International Journal of Electrical and Computer Engineering (IJECE)
Vol. 10, No. 4, August 2020, pp. 3550~3557
ISSN: 2088-8708, DOI: 10.11591/ijece.v10i4.pp3550-3557

Text hiding in text using invisible character

Nada Abdul Aziz Mustafa
College of Languages, University of Baghdad, Iraq

Article Info

Article history:

Received Jun 24, 2019
Revised Jan 29, 2020
Accepted Feb 10, 2020

Keywords:

Invisible character
Steganography
Stego-cover
Stego-object
Stego-text

ABSTRACT

Steganography can be defined as the art and science of hiding information in the data that could be read by computer. This science cannot recognize stego-cover and the original one whether by eye or by computer when seeing the statistical samples. This paper presents a new method to hide text in text characters. The systematic method uses the structure of invisible character to hide and extract secret texts. The creation of secret message comprises four main stages such using the letter from the original message, selecting the suitable cover text, dividing the cover text into blocks, hiding the secret text using the invisible character and comparing the cover-text and stego-object. This study uses an invisible character (white space) position of in the cover text that used to hide the the secrete sender masseges. The experiments results show that the suggested method presents highly secret due to use the multi-level of complexity to avoid the attackers.

Corresponding Author:

Nada Abdul Aziz Mustafa,
College of Languages,
University of Baghdad,
Karrada, Al-Jadriya, Baghdad, Iraq.
Email: nada@colang.uobaghdad.edu.iq

1. INTRODUCTION

Text steganography technique is applied for saving privacy and originality of text-based documents. Thereby, text steganography connsiders as a challenging mission that tenuous adjustment in text file can be specified. Invisible character technique is used to hide the text into text or image without anyone can be seen the hidden processing [1]. There are different techniques like steganography, cryptography, coding, etc have been utilized. Currently, steganography is commonly used to hide text in text. Steganography implies "covered writing" in Greek as well as it considers the flag of communicating in a hidden style [2].

In view of this digital steganography conceals even the evidence of encrypted messaging [3]. Steganography is one of the oldest arts that people were eager to have since they started communication with each other [4]. Many methods have been used to hide information by using the recorder with tales of steganography and cryptography through times of war or peace [5]. Moreover, steganography is the art and science which hide information in any computer readable data in a way that a stego-object should be not distinguishable from origin cover neither by a human nor by computer looking for statistical pattern [6]. While steganography is an early topic, the current preparation of it arrives via the prisoner's issues suggested by [7], Alice and Bob are the two prisoners, they communicate via secret to access an escape strategy. Several message sent via a warden called Eve that may toss the private limitation when it suspects any type of secret message [8]. Thus, the secret letter provides the natural of steganography to realize the hiding information between them. They used a warden method to test the whole messages replaced between them may either be direct or indirect. A direct warden method attempts for adjusting the message and the supposed covered data calculatingly via eliminating the data while an indirect warden method that is used to take the meaning of covered message, updates the others and lets the message to permit [9]. A supposition may be completed based on this typical when both the sender and receiver will share particular public secret data then the conforming steganography procedure is recognized as then the secret key steganography where

as pure steganography proceeds that there is nothing previous data mutual by sender and receiver. While the public key of the receiver is identified to the sender, the steganographic procedure is named public key steganography [10].

