ	ا
uFE8E	u0628	uFE90	uFE91	uFE92	u067E	uFB57	uFB58	uFB59	u062A	uFE96	uFE97	uFE98	u062B	uFE9A
ﺎ	ب	ﺐ	ﺑ	ﺒ	پ	ﭗ	ﭘ	ﭙ	ت	ﺖ	ﺗ	ﺘ	ث	ﺚ
uFE9B	uFE9C	u062C	uFE9E	uFE9F	uFEA0	u0686	uFB7B	uFB7C	uFB7D	u062D	uFEA2	uFEA3	uFEA4	u062E
ﺛ	ﺜ	ج	ﺞ	ﺟ	ﺠ	چ	ﭻ	ﭼ	ﭽ	ح	ﺢ	ﺣ	ﺤ	خ
uFEA6	uFEA7	uFEA8	u062F	uFEAA	u0630	uFEAC	u0631	uFEAE	u0632	uFEB0	u0698	uFB8B	u0633	uFEB2
ﺦ	ﺧ	ﺨ	د	ﺪ	ذ	ﺬ	ر	ﺮ	ز	ﺰ	ژ	ﮋ	س	ﺲ
uFEB3	uFEB4	u0634	uFEB6	uFEB7	uFEB8	u0635	uFEBA	uFEBB	uFEBC	u0636	uFEBE	uFEBF	uFEC0	u0637
ﺳ	ﺴ	ش	ﺶ	ﺷ	ﺸ	ص	ﺺ	ﺻ	ﺼ	ض	ﺾ	ﺿ	ﻀ	ط

Fig. 2. Types of Persian letter forms.

TABLE II. THE METHOD OF ENCRYPTION USING THE ABILITY TO STRETCH LETTERS

Confidential Bits	**110010**
Cover text	من ترکیه را دوست ندارم
Hidden text	مـن تـرکیـه را دوسـت نـدارم

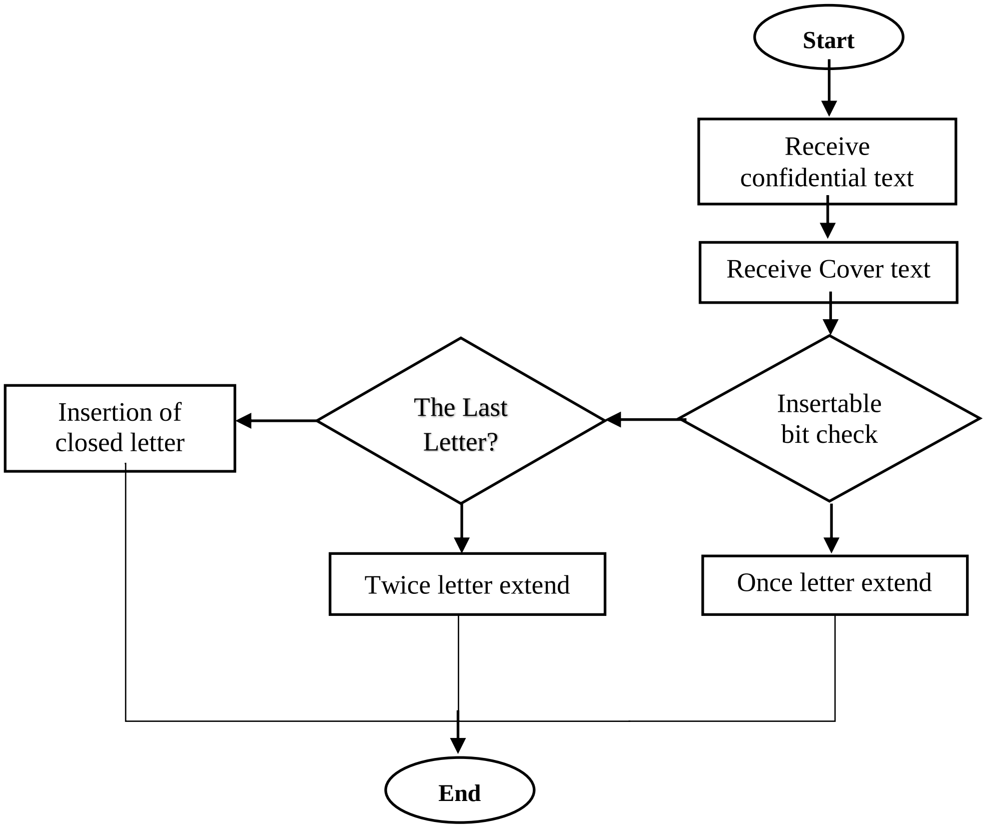


Fig. 3. Proposed method flowchart.

Now, in order to increase the efficiency of the solution, changes can be made in the letters with closed spaces so that the capability of this solution increases. The letters with closed space include the characters "ص،ض،ط،ظ،ع،غ،ف،ق،م،و،ه ". Of course, for the letters "ع،غ" only when these letters are used in the middle of the word will they have a closed space; therefore, in this method, only one position of the letters "ع،غ" (in the middle of the word) can be used to hide the information in the letters "ع،غ". Accordingly, in order to hide these letters, the amount of empty space is slightly smaller so that it has the least effect on changing the font of the desired letters and through this change if the last letter of a word cannot be it is drawn, it is part of the letters with closed space, it is used. As a result, if the insertable bit is zero, the character will remain unchanged. But if the insertable bit was 1, the character is inserted with a smaller enclosed space. In this case, the complete flowchart of the proposed solution is shown in Fig. 3.

