A FRAMEWORK FOR SECURING SECRET MESSAGES USING UNICODE IN TEXT AND IMAGE STEGANOGRAPHY

[1]Mrs. Sumathy Kingslin, [2]R. Anusuya
[1]Research Supervisor, PG and Research, [2]Research Scholar, PG and Research
[1,2]Department of Computer Science
[1,2]Quaid E Millath Government College for Women, Chennai-02

Abstract - Message Security - Hiding a secret message when* sending *it to an authorized user without any attack is the most difficult task. Steganography is one such method for sending a secret text across the secure communication medium and protect the information throughout transmission. Background: Communication is widely carried out using network through email, sms, etc., but a sending of secret text through insecure channel is a big challenge. The limitations of text being sent via the communication medium and overhead involves are in terms of time, payload, etc., in this paper, text* steganography *technique are used to protect the secret text using Unicode characters. The similarities of glyphs provide same as alphabetic characters and the overall increase of hiding capacity. Result: The results prove that secret data, embedded in the cover text using Unicode characters preserves secrecy. The embedding and extraction done efficiently. The secret text (stego-text) is again embedded under a cover image it's provides double security to the secret text. Conclusion, the proposed technique successfully secure the secret data and achieves the high payload capacity, the text length are unlimited.

Keywords: Text Steganography, Unicode Characters, Binary Value, QuickStego Tool, E-Mail, Text Hiding.

1. INTRODUCTION

A Steganography (pronounced STEHG-uh-NAH-gruhf-ee, from Greek steganos, or covered," and graphic, or "writing") is the hiding of a top-secret message within an ordinary message and the extraction of it at its destination [14]. Steganography outplays cryptography a step beyond by hiding an encrypted message so that anybody will definitely fail to recognize the presence of secret message under it.

Digital steganography, information is first encrypted by the common methods and then inserted; using a different algorithm, into redundant data that is part of a specific file format such as a JPEG image, text format, audio format and video format. Information is more secure when it's hidden under various cover file formats [14]. The secret data hidden under the cover media will not be visible to the naked eye. So, unsuspicious users fail to notice the presence of secret message. The main goal of steganography is to protect the secret data from unauthorized users without disturbing the cover media.

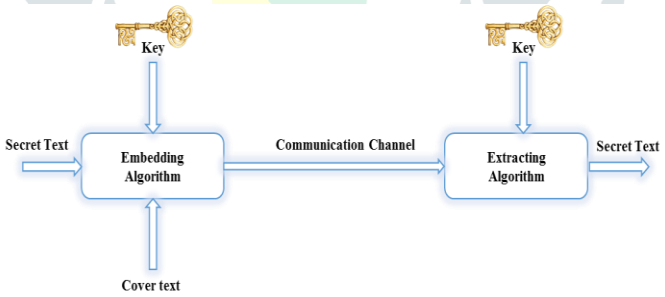


Figure 1. Basic Steganography process

1.1 Types of Steganography

There are many different form of steganography based on the cover media used for hiding the secret text. The various file format used decide the types of steganography on show in figure 2.

- Text Steganography
- Image Steganography
- Audio Steganography
- Video Steganography

Figure 2. Types of Steganography

1.1.1. Text Steganography:

Text steganography is a form of steganography that hides the secret under other cover text file. The commonly used methods in the scheme are [4]

- **Line-shift coding:** Text lines are vertically shifted to encode the document uniquely.
- **Word-shift coding:** The code words are coded into the document by shifting the horizontal locations of words within text lines, While maintaining a natural spacing appearance.
- **Feature coding:** In certain text features are altered, or not altered, depending on the code word.

1.1.2 Image Steganography:

Image Steganography is widely used method where in a image of a picture is used to hide the secret. [4],

The commonly used encoding techniques are

- **Least significant bit insertion (LSB)** - The technique works by substituting some of the data in a given pixel with data from the Information in the image. It has potential to embed data into an image on any bit-plane; LSB embedding is implemented on the least significant bit(s) [4]
- **Most significant bit insertion (MSB)** – This technique used to substituting some of the data in a given pixel with data from the Information in the image. It works in higher order bit. It has potential to embed the data into an image on any bit plane.

1.1.3 Audio Steganography

Audio Steganography is a technique used to communicate hidden information by modifying an audio signal in an undetectable manner. [4] It is the science of hiding some secret text or audio information in a host message [4].