2. RELATED WORK

Steganography suppose the opposition may capture the cover, nevertheless may be not observe any data further the original cover satisfied [11]. The data is covered and can have no supplementary safety further the real letter embedding [12]. A role principle for performance measurement of a steganographic structure is the statistical hiddenness of the secret data [13]. Digital text image have higher grade of redundancy and hence the most appropriate for steganography. There exist many steganographic methods that allow hiding secret data in text image having their own pros and cons. Least Significant Bit modification (LSB) method is the commonly used steganographic procedure that uses the information at least significant bits of a text image are random noise and shifting them does not shift the perceptual worth of the text images. Moreover, LSB depend on approaches may be either LSB Matching or LSB Replacement where the former swaps the LSB of the pixels with the communication to be sent and the advanced increment/decrement the pixels arbitrarily with secret bits [14]. Hiding information within places appears to be potential as human hardly may distinguish about the being of the hidden bits. Presently, operation of whitespaces appears useful and has its possible in information hiding due to whitespaces seem in a document more than the occurrence of words. It is uniform a benefit when anyone will distinguish that a blank piece of text is really role secret information [15]. In addition, a new technique of text steganography was presented. Their proposed algorithm uses text media to embed their secret text file depending on a dictionary. This dictionary contains English. Words sorted in alphabetical order to be selected by user in order to build the cover text [16]. The statistical properties of dithered imagery was used. With this method, the information bits to be concealed control the dot patterns of the ordered dither pixel. The system adapts two kilobytes of hidden data for a bit level (256*256) images, compliant a volume of or data-hiding ratio of one data bit to four cover image bits [17]. In the same context, the term steganography factually incomes “covered writing” as stemmed from Greek. It can be defined as the art and science of communication in a way that it used to hide the actuality of communication. A general steganography system is shown in Figure 1 as given by [18] in the following:

In this section, it is explained the results of research and at the same time is given the comprehensive discussion. Results can be presented in figures, graphs, tables and others that make the reader understand easily [2, 5]. The discussion can be made in several sub-chapters.

- Embedded <data-type>: Something to be hidden in something else.
- Stego-<data-type>: The output of the hiding process, something that has the embedded.
- Cover <data-type>: An input with “original” form of the stego-message. In some applications such a cover message is given from the outside, it can be chosen during the hiding process. The letter is represented by the dashed extension to the inner hiding process.
- Stego-object: the output from stego-system, something hidden in it.
- Stego-key: Extra secret information can be hidden in the stego-system the same key is typically required to extract the embedded message again.
- The process of hiding the embedded message is called Embedding.
- Getting the embedded message out of the stego-message again is called Extracting [19].

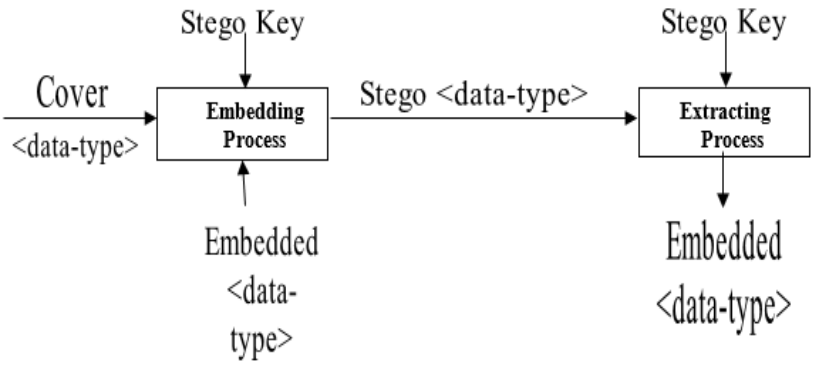


Figure 1. General steganography system

3. PROPOSED SYSTEM

The systematic approach is used invisible character that can be classified into three chief stages: Hiding (secret) text in (cover) text using the invisible character (hiding keys), extracting the (original) text by using the hidden keys and performance evaluation as shown in Figure 2.