IV. EVALUATION

MATLAB 2018 program was used to implement the solution. MATLAB has considered two environments, GUIDE and App Designer, for designing graphic interfaces. The GUIDE environment has been used since the old versions of MATLAB, but the App Designer environment has recently been presented in the new versions of MATLAB and is being developed. Based on this, the user interface related to the proposed solution has been designed with the help of the advanced App Designer tool so that by using it, it is possible to receive cover text and secret text and carry out operations related to encryption or secret text extraction. Fig. 4 shows a view of the program environment implemented in MATLAB.

Based on this and as seen in Fig. 4, the user enters the cover text in the relevant box and then, in the next box, the confidential text that he intends to hide. Then, by clicking the hide button, the confidential text will be hidden inside the cover text according to the proposed solution. An example of the implementation of the solution is shown in Fig. 2 to 4.

The part related to the recovery of encrypted text is also shown in Fig. 5. For this purpose, it is sufficient for the user to click on the button to copy the encrypted text after performing the encryption operation. In this case, the page corresponding to Fig. 6 will open. Now the user clicks on the "decode" button so that the encrypted text will be shown along with the original cover text.

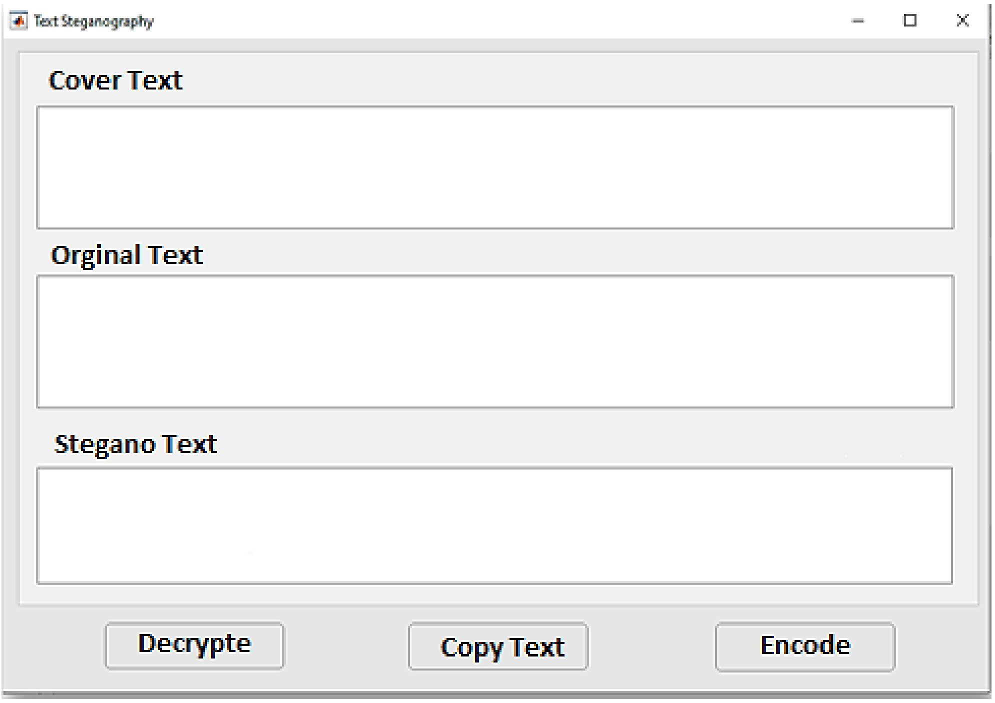
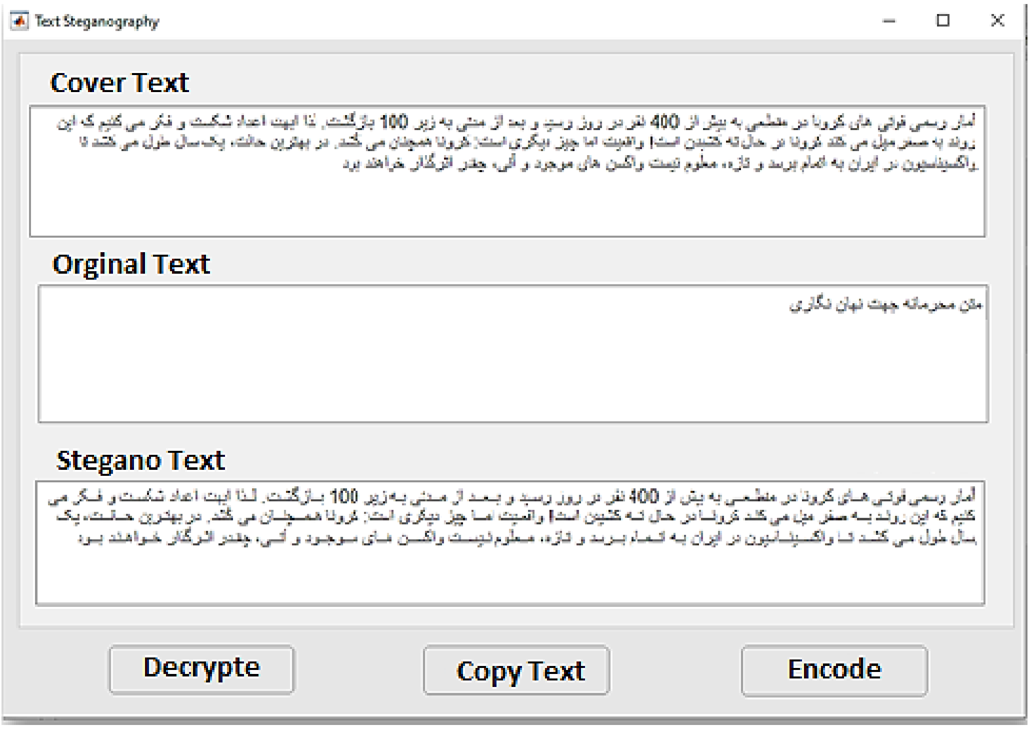