1.1.4 Video Steganography

The Video Steganography makes use of a video as a cover (carrier file) to hide some secret data inside the video file by using some special embedding process. [4]

1.2 UNICODE Standard

The Unicode Standard is a Worldwide Character encoding scheme for writing characters and text, it has the ability to maintain a million characters and it can be implemented by various characters, symbols, and numbers to be encoded. Utmost commonly used encodings are UTF-8, UTF-16. UTF-8 is comparable to ASCII coding that uses the same 8-bit code [9] UTF-16 or UCS-2 uses two bytes for each character. Unicode characters are notable by code points, which are predictably represented by the letter U followed by four hexadecimal digits, for example, U+0043 or U+FF52. Unicode characters can range in scalar values from 0 to above a million.

2. PROPOSED METHOD IN TEXT STEGANOGRAPHY USING UNICODE GLYPHS

The proposed method makes use of Unicode character to secure the transmission of secret message over insecure channel. A glyph is an appearance of an exact shape which a character may have when rendered or displayed [9]. The glyph of English Alphabetical characters must be corresponding to the original characters. Proposed method uses all characters, uppercase, lowercase, special characters.

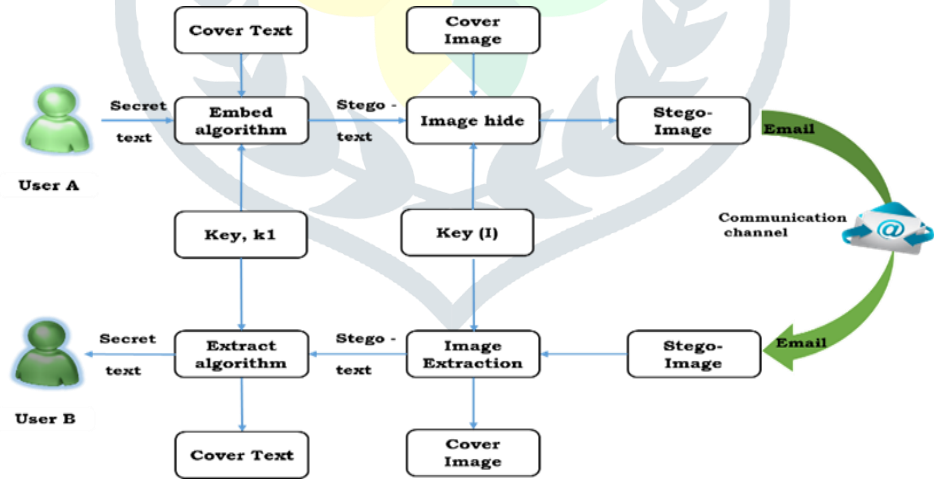


Figure 3. Proposed Methodology

The basic methodologies that describe the working of the proposed scheme is explained in figure 3. A sender hides a secret message using cover text with the help of Stego-Key, k1 to secure the data. The output of the embed algorithm is known as stego-text. Text encoded in ASCII or UTF-8 denotes every character as a series of bytes. Information must be serialized, so that it may be embedded bit by bit into the cover. The secret message must be no longer than the cover. The data may be embedded include text message encoded with ASCII values of characters. That Stego-text is again embed in the cover image. Image can send into a communication channel, E-mail and the receiver (User B) get the image from the e-mail. Receiver extracts the image using key (I), they get a stego-text and Key (k1). Again, the stego-text is extracted using of stego-key, finally, the receiver get the secret message from the stego-text.

2.1. Unicode Table

Unicode is a recent standard for text representation that describes each of the letters and symbols commonly used in today's digital and print media. Unicode has grown into the top standard for identifying characters in text in nearly any language [17]. It range start from

U+0000 to U+1F9FF which includes all kinds of language it has been supported [16]. But, these algorithms are used particular values to encode and decode the secret data. The ranges used in this process are U+0021 to U+FF5E.

2.2. Embedding Process

The two algorithms are mainly used in proposed methodology which is embedded procedure (hiding the secret data) and another one is extracted procedure (revealing the secret data). When the secret message is hidden in the cover text, the secret message is converted into the binary values and each character contains 8-bits, that binary value is substituted in the cover text using text steganography methods, each bit assigned to every single character in the cover text. Besides, the Unicode progression happens in this phase, the Unicode is a two-way process U+0043 or U+FF52, if the cover text of the binary value is '0' means it assumes the value U+0043. If the cover text of the binary value is '1' means it assumes the U+FF52. It has a list of Unicode tables for all the characters - uppercase, lower case, special characters, and numbers. It's a pre-defined value of the computer system. U+0043 and U+FF52 are similar to the original English alphabets.