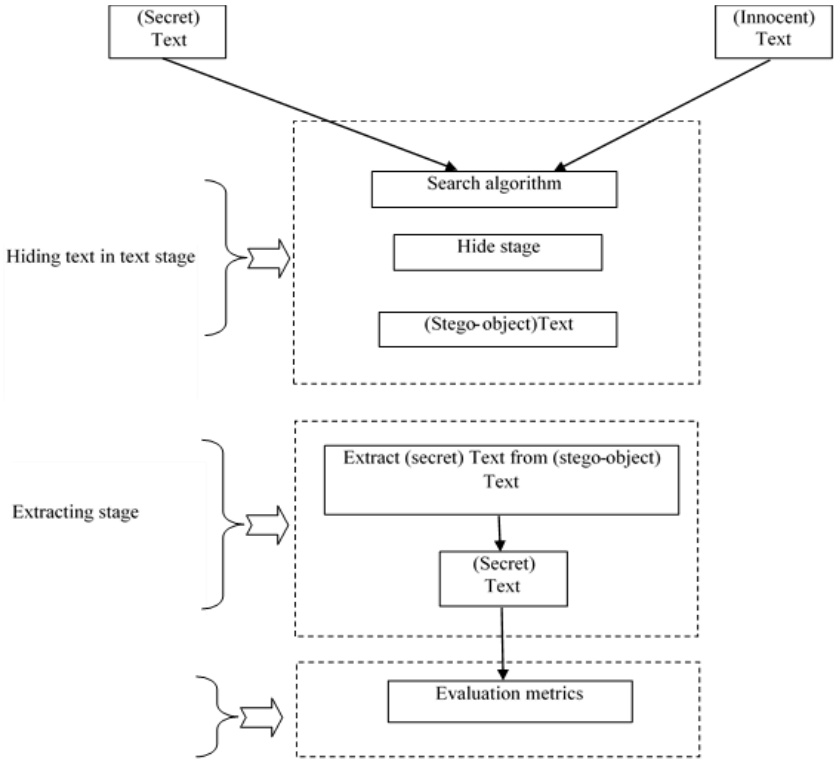


Figure 2. Flow chart of the systematic approach

3.1. Hiding text in text

Below the steps for hiding the (secret) text in (innocent) text. Choose any text with enough size as innocent text to use it as cover text. The hide procedures are as follows:

- Divide the (cover-Text) to blocks (number of characters or bytes).
- In each block hides one character of the secret message (Text).
  (Note: both of them cover and secret message characters are represented in ASCII code).
- The calculations of each block length is:

$$\text{Block length} = \frac{No.of\ characters\ in\ cover-text}{No.of\ characters\ in\ secret-text}\ (\text{Byte})$$

- From each block we leave two bytes (two from the beginning and at the end of the block) that will not be used to hide the secret message. (Note: more than two bytes can be leaved to increase the hiding capacity). Search the first block to find an identical (ASCII) value to that of the first secret message character.
- If we found it then insert invisible character to the right of it to indicate the hidden character.
- Else, if we not found an identical (ASCII) value then we search again for the character that is either greater than secret character (ASCII) value by (1,3,or 5) or less than secret character (ASCII) value by (3 or 5); if we found one then insert an invisible character to the right of it.
- To indicate whether the character to the left of invisible character was identical increase, or decrease by some value. The following protocol (depending on the two bytes that we left at the beginning and end of the block).
- The comparison between the cover-text and steg-object doesn't find any modification (before and after hiding text) and it should be identical to prevent any drawn suspicion of the attackers existed into innocent text as shown in Figure 3.