Fig. 4. A view of the user interface of the implemented program.

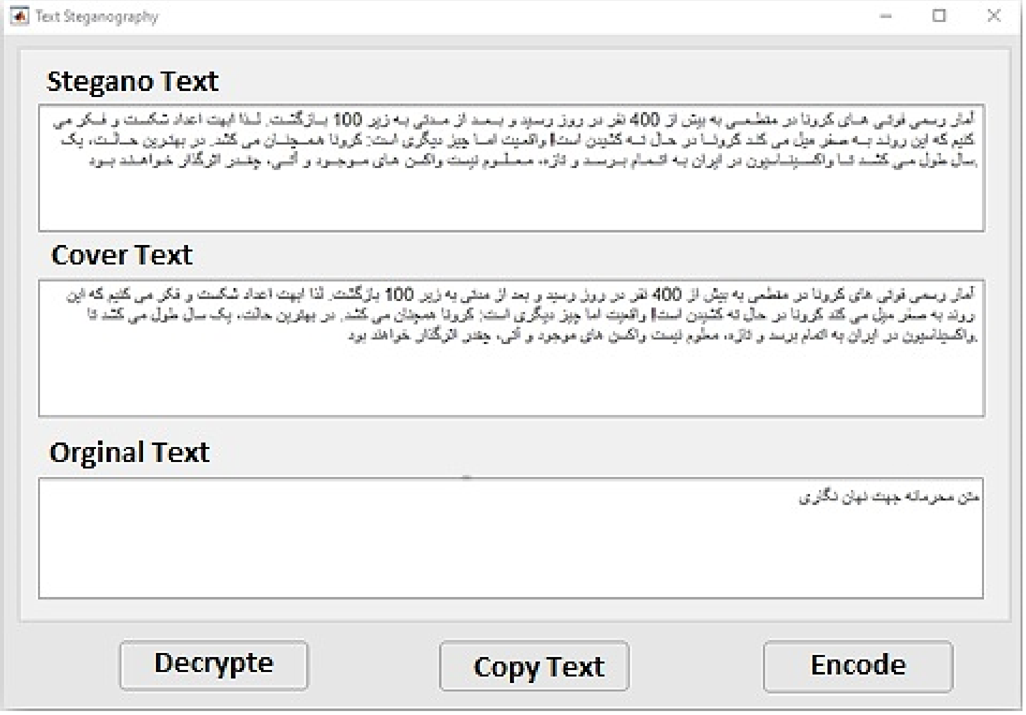


آمار رسمی فوتی های کرونا در مقطعی به بیش از 400 نفر در روز رسید و بعد از مدتی به زیر 100 بازگشت. لذا ابتدا اعداد شکست و فکر می کنیم که این
روند به صفر میل می کند کرونا در حال ته کشیدن است! واقعیت اما چیز دیگری است: کرونا همچنان می کُشد. در بهترین حالت، یک سال طول می کشد تا
واکسیناسیون در ایران به اتمام برسد و تازه، معلوم نیست واکسن های موجود و آتی، چقدر اثرگذار خواهند بود

متن محرمانه جهت نهان نگاری

آمار رسمی فوتی هـای کرونا در مقطعـی به بیش از 400 نفر در روز رسید و بـعـد از مـدتی بـه‌زیر 100 بـازگشت. لـذا ابتدا اعداد شکست و فـکر می
کنیم که این روند بـه صفر میل می کند کرونـا در حال تـه کشیدن است! واقعیت امـا چیز دیگری است: کرونا همـچـنـان می کُشد. در بهترین حـالت، یک
سال طول می کشد تـا واکسـینـاسیون در ایران بـه اتـمام بـرسد و تازه، مـعلوم‌نیست واکســن هـای مـوجـود و آتـی، چقدر اثرگذار خـواهند بـود

Fig. 5. How to hide the text by the solution.



Stegano Text

آمار رسمی فوتی هـای کرونا در مقطعـی به بیش از 400 نفر در روز رسید و بـعد از مـدتی به زیر 100 بـازگشت. لـذا ابهت اعداد شکـست و فـکر می کنیم که این روند بـه صفر میل می کند کرونـا در حال تـه کشیدن است! واقعیت امـا چیز دیگری است: کرونا همـچنـان می کشد. در بهترین حـالـت، یک سال طول می کشد تـا واکسـیناسیون در ایران به اتـمام بـرسـد و تازه، مـعـلـوم نیست واکسن های مـوجـود و آتـی، چقـدر اثرگذار خواهـند بـود.