Cover text: "Research comprises "creative work undertaken on a systematic basis in order to increase the stock of knowledge. A research project may also be an expansion of past work in the field. [14]**"**

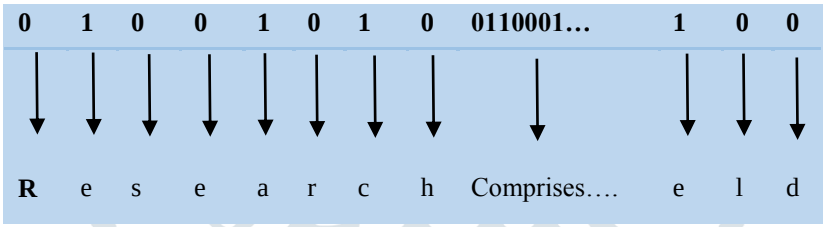
Secret Message: **"**Jai Hind**"**

A cover text is given as a paragraph text; the secret message is hidden in the Cover text.

For example: (J-01001010, a- 01100001, i-01101001)

[Jai Hind = 01001010 01100001 01101001 01001000 01101001 01101110 01100100]

If the secret text is hidden in the cover text using this kind of text steganography method.



This Stego-text is hiding into an image using a quick stego tool.

Embedding Algorithm

Input: *Cover text, C, Secret Message, Stegokey, k1*
Output: *Stego-text*

(1)Compute the number of characters required to hide the secret text, [n]
(2)Choose a cover text of length l, such that n<=l.
(3)Convert the secret text into the ASCII bits (8 bits) [b] equivalent.
*b = (n * 8) <= l,*
(4)Read a bit of the message and Read a letter from the cover letter.
(5) for each bits in the cover text, b
***if** bit = 1 then*
Replace with Unicode equivalent from the Unicode table 1 using glyphs
***else**, bit=0 no change to the cover*
*(6) Repeat **step 5**, until all the bits of binary value embedded*
(7) Return the stego-text

2.3. Extraction Process

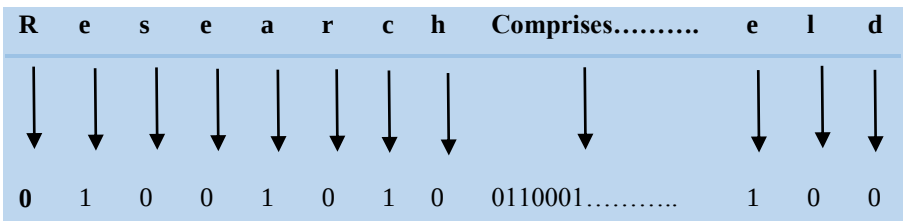
The Stego-Image is received by the authorized user. The extraction process is performed at the user system using the same table used for embedding. A Stego-Image is extracted using a tool, and then the user gets the Stego-text. The stego-text is carrying the secret message and check for the Unicode process. The secret message is extracted with the help of secret Key used by the sender.

The output of extraction process,

For example: (J-01001010, a- 01100001, i-01101001)

[Jai Hind = 01001010 01100001 01101001 01001000 01101001 01101110 01100100].

The value of secret message is given below in the table. 01001010 are allotted for the cover text as a sentence.



The Unicode character is converted into the binary value and then the binary values are replacement into a secret message from the cover text.

Extraction Algorithm

Input: *Stego-Image, k1*
Output: *Secret Text*

(1) Open the Stego-text
(2) Read a number of words (n) from the stego file
(3) l = length of the word,
for each bit in the cover text
*b = (n * 8) <= l,*
Check the code of the selected characters.
(4) ***If*** *the code is within U+0043 to U+FF52*
then *the secret message is extracted*
else *secret message is nothing based on a binary table or table 1.*
(5) Repeat step 4, until all the bits of binary value extracted.
(6) Return Secret Message.

3. RESULTS AND ANALYSIS

The sample outputs of the embedding algorithm are shown in figure. Steganography involves two portions of data: the cover text, and the secret data to be hidden. The approaches for embedding information vary in their complexity and its purpose.

Table 1. Unicode table

Binary Value	Unicode	Original Alphabet	Substituted Alphabet
0	U+0052	R	R
1	U+FF45	e	e
0	U+0073	s	s
0	U+0065	e	e
1	U+FF41	a	a
0	U+0072	r	r
1	U+FF43	c	c
0	U+0068	h	h

3.1. Results

Sample 1:

"Research comprises "creative work undertaken on a systematic basis in order to increase the stock of knowledge".

Figure 4. Cover text

Jai Hind

Figure 5. Secret text

"Research comprises creative work undertaken on a systematic basis in order to increase the stock of knowledge".