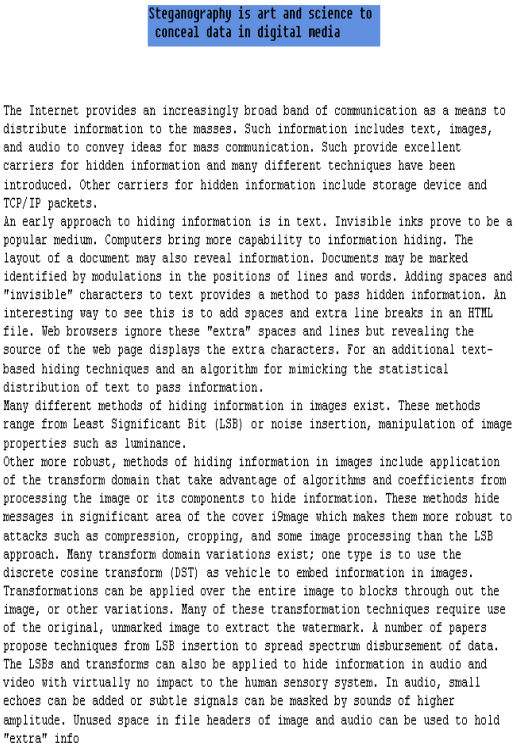


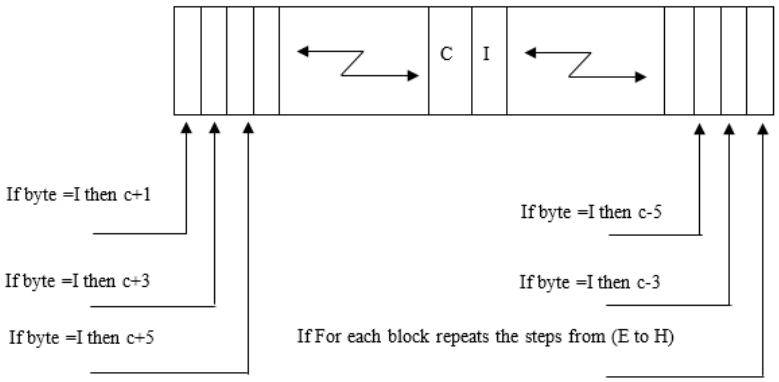
Figure 3. Secret text in the top and the cover text at the bottom (before/after hiding text)

3.2. Hiding processing

The whole steps for each point is explained as follows:

- Insert invisible character at the end of the block to indicate an identical character value.
- Insert invisible character at the beginning of the block, means that the character value is increased by one.
- Insert invisible character after the first byte of the block means, that the character value is increased by three.
- Insert invisible character after the second byte of the block, means that the character value is increased by five.
- Insert invisible character before the first byte from the end of the block, means that the character value is decreased by three.
- Insert invisible character before the byte before the second byte from the end of the block means that the character value is decreased by five.

For each block repeats the steps from (E to H) until all secret characters are hidden Figure 4, The flow chart for hiding text in text as shown Figure 5.



When I: is the invisible character and C: is any value (ASCII).

Figure 4. Hiding text in text using invisible character

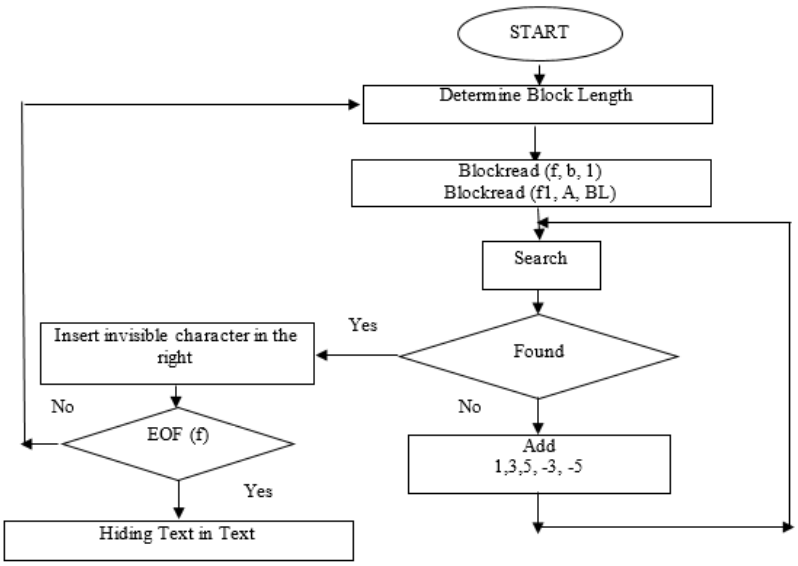


Figure 5. Flow chart to hide text in text f: secret message file, f1: cover file, b: byte, A: Array of character BL: Block Length

While the algorithm for hiding (secret) text in (innocent) text as follows:

```
Start
    Loop: while (not Eof stego-object file) do
     Begin
  Blockread (stego-object file, b, Block length)
     Blockread (secret file, c, 1)
    Loop: Repeat i=3;
              If C=(b [i] or (b [i]+1) or (b [i]+3) or (b [i]+5 or (b [i]-3 or (b[i]-5)) then
 Begin
    b [i+1] = Invisible character;
    (b[1]or b[2] or b[3] or b[block length] or b[block length-1]or b[block length-2] ) = invisible
    character {according to above}
    Found = true;
  End;
    Until (found)
```