Cover Text

آمار رسمی فوتی های کرونا در مقطعی به بیش از 400 نفر در روز رسید و بعد از مدتی به زیر 100 بازگشت. لذا ابهت اعداد شکست و فکر می کنیم که این روند به صفر میل می کند کرونا در حال ته کشیدن است! واقعیت اما چیز دیگری است: کرونا همچنان می کشد. در بهترین حالت، یک سال طول می کشد تا واکسیناسیون در ایران به اتمام برسد و تازه، معلوم نیست واکسن های موجود و آتی، چقدر اثرگذار خواهند بود.

Orginal Text

متن محرمانه جهت نهان نگاری

Fig. 6. The part of the recovery of the encrypted text.

A. Evaluation Parameters

In order to measure the proposed solution, a series of parameters are needed for review and evaluation. According to the studies that have been done on the articles written in the field of cryptography of Arabic and Persian texts, the proposed solution has been examined according to the parameter of cryptography capacity. The following relationship is used to calculate the hiding capacity:

$$CapacityRatio = \frac{\text{hidden text size}}{\text{covered text size}}$$

1) The solutions to be evaluated: In order to check the effectiveness of the solution, it is necessary to compare it with other common methods in the field of text encryption. Accordingly, in this research, the proposed solution is compared and evaluated with the following solutions, which were also examined in the research conducted in [27, 40]:

- Enhanced Kashida [3]
- ZWC and space [8]
- Enhanced capacity [9]
- Pseudo-space [10]
- Method 2 [11]
- Alhusban's method [12]
- Moon and Sun letters [24].

V. Results

Considering that only Arabic texts were used in the evaluation of the solutions mentioned in the previous section, as a result, in order to have the same data in the proposed solution, the same Arabic texts with the same size determined in the evaluated solutions were used. It has been done so that the checks and comparisons are done in a standard and uniform way. First, in order to compare the output of the solutions in Tables III and IV, an example of the output of the solutions has been implemented according to the number of different words. The purpose of this evaluation is to check the output of text encryption in each of the solutions.

Also, in Table IV, another example of the application of encryption solutions is shown on a cover text with a length of 155.

As can be seen in the output of Tables III and IV, the proposed solution is able to perform the encryption operation by applying the least noticeable change in the letters. Now, according to the ratio of encryption capacity, solutions with different lengths have been examined. The purpose of this evaluation is to check the percentage of optimality in hiding the message in envelopes of different sizes. The results of this evaluation are shown in Tables V to VIII. Also, the summarization of the overall performance of the proposed method is shown in Table IX.

As can be seen in Table VI, in the case where in the proposed solution only the encryption technique based on closed letters is used, the encryption capacity has decreased to a great extent. Based on this, the results of this evaluation indicate that the use of the extended technique (keshida) along with closed letters can effectively increase the capacity of cryptography.

As seen in the output of Tables III to VIII, in all three evaluations, the proposed solution has a higher encryption capacity compared to other methods, and with the increase in the length of the cover text, this amount has become more than optimal. The reason for this is the use of two techniques in the proposed solution. So that at first, using the technique of stretching the letters, the encryption operation was done, and as can be seen, this feature did not have the slightest effect on the readability or changing the meaning of the text. In addition, the utilization of the closed space available in Persian and Arabic letters has made the ability of the solution for encryption higher and the more the number of these letters, the higher the encryption capacity. Also, as it is clear from the results of these evaluations, in the case where the proposed solution, only the encryption technique based on the closed letter space is used, the encryption capacity has decreased to a great extent. Based on this, the results of this evaluation indicate that the use of the extended technique along with closed letters can increase the cryptography capacity in a more effective way.

TABLE III. COMPARISON OF THE OUTPUT OF THE IMPLEMENTATION OF 8 CRYPTOGRAPHY TECHNIQUES ON THE SAME ARABIC TEXT WITH THE COVER TEXT OF LENGTH 109

Output	Method
يـبـحث عـلم الحاسـوب استخدام الـحوسبـه بجـميع اشـكـالها لحـل المـشكلات من مـنـظورعـلمي رياضـي وغالـبا مـا يـشـمل ذلـك تصميم	Enhanced Kashida
يبحث علم الحاسوب استخدام الحوسبه بحميع اشكالها لحل المشكلات من منظور علمي رياضي وغالبا ما يشمل ذلك تصميم	ZWC and space
يـبحث عـلم الحاسـوب اسـتخدام الحوسـبـه بـحميـع اشـكـالها لحل الـمشكلـات من مـنـظور علمـي ريـاضـي وغـالـبـا ما يـشمل ذلـك تصميم	Enhanced capacity
يبحث علم الحاسوب استخدام الحوسبه بجميع اشكالها لحل المشكلات من منظور علمي رياضي وغالبا ما يشمل ذلك تصميم	Pseudo-space
يبحث علم الحاسوب استخدام الحوسبه بحميع اشكالها لحل المشكلات من منظور علمي رياضي وغالبا ما يشمل ذلك تصميم	Method 2
يـبـحث عـلم الحاسوب فـي استخـدام الحوسبة بجميـع أشكـالها لحل المشكلات من منظور عـلـمي ريـاضي. وغالـًبا ما يشمل ذلك تصميم	Alhusban's method
يبحث علـم الـحاسـوب في اسـتـخدام الـحـوسـبـة بـجميع أشـكالـهـا لـحـل المـشـكلـات من مـنـظـور عـلـمـي رياضـي. وغالـًبا ما يشـمل ذلـك تصـميـم	Moon and Sun letters
يبحث علم الـحاسوب استخـدام الحوسبه بـحـمـيـع اشكالـهـا لـحـل الـمـشـكـلات مـن مـنـظـور عـلـمـي ريـاضـي وغـالـبـا مـا يـشـمـل ذلك تـصـمـيـم	Proposed method