Figure 6. Stego-text

Sample 2:

Physical fitness is a general state of health and well-being and, more specifically, the ability to perform aspects of sports, occupations and daily activities. Physical fitness is generally achieved through proper nutrition, moderate-vigorous physical exercise.

Figure 7. Cover text

Meet 5'o clock, & War started

Figure 8. Secret text

Physical fitness is a general state of health and well-being and, more specifically, the ability to perform aspects of sports, occupations and daily activities. Physical fitness is generally achieved through proper nutrition, moderate-vigorous physical exercise.

Figure 9. Stego-text

3.2. Analysis

The goodness of the proposed method is analyzed using similarity ration and security measures of steganography. The similarity of the cover and stego-text speaks about how invisible is the secret message to the naked eye. This is measured using two techniques of Histogram measures and PSNR ratio.

(1) Using a Histogram
(2) Using PSNR ratio

3.2.1. Similarity Ratio

The RATIO OF SIMILARITY between any two similar figures is the ratio of any pair of equivalent sides. Once it is determined that two figures are similar, all pairs of an image is similar to ratio.

3.2.1.1 Similarity measures using Histogram

The similarity ratio between the original image and stego-Image. The similarities of the cover and stego-text are compared using histograms generated from figure. It is clearly seen that both histogram are very much identical there by there are minimal differences between cover and stego-text.

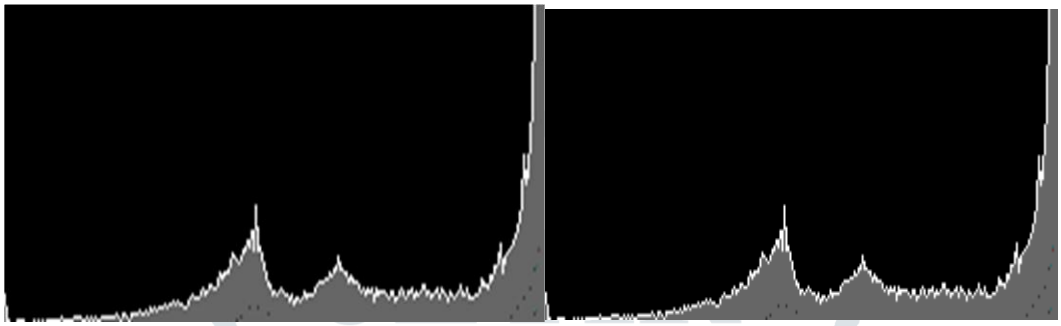


Figure 10. Original Image **Figure 11. Stego-Image**

3.2.1.2 Using PSNR ratio

Peak signal-to-noise ratio, (PSNR) explains how much the secret text has influence on the cover text. The higher the value indicates that there is less noise thereby increases the quality of the image and finding the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the reliability of its representation. Some of the noise is very high dynamic range. This is mainly used for image compression process and identified the noise occurred in the Image. The signal in a mean that is the original data, and the noise is the error during data compression in an image So that, PSNR is used to identify the highest quality of a compression using equations, PSNR is most easily defined via the mean squared error (MSE) [7].

$$MSE = \frac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \quad \textbf{(1)}$$

The PSNR (in dB) is defined as:

$$\begin{aligned} PSNR &= 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \\ &= 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \\ &= 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE) \end{aligned} \quad \textbf{(2)}$$

Table 2. Calculating the PSNR value and quality of Image

Sample Images	Cover Text Length	Secret Text Length	MSE Value	PSNR Value
Shield	70	8	0.6224	36.21
Books	175	20	3.2001	44.23
House	100	12	1.0258	52.34
Bike	480	60	16.9356	42.22
Cat	240	29	5.7369	34.31
Lion	300	37	10.4245	31.53
Whale	400	50	12.7891	21.58
Rupees	221	28	9.2553	26.52
Window	698	87	18.0365	39.02
Phone	352	44	12.4454	59.25

The measure of the MSE value and PSNR value for an Stego-Image and cover image. The MSE and PSNR is commonly measures for two error metrics are using to compare the stego-image. The MSE ration represent the cumilative squared error between original and stego-image. The table 2. Represent the value of stego-image. If the value for stego-image is high then the quality of the image is also high. Error must be depend upon the image quality.