3.3. Extractin stage

The extracting stage is intended to recover the hidden text from the stego-object. Extracting stage includes extracting secret-text from the stego-object text that it is indicated as the opposites of hide procedure algorithms [20]. The whole algorithm is used to extract the secret text as shown in details as bellows:

```
Start
    Blockread (stego-object file, b, 6o)
    j= b [1]
    No. Of part = b [1+j]
    No. Of Byte in each part = b [1+2j] and b [2+2j]
    No. Of byte in last part = b [1+3J]
    Length of text block = b [1+4j]
Loop: repeat
Loop: For i= 1 to (No. of byte hide in each part) NS
            Block read (stego-object, c, 1)
            If (i =A [i]) then
            Block write (extract file, c, 1)
  Until (No. Of parts)
```

4. EXPERIMENTAL RESULT

Prefect steganography is when gets stego-object similar to original innocent text by both perceptually and computer reading. This may be impossible to reach. For experimental results, it was found that the cover when it changes to stego-object gives the closest criteria to original cover.

There are two different schemes that is used to give an idea if these are a hidden message or not:

- Blind scheme. One is the process that should be possible without the original cover called (blind schemes) which need a high technique to discover and detect the hidden message [21].
- Escrow scheme. The second scheme where the original cover signal is needed to reveal the hidden information called (escrow scheme).

The second scheme compare between the origin cover and stego-object and cheek if there are any different. The second scheme is used to evaluate the performance evaluation by taking several criteria that it was used to evaluate the covers [22].

4.1. Avoiding stego-analysis

The developed algorithm for information hiding in this study was implemented taking into account stego-analysis. This algorithm includes some features that may help in avoiding stego-analysis attacks. These features are:

- Using private stego-object to prevent drawing suspicion about the hidden information. A comparison between stego-object and the innocent cover may be inefficient, and any modification on the original cover may be imperceptible [23]. Figure 6 explain comparing between the text-cover and its corresponding stego-text. From the result of comparing between the innocent text and the stego-object, it observed that anyone cannot see any modification in stego-object.
- Using multiple stages of encryption to prevent the extraction of the hidden information as most as possible [24], i.e. if there is any suspicion about the hidden information, they cannot be extracted.

Hiding the secret text in innocent text at first by using the method of the hypertext markup language (HTML) files. The files utilized to convey the information meanwhile adding space, tabs, and “invisible” character. For example, hiding an English secret text Figure 7 in an innocent text Figure 8, stego-object with hidden information, and without drawn suspicion by using image (1), while Figure 9 explains the evaluation results with hidden information when using stego-object.

From the experimental results, it can clearly observed that the comparison result of performance evaluation (average, variant, energy, entropy, similarity, secrecy and SNR) between the cover-text and the stego-object is a little bit different that it indicate anyone can not seen the hidden operations. The average value of the proposed technique carried out better performance with an increase of 15%% compared to the average value of of energy and sensitivity [25] using the BMP image steganography.

The Internet provides an increasingly broad band of communication as a means to distribute information to the masses. Such information includes text, images, and audio to convey ideas for mass communication. Such provide excellent carriers for hidden information and many different techniques have been introduced. Other carriers for hidden information include storage device and TCP/IP packets.
An early approach to hiding information is in text. Invisible inks prove to be a popular medium. Computers bring more capability to information hiding. The layout of a document may also reveal information. Documents may be marked identified by modulations in the positions of lines and words. Adding spaces and "invisible" characters to text provides a method to pass hidden information. An interesting way to see this is to add spaces and extra line breaks in an HTML file. Web browsers ignore these "extra" spaces and lines but revealing the source of the web page displays the extra characters. For an additional text-based hiding techniques and an algorithm for mimicking the statistical distribution of text to pass information.
Many different methods of hiding information in images exist. These methods range from Least Significant Bit (LSB) or noise insertion, manipulation of image properties such as luminance.
Other more robust, methods of hiding information in images include application of the transform domain that take advantage of algorithms and coefficients from processing the image or its components to hide information. These methods hide messages in significant area of the cover i9mage which makes them more robust to attacks such as compression, cropping, and some image processing than the LSB approach. Many transform domain variations exist; one type is to use the discrete cosine transform (DST) as vehicle to embed information in images. Transformations can be applied over the entire image to blocks through out the image, or other variations. Many of these transformation techniques require use of the original, unmarked image to extract the watermark. A number of papers propose techniques from LSB insertion to spread spectrum disbursement of data. The LSBs and transforms can also be applied to hide information in audio and video with virtually no impact to the human sensory system. In audio, small echoes can be added or subtle signals can be masked by sounds of higher amplitude. Unused space in file headers of image and audio can be used to hold "extra" info