TABLE IV. COMPARISON OF THE OUTPUT OF THE IMPLEMENTATION OF 8 CRYPTOGRAPHY TECHNIQUES ON THE SAME ARABIC TEXT WITH THE COVER TEXT OF LENGTH 155

Output	Method
تـنـقسم علوم الصـحـة إلى قـسميـن دراسة جسم الإنـسـان والبحث لتـعـزيز مـعرفتـنـا بالآلـيـات الـتـي يعمـل بـهـا الـجسم الحي ومـمـرضـاتـه وعـلـم الـصـحة التـطبيـقـي الـذي يهتم بـتطـبـيق هذه الـمعرفه	Enhanced Kashida
تنقسم علوم الصحة إلى قسمين دراسة جسم الإنسان والبحث لتعزيز معرفتنا بالآليات التي يعمل بها الجسم الحي وممرضاته وعلم الصحة التطبيقي الذي يهتم بتطبيق هذه المعرفه	ZWC and space
تـنـقسم علوم الصـحـة إلى قـسميـن دراسة جسم الإنـسـان والبحث لتـعـزيز مـعرفتـنـا بالآلـيـات الـتـي يعمـل بـهـا الـجسم الحي ومـمـرضـاتـه وعـلـم الـصـحة التـطبيـقـي الـذي يهتم بـتطـبـيق هذه الـمعرفه	Enhanced capacity
تنقسم علوم الصحة إلى قسمين دراسة جسم الإنسان والبحث لتعزيز معرفتنا بالآليات التي يعمل بها الجسم الحي وممرضاته وعلم الصحة التطبيقي الذي يهتم بتطبيق هذه المعرفه	Pseudo-space
تنقسم علومالصحة إلىقسمين دراسة جسم الإنسانوالبحث لتعزيزمعرفتنا بالآلياتالتي يعملبها الجسمالحي وممرضاتهوعلم الصحةالتطبيقي الذييهتم بتطبيقهذهالمعرفة	Method 2
تنقسم علوم الصحة إلى قسمين: دراسة جسم الإنسان والبـحث لتعزيز معـرفـتـنـا بالآليات الـتي يـعمل بهـا الـجسم الحي وممرضـاتـه وعـلم الصحة الـ تـطبيقي الذي يهتم بتطبيق هذه المعرفـه	Alhusban's method
تـنقسـم عـلـوم الصحـــة إلى قـسمـــين دراســة جـــسم الإنسان والبـحـث لــتــعـزيز معرفتـنا بـالــآليات الــتــي يـعمل بها الــجسم الحـــي ومـمــمــرضاته وعـلـم الصحـــة التطبـيقي الــذي يــهتم بـتــطبـيق هـذه الــمعــرفــه	Moon and Sun letters
تنقسم عـلـوم الصحة إلـى قـسمين دراسة جسـم الــإنـسان والـبـحـث لـتعزيـز مـعرفتنا بـالــآلـيـات التي يعمـل بـهــا الجسـم الـحــي وممرضاته وعـلـم الــصــحــة الــتــطــبــيــقــي الذي يــهــتــم بــتـطـبـيــق هذه الــمــعــرفــه	Proposed Method

TABLE V. COMPARISON OF THE EVALUATION OF SOLUTIONS WITH A COVER TEXT OF 155 LENGTH

Capacity rate	The maximum length of hidden text	Method
46%	71	Enhanced Kashida [3]
31%	48	ZWC and space [8]
60%	93	Enhanced capacity [9]
15.4%	24	Pseudo-space [10]
61%	96	Method 2 [11]
13%	21	Alhusban's method [12]
58%	91	Moon and Sun letters [41]
34%	53	The proposed method, without using the extended technique
74%	115	proposed method

TABLE VI. COMPARISON OF THE EVALUATION OF SOLUTIONS WITH A COVER TEXT OF 600 LENGTH

Capacity rate	The maximum length of hidden text	Method
44.5%	264	Enhanced Kashida [3]
34%	48	ZWC and space [8]
61%	369	Enhanced capacity [9]
17%	103	Pseudo-space [10]
68%	408	Method 2 [11]
14%	84	Alhusban’s method [12]
67%	405	Moon and Sun letters [24]
27.5%	165	The proposed method, without using the extended technique
76.6%	460	proposed method

TABLE VII. COMPARISON OF THE EVALUATION OF SOLUTIONS WITH A COVER TEXT OF 1100 LENGTH

Capacity rate	The maximum length of hidden text	Method
46%	505	Enhanced Kashida [3]
32.7%	360	ZWC and space [8]
62.2%	685	Enhanced capacity [9]
16.3%	180	Pseudo-space [10]
65.4%	720	Method 2 [11]
12.3%	132	Alhusban’s method [12]
65%	716	Moon and Sun letters [24]
38%	420	The proposed method, without using the extended technique
79%	870	proposed method