3.2.2 Security Measure of text steganography

The Security Measure of text steganography is

- **Confidentiality** –The proposed method uses using Unicode characters the secret message highly invisible, because, the secret text Character is same as English alphabetic.
- **Integrity** – User sends the stego message along the communication medium which can be opened only by the authorized user. Unauthorized user unable to change or attack the secret message.
- **Availability** – The proposed method uses Unicode character for generating the stego-text which is very much similar to the English alphabets.
- **Security** - The secret message sends by using Unicode techniques and it's provide the security of the secret data and its hide using an image as a cover text. So the unauthorized user should know the key of an image and stego text. Otherwise, the hacker fails to access the secret message.

4 CONCLUSION AND FUTURE WORK

Steganography is the process securing the secret by the method of hiding, such that its presence cannot be noticed. In proposed method, the main objective of text steganography is confidentiality, integrity, availability, invisibility, and security. To examine the results are the Unicode process, Binary value produces the double the security of the top secret messages. This technique helps to hide a maximum number of character and high payload capacity and redundant performance. So, it can't easily extracted by the unauthorized user. The embedded secret text (Stego-Image) is sent using an e-mail. E-mails are sent to the authorized user without any interruption and disclosures. The character is not restricted; a user can give a cover of unlimited length and also a secret message. These criteria are satisfied in this proposed methodology. It is more secure through invisibility; the main factor of secret text is invisible to the unauthorized users, because the character of the cover text is interchanged as a Unicode character which is similar to the glyphs of the original Message. The stego-text is hidden in an image gives the double security for the secret message. The hiding capacity of the secret message is very high. Accordingly, the proposed method sustains the successful text steganography method by using Unicode, and an exceptional way to obtain safe and secure data transmission. This method can be extended to different languages and it can be implemented in cloud computing process or virtual machine process.

5 ACKNOWLEDGMENT

I would like to express my deepest gratefulness to Mrs. Sumathy Kingslin for her guidance and expert contributions to this paper. Without her support it would be impossible to complete this paper. I would also have to appreciate the guidance given by supervisor as well as the panels especially in my project presentation that has improved a lot. Thanks for their comments and advices.

REFERENCES

[1] Abdul Monem S. Rahma, Wesam S.Bhaya, Dhamyaa A. Al-Nasrawi, "Text Steganography Based On Unicode of Characters in Multilingual" Proceedings of International Journal of Engineering Research and Applications (IJERA) , Jul-Aug 2013, pp.1153-1165.

[2] Ananthi S, Sumathy K, "A Systematic Approach towards the techniques of Text Steganography", Proceedings of International Conference on Research Trends in Computer Science, 2013 Aug, pp. 88–96.

[3] Navneet Kaur, Sunny Behal, "A Survey on various types of Steganography and Analysis of Hiding Techniques" published on International Journal of Engineering Trends and Technology (IJETT), May 2014, pp.388-392

[4] Mohammad Shirali-Shahreza, M. Hassan Shirali-Shahreza, "Text Steganography in SMS", Proceedings of International Conference on Convergence Information Technology, 2007.

[5] Ananthi Sheshasaayee, D. Sumathy, "Text Steganography in SMS Using Similarity of Glyphs in Unicode Characters", Proceedings of Indian Journal of Science and Technology, November 2015.

[6] Wesam S. Bhaya, "Text Hiding in Mobile Phone Simple Message Service Using Fonts", Proceedings of Journal of Computer Science, 2011.

[7] https://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio

[8] Petitcolas, F.A., P. Anderson, R.M.G. Kuhn," Information hiding-a survey", 2007.

[9] M.Grace Vennice, et. Al, "Hiding the Text Information using Steganography", International Journal of Engineering Research and Applications (IJERA). Jan-Feb 2012, pp.126-131.

[10] Text to Unicode convertor. https://www.branah.com/unicode-converter.

[11] Vidhya P.M, Varghese Paul, "A Method for Text Steganography Using Malayalam Text", International Conference on Information and Communication Technologies (ICICT 2014) pp. 524 – 531.

[12] K. F. Rafat and M. Sher, StegRithm: "Steganographic Algorithm for Digital ASCII Text Documents". IACSIT International Journal of Engineering and Technology/December 2012.

[13] F. A. Haidari, A. Gutub, K. A. Kahsah, and J. Hamodi, "Improving security and capacity for Arabic text steganography using Kashida extensions," IEEE/ACS Int. Conf. On Computer Systems and Applications, 2009, pp.: 396-399.

[14] M. Agarwal, "Text Steganographic Approaches: A Comparison", International Journal of Network Security & Its Applications, 2013, pp: 91-106.

[15] https://en.wikipedia.org/wiki/Information_security.

[16] http://unicode-table.com/en,

[17] https://www.techopedia.com/definition/974/unicode