ORIGIN TEXT

The Internet provides an increasingly broad band of communication as a means to distribute information to the masses. Such information includes text, images, and audio to convey ideas for mass communication. Such provide excellent carriers for hidden information and many different techniques have been intr-oduced. Other carriers for hidden information include storage device and TCP/IP packets.
An early approach to hiding information is in text. Invisible inks prove to be a popular medium. Computers bring more capability to information hiding. The layout of a document may also reveal information. Documents may be marked ident-ified by modulations in the positions of lines and words. Adding spaces and "invisible" characters to text provides a method to pass hidden information. An interesting way to see this is to add spaces and extra line breaks in an HTML file. Web browsers ignore these "extra" spaces and lines but revealing the source of the web page displays the extra characters. For an additional text-oased hiding techniques and an algorithm for mimicking the statistical distribution of text to pass information.
Many different methods of hiding information in images exist. These methods range from Least Significant Bit (LSB) or noise insertion, manipulation of image properties such as luminance.
ther more robust, methods of hiding information in images include application of the transform domain that take advantage of algorithms and coefficients from processing the image or its components to hide information. These methods hide messages in significant area of the cover i9mage which makes them more robust to attacks such as compression, cropping, and some image processing than the LSB approach. Many transform domain variations exist; one type is to use the discrete cosine transform (DST) as vehicle to embed information in images. Transformations can be applied over the entire image to blocks through out the image, or other variations. Many of these transformation techniques require use of the original, unmarked image to extract the watermark. A number of papers propose techniques from LSB insertion to spread spectrum disbursement of data. The LSBs and transforms can also be applied to hide information in audio and video with virtually no impact to the human sensory system. In audio, small echoes can be added or subtle signals can be masked by sounds of higher amplitude. Unused space in file headers of image and audio can be used to hold "extra" info

STEGO TEXT

Figure 6. Comparison between the origin (text-cover) and its corresponding stego-object text