TABLE VIII. COMPARISON OF THE EVALUATION OF SOLUTIONS WITH A COVER TEXT OF 1200 LENGTH

Capacity rate	The maximum length of hidden text	Method
46%	737	Enhanced Kashida [3]
32%	520	ZWC and space [8]
63%	1010	Enhanced capacity [9]
16%	260	Pseudo-space [10]
65%	1040	Method 2 [11]
13%	240	Alhusban’s method [12]
64%	1029	Moon and Sun letters [24]
40%	650	The proposed method, without using the extended technique
81.7%	1310	proposed method

TABLE IX. THE SUMMARIZATION OF THE OVERALL PERFORMANCE OF THE METHODS

Text Length 1200	Text Length 1100	Text Length 600	Text Length 155	Method
46%	46%	44.5%	46%	Enhanced Kashida [3]
32%	32.7%	34%	31%	ZWC and space [8]
63%	62.2%	61%	60%	Enhanced capacity [9]
16%	16.3%	17%	15.4%	Pseudo-space [10]
65%	65.4%	68%	61%	Method 2 [11]
13%	12.3%	14%	13%	Alhusban’s method [12]
64%	65%	67%	58%	Moon and Sun letters [24]
81.7%	**79%**	**76.6%**	**74%**	**Proposed Method**

A. *Coverage Capacity*

To assess the capability of the suggested approach, the formula for determining the number of secret message encodings based on letter is presented in [28]. The results are illustrated in Fig. 7, representing the coverage capacity (CC) values computed through the application of Eq. (1).

$$CC(M) = \frac{1}{M}\sum_{i=1}^{M}\frac{(L_i).k}{D(S_i)} \times 100$$

In the given context, "M" denotes the count of covered secret messages, representing the number of matched segments at the maximum length of "N" bits. "Li" refers to the length of each matched segment, measured in the number of letters within it. The average number of embedded bits in each letter of the segment is denoted by "K," and the number of bits in a segment is represented by "D(Si)." For the Arabic letter, the standard encoding system utilizes 8 bits; therefore, D(Si) = 8 * Li.

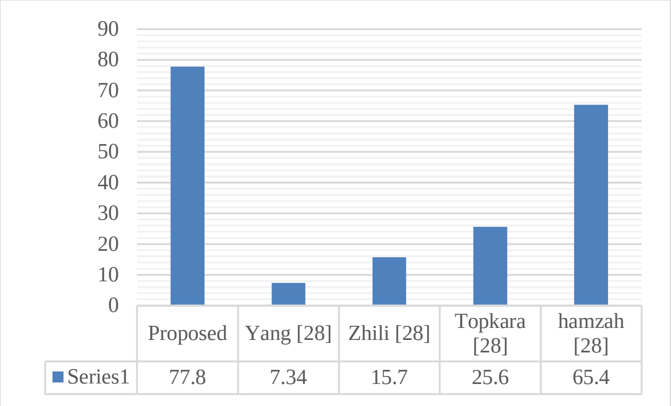


Fig. 7. Coverage capacity compared to other approaches.

Fig. 7 shows the coverage capacity between proposed method and other methods [28]. The coverage rate of the proposed method, which utilizes multiple shapes of letters, significantly surpasses the capacity rate of other methods. These CC values are derived from the proposed and other methods are based on a dataset sample. Furthermore, a comparison is made between these CC values and those obtained from previous works and similar studies to gauge the efficacy and efficiency of the proposed methodology. Upon applying this equation, the coverage capacity of our proposed method, which employs multiple shapes of letters, reached average at level of 77.8%. As can be seen, in this case, the proposed method has performed better than all the other four solutions. It has performed more than 70% better than Yang method [28], and more than 60% better than Zhili method [28], and more than 25% better than Hamzah [28] method. The findings demonstrate that the proposed solution, with employing both stretching and altering closed letters, has achieved a favorable capacity for concealing text.

VI. Conclusion

In Persian and Arabic languages, each letter can have a unique code according to its position. On the other hand, in some cases, in order to make a letter more beautiful, it can also be written with a stroke. In fact, stretching letters is one of the advantages of Persian and Arabic languages; this feature does not have the slightest effect on the readability or changing the meaning of the text. The only problem is that not all letters have this ability, and it is limited to the position in which that letter is placed. Accordingly, this feature has been used in Persian and Arabic languages in this research, and a hybrid encryption solution has been presented. Based on this, minor changes were made in the letters with closed spaces in addition to stretching the letters. The letters with closed space include the characters "ص،ض،ط،ظ،ع،غ،ف،ق،م،و،ه ”. Of course, for the letters "ع،غ " only when these letters are used in the middle of the word will they have a closed space; therefore, in this method, only one position of the letters "ع،غ " (in the middle of the word) can be used to hide the information in the letters "ع،غ ". Accordingly, in order to hide these letters, the amount of empty space is slightly smaller so that it has the least effect on changing the font of the desired letters and through this change, if the last letter of a word that cannot be if is drawn, it is part of the letters with closed space, it is used. As a result, if the insertable bit is zero, the character will remain unchanged. But if the insertable bit was 1, the character is inserted with a smaller enclosed space. In the end, the proposed solution was implemented in the MATLAB program environment and was evaluated by the rate parameter of the encryption capacity compared to other solutions. The evaluations were carried out according to the sentences with different lengths, and the relevant results indicated that in this case, the proposed solution through the use of two techniques of stretching and changing closed letters has been able to have a suitable capacity of hiding to get writing. So in all the evaluations, it has a noticeable superiority compared to other methods.

VII. Future Works

In the future work, we can focus on the development a mobile application. Also, more improvements can be made by using the available space in Arabic and Persian letters. As a result, in this way, the coverage rate can be increased to a greater extent in the proposed solution.

VIII. Funding

This work was supported by the Education Department of Jilin Province 2022 Annual vocational education and adult education teaching reform research project. Project name: Design and Research of Thin Client Online Training System Based on 5G and Cloud Computing Virtualization Technology. Project number: 2022ZCZ011.

References

[1] S. R. Yaghobi and H. Sajedi, "Text steganography in webometrics," International Journal of Information Technology 13,pp. 621–635 (2021).

[2] A. Majumder and S. Changder, "An Automated Cover Text Selection System for Text Steganography Algorithms," in Intelligent Cyber-Physical Systems Security for Industry 4.0 (Chapman and Hall/CRC, 2022), pp. 33–55.

[3] N. A. Roslan, N. I. Udzir, R. Mahmod, and A. Gutub, "Systematic literature review and analysis for Arabic text steganography method practically," Egyptian Informatics Journal (2022).

[4] Kunhoth, J., Subramanian, N., Al-Maadeed, S., & Bouridane, A. (2023). Video steganography: recent advances and challenges. Multimedia Tools and Applications, 1-43.

[5] Marszałek, P., & Bilski, P. (2023). Steganography in Audio Files: COTS Software Analysis. International Journal of Electronics and Telecommunications, 69(1).

[6] Peng, W., Wang, T., Qian, Z., Li, S., & Zhang, X. (2023). Cross-Modal Text Steganography Against Synonym Substitution-Based Text Attack. IEEE Signal Processing Letters, 30, 299-303.

[7] Xiang, L., Wang, R., Yang, Z., & Liu, Y. (2022). Generative Linguistic Steganography: A Comprehensive Review. KSII Transactions on Internet & Information Systems, 16(3).

[8] S. M. A. Al-Nofaie and A. A.-A. Gutub, "Utilizing pseudo-spaces to improve Arabic text steganography for multimedia data communications," Multimed Tools Appl 79, pp.19–67 (2020).

[9] W. Peng, T. Wang, Z. Qian, S. Li, and X. Zhang, "Cross-Modal Text Steganography Against Synonym Substitution-Based Text Attack," IEEE Signal Process Lett 30, pp.299–303 (2023).

[10] Chen Cao, Jianhua Wang, Devin Kwok, Zilong Zhang, Feifei Cui, Da Zhao, Mulin Jun Li, Quan Zou. webTWAS: a resource for disease candidate susceptibility genes identified by transcriptome-wide association study. Nucleic Acids Research.2022, 50(D1): D1123-D1130.

[11] Ning Xu, Zhongyu Chen, Ben Niu, and Xudong Zhao. Event-Triggered Distributed Consensus Tracking for Nonlinear Multi-Agent Systems: A Minimal Approximation Approach, IEEE Journal on Emerging and Selected Topics in Circuits and Systems, DOI: 10.1109/JETCAS.2023.3277544, 2023.

[12] Trick, M., & Boukani, B. (2014). Placement algorithms and logic on logic (LOL) 3D integration. *Journal of mathematics and computer science*, *8*(2), 128-136.

[13] Haoyu Zhang, Quan Zou, Ying Ju, Chenggang Song, Dong Chen. Distance-based Support Vector Machine to Predict DNA N6-methyladine Modification. Current Bioinformatics. 2022, 17(5): 473-482.

[14] B. Das, S. Mondal, and K. K. Mandal, "Combined Cryptography and Text Steganography for Enhanced Security Based on Number System," in Machine Learning, Image Processing, Network Security and Data Sciences: Select Proceedings of 3rd International Conference on MIND 2021 (Springer, 2023), pp. 839–849.

[15] M. A. Majeed, R. Sulaiman, Z. Shukur, and M. K. Hasan, "A review on text steganography techniques," Mathematics 9(21),p. 2829 (2021).

[16] O. F. A. Adeeb and S. J. Kabudian, "Arabic text steganography based on deep learning methods," IEEE Access 10, pp.94403–94416 (2022).

[17] Khezri, E., Zeinali, E., & Sargolzaey, H. (2022). A novel highway routing protocol in vehicular ad hoc networks using VMaSC-LTE and DBA-MAC protocols. *Wireless Communications and Mobile Computing*, *2022*.

[18] Roslan, N. A., Udzir, N. I., Mahmod, R., & Gutub, A. (2022). Systematic literature review and analysis for Arabic text steganography method practically. Egyptian Informatics Journal.

[19] Majeed, M. A., Sulaiman, R., Shukur, Z., & Hasan, M. K. (2021). A review on text steganography techniques. Mathematics, 9(21), 2829.