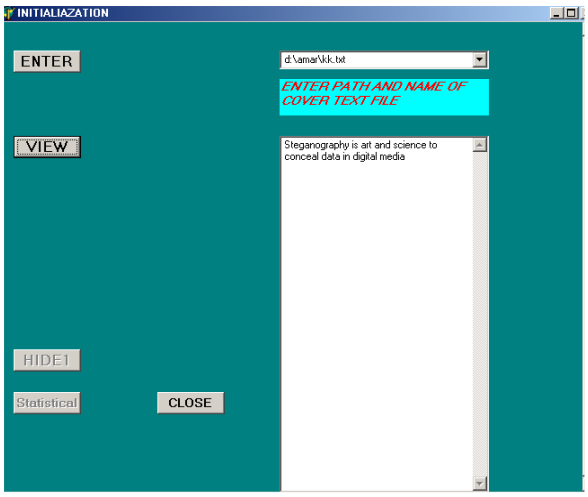


Figure 7. English text (secret message), which will be hidden in text

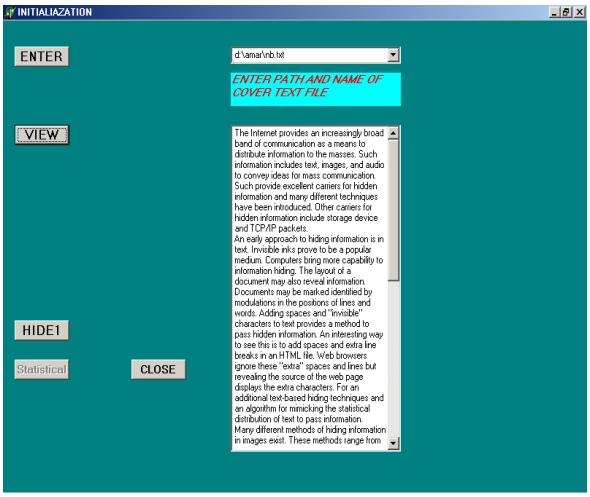


Figure 8. The innocent text cover, in which information will be hidden

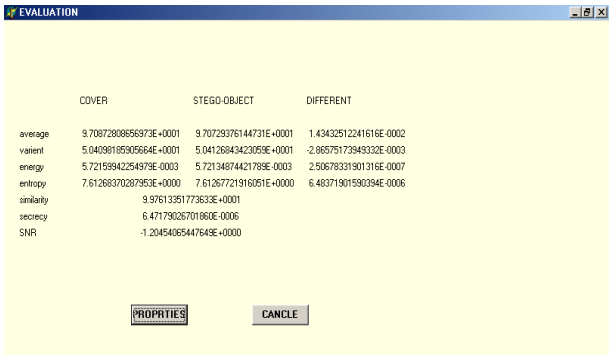


Figure 9. Comparison between text-cover and stego-object text depends on the evaluation metrics

5. CONCLUSION

Hiding text in text using invisible character provides an efficient method. This study was performed in three stages, hide text in text, extract the secret text from the stego-object text and evaluation metrics. The hide procedure can perform a security operation in order to hide the secret text. The extract stage algorithm reverses the hide procedure effects in order to extract the hidden text. Moreover, using public stego-object to prevent knowing the original cover, in order to prevent any comparison between the innocent cover and stego-object. There are many levels of complexity with the proposed system, which prevent attacker to reach the secret message. Consequently, the capability to increase the capacity of hiding. Proposed system depends on modifying (2-3) bit of secret message. Also can instead of represent each character with (8-bit) can represent it with (5-bit) (no. of character are 26) and hence (3-bit) save can accomplished.

REFERENCES

[1] V.L. Narayana, A.P. Gopi, and N.A. Kumar, "Different techniques for hiding the text information using text steganography techniques: A survey," *Ingenierie des Systemes d'Information*, vol. 23, pp. 115, 2018.
[2] A.F.M.A.K. Sana and S. Mohmmad, "An FPGA implementation of secured steganography communication system," *Tikrit Journal of Engineering Sciences*, vol. 19, pp. 14-23, 2012.
[3] A. Kumar and K. Pooja, "Steganography-A data hiding technique," *International Journal of Computer Applications*, vol. 9, pp. 19-23, 2010.
[4] H. Mohajeri Moghaddam, B. Li, M. Derakhshani, and I. Goldberg, "Skypemorph: Protocol obfuscation for tor bridges," in *Proceedings of the 2012 ACM conference on Computer and communications security*, pp. 97-108, 2012.
[5] I. Sedeeq, "HTML Steganography Algorithms and Detection Methods," University of Liverpool, 2018.
[6] S.M.A. Al-Nofaie and A.A.-A. Gutub, "Utilizing pseudo-spaces to improve Arabic text steganography for multimedia data communications," *Multimedia Tools and Applications*, pp. 1-49, 2019.

[7] G.J. Simmons, "The prisoners' problem and the subliminal channel," in *Advances in Cryptology*, pp. 51-67, 1984.
[8] T. Pevný, T. Filler, and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *International Workshop on Information Hiding*, pp. 161-177, 2010.
[9] M. Kharrazi, H.T. Sencar, and N. Memon, "Image steganography and steganalysis: Concepts and practice," in *Mathematics and Computation in Imaging Science and Information Processing, Ed: World Scientific*, pp. 177-207, 2007.
[10] M. Mishra, G. Tiwari, and A.K. Yadav, "Secret communication using public key steganography," in *International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014)*, pp. 1-5, 2014.
[11] G. Swain and S. K. Lanka, "A quick review of network security and steganography," *International Journal of Electronics and Computer Science Engineering*, vol. 1, pp. 426-435, 2012.
[12] M. Douglas, K. Bailey, M. Leeney, and K. Curran, "An overview of steganography techniques applied to the protection of biometric data," *Multimedia Tools and Applications*, vol. 77, pp. 17333-17373, 2018.
[13] R. Böhme, "Principles of modern steganography and steganalysis," *in Advanced Statistical Steganalysis*, Ed: *Springer*, pp. 11-77, 2010.
[14] S. Tan and B. Li, "Targeted steganalysis of edge adaptive image steganography based on LSB matching revisited using B-spline fitting," *IEEE Signal Processing Letters*, vol. 19, pp. 336-339, 2012.
[15] J. Zhang, H. Huang, L. Wang, H. Lin, and D. Gao, "Coverless Text Information Hiding Method Using the Frequent Words Hash," *International Journal Network Security*, vol. 19, pp. 1016-1023, 2017.
[16] M.S. Hana'a, "A Natural Language Steganography Technique for Text Hiding Using LSB's," *Engineering and Technology Journal*, vol. 26, pp. 351-364, 2008.
[17] R. Poornima and R. Iswarya, "An overview of digital image steganography," *International Journal of Computer Science and Engineering Survey*, vol. 4, pp. 23, 2013.
[18] L.Y. Por and B. Delina, "Information hiding: A new approach in text steganography," in *7th WSEAS Int. Conf. on Applied Computer & Applied Computational Science (Acacos '08)*, Hangzhou, China, 2008
[19] G. Nehru and P. Dhar, "A detailed look of audio steganography techniques using LSB and genetic algorithm approach," *International Journal of Computer Science Issues (IJCSI)*, vol. 9, pp. 402, 2012.
[20] C. Baldwin, A. MacCormack, and J. Rusnak, "Hidden structure: Using network methods to map system architecture," *Research Policy*, vol. 43, no. 8, pp. 1381-1397, 2014.
[21] G.C. Kessler and C. Hosmer, "An overview of steganography," in *Advances in Computers*, Ed: *Elsevier*, vol. 83, pp. 51-107, 2011.
[22] A.A.J. Altaay, S.B. Sahib, and M. Zamani, "An introduction to image steganography techniques," in *International Conference on Advanced Computer Science Applications and Technologies (ACSAT)*, pp. 122-126, 2012.
[23] C. Sumathi, T. Santanam, and G. Umamaheswari, "A study of various steganographic techniques used for information hiding," *arXiv preprint arXiv:1401.5561*, 2014.
[24] A. Biryukov, I. Pustogarov, and R.-P. Weinmann, "Trawling for tor hidden services: Detection, measurement, deanonymization," in *2013 IEEE Symposium on Security and Privacy*, pp. 80-94, 2013.
[25] S. Alsaleem, "Automated Arabic Text Categorization Using SVM and NB," *Int. Arab J. e-Technol.*, vol. 2, pp. 124-128, 2011.

BIOGRAPHY OF AUTHOR

Nada Abdul Aziz Mustafa, her born at September 1st 1966, Scientific Degree: Msc. in computer science. Research Projects in The Felid of Specialization to The Environment and Society or the Development of Education: (a) Design and Implementation: Proposed Encoding and Hiding Texts in an Image, University of Sulymaina, 2010. (b) Utilizing Computers in Teaching, University of Baghdad, Conference, 2010. (c) A proposed Technique for Information Hiding Based on DCT, IJACT, South Korea, 2011. (d) Encryption a text using affine cipher and hiding it in the colored image by using the Quantization stage, University of Mustansiriya, Conference, 2013. (e) Face detection and the effect of contrast and brightness, JNIT, South Korea, 2014. (f) The Effect of the Smoothing Filter on an Image Encrypted By the Blowfish Algorithm Then Hiding It in A BMP Image, JNIT, South Korea, 2014.