[20] Xiang, L., Guo, G., Yu, J., Sheng, V. S., & Yang, P. (2020). A convolutional neural network-based linguistic steganalysis for synonym substitution steganography. Mathematical Biosciences and Engineering, 17(2), 1041-1058.

[21] Li, M., Mu, K., Zhong, P., Wen, J., & Xue, Y. (2019). Generating steganographic image description by dynamic synonym substitution. Signal Processing, 164, 193-201.

[22] Wang, J., Zhu, Y., Ni, J., Wang, H., & Yao, Y. (2023). Text Coverless Information Hiding Based on the Combination of Chinese Character Components. Journal of Circuits, Systems and Computers, 32(03), 2350055.

[23] Haoyan Zhang, Xudong Zhao, Huangqing Wang, Ben Niu, Ning Xu, Adaptive Tracking Control for Output-Constrained Switched MIMO Pure-Feedback Nonlinear Systems with Input Saturation, Journal of systems science & complexity, 36: 960–984, 2023.

[24] Heng Zhao, Huanqing Wang, Ben Niu, Xudong Zhao, K. H. Alharbi, Event-Triggered Fault-Tolerant Control for Input-Constrained Nonlinear Systems With Mismatched Disturbances via Adaptive Dynamic Programming, Neural Networks, 164: 508-520, 2023.

[25] Trik, M., Molk, A. M. N. G., Ghasemi, F., & Pouryeganeh, P. (2022). A Hybrid Selection Strategy Based on Traffic Analysis for Improving Performance in Networks on Chip. *Journal of Sensors*, *2022*.

[26] Zhongwen Cao; Ben Niu; Guangdeng Zong; Xudong Zhao; Adil M. Ahmad, "Active Disturbance Rejection-Based Event-Triggered Bipartite Consensus Control for Nonaffine Nonlinear Multiagent Systems", International Journal of Robust and Nonlinear Control, DOI:10.1002/rnc.6746.

[27] Yu, L., Lu, Y., Yan, X., & Yu, Y. (2022). Mts-stega: linguistic steganography based on multi-time-step. Entropy, 24(5), 585.

[28] Hamzah, A. A., & Bayomi, H. (2020). Text steganography with high embedding capacity using arabic calligraphy. In Emerging Trends in Intelligent Computing and Informatics: Data Science, Intelligent Information Systems and Smart Computing 4 (pp. 127-138). Springer International Publishing.

[29] Cao, Y., Zhou, Z., Chakraborty, C., Wang, M., Wu, Q. J., Sun, X., & Yu, K. (2022). Generative steganography based on long readable text generation. IEEE Transactions on Computational Social Systems.

[30] Khezri, E., Zeinali, E., & Sargolzaey, H. (2023). SGHRP: Secure Greedy Highway Routing Protocol with authentication and increased privacy in vehicular ad hoc networks. *Plos one*, *18*(4), e0282031.

[31] Alanazi, N., Khan, E., & Gutub, A. (2021). Efficient security and capacity techniques for Arabic text steganography via engaging Unicode standard encoding. Multimedia Tools and Applications, 80, 1403-1431.

[32] Alyousuf, F. Q. A., & Din, R. (2020). Analysis review on feature-based and word-rule based techniques in text steganography. Bulletin of Electrical Engineering and Informatics, 9(2), 764-770.

[33] Mokhlesi Ghanevati, D., Khorami, E., Boukani, B., & Trik, M. (2020). Improve replica placement in content distribution networks with hybrid technique. *Journal of Advances in Computer Research*, *11*(1), 87-99.

[34] Bukhelli, A. A. (2023). Manipulating the Perception of Paragraph Breaks: A New Theoretical Model of Textual Steganography Using Paragraphs (Doctoral dissertation, University of Portsmouth).

[35] Khezri, E., & Zeinali, E. (2021). A review on highway routing protocols in vehicular ad hoc networks. *SN Computer Science*, *2*, 1-22.

[36] Khot, S., Thakur, S., Patil, S., & Bhandari, K. Emoji Steganography Using AES & LSB Technique. JOURNAL OF ENGINEERING AND SCIENCES, 18.

[37] N. A. Roslan, N. I. Udzir, R. Mahmod, and A. Gutub, "Systematic literature review and analysis for Arabic text steganography method practically," Egyptian Informatics Journal (2022).

[38] Samiei, M., Hassani, A., Sarspy, S., Komari, I. E., Trik, M., & Hassanpour, F. (2023). Classification of skin cancer stages using a AHP fuzzy technique within the context of big data healthcare. *Journal of Cancer Research and Clinical Oncology*, 1-15.

[39] Sun, J., Zhang, Y., & Trik, M. (2022). PBPHS: a profile-based predictive handover strategy for 5G networks. Cybernetics and Systems,53(6), 1-22.

[40] Trik, M., Akhavan, H., Bidgoli, A. M., Molk, A. M. N. G., Vashani, H., & Mozaffari, S. P. (2023). A new adaptive selection strategy for reducing latency in networks on chip. *Integration*, *89*, 9-24.

[41] S. N. Al Azzam and F. A. Al-Garni, "The use of binary digit mapping on ASCII characters to create a high-capacity, undetectable text steganography," Journal of Advanced Sciences and Engineering Technologies 5(2), pp.51–59 (2023).