ISSN No. 0976-5697

Volume 2, No. 6, Nov-Dec 2011

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Word-Based LSB Image Steganography

Dr. Mohammed Abbas Fadhil Al-Husainy
Department of multimedia systems, faculty of science and information technology,
Al-Zaytoonah University of Jordan.
Amman-Jordan
dralhusainy@yahoo.com, alhusainy@alzaytoonah.edu.jo

Abstract: Steganography is the art and science of hiding important information by embedding message within other file. A new Least Significant Bit (LSB) steganography technique is presented in this work by treating the secret message based on its words content instead of its characters content. By using a specific small words dictionary by the sender and the receiver of the secret message, and representing each word in the secret message with an index in the words dictionary. Then embed these words indices, of the secret message, in the LSB of the pixels in the stego-image. The experimental results shown that the Word Based LSB technique will add more security to the secret message, reduces the distortion that will occur in the stego-image, increase the capability to hide very long secret message in a small stego-image, and minimize the time that is needed to hide and extract the secret message.

***Keywords*:** Security, Distortion, Embedding, Word Index, Collision

I. INTRODUCTION

Steganography can be defined as the technique used to embed data or other secret information inside some other object commonly referred to as cover, by changing its properties. The purpose of steganography is to set up a secret communication path between two parties such that any person in the middle cannot detect its existence; the attacker should not gain any information about the embedded data by simply looking at cover file or stego file. Steganography is the art of hiding information in ways that prevent the detection of hidden messages. Steganography, derived from Greek, literally means "covered writing." It includes a vast array of secret communications methods that conceal the message's very existence. These methods include invisible inks, microdots, character arrangement, digital signatures, covert channels, and spread spectrum [1, 2].

The basic model of steganography uses a cover object (any object that can be used to hold secret information inside), the secret message (the secret information that is to be sent to some remote place secretly), a stego key that is used to encode the secret message to make its detection difficult and a steganography algorithm/technique (the procedure to hide secret message inside cover object). The outcome of the process is the stego object which is the object that has the secret message hidden inside. This stego object is sent to the receiver where receiver will get the secret data out from the stego image by applying decoding algorithm/technique [1].

Recently, steganography is implemented by using digital media. Secret message is embedded inside digital cover media like text, images, audio, video or protocols depending upon the requirement and choice of the sender. Compared with the other types of steganography, the image steganography is most widely used. The reason behind the popularity of image steganography is the large amount of redundant information present in the images that can be easily altered to hide secret messages inside them, and because it can take advantage of the limited power of the human visual system (HVS). With the continued growth of strong graphics power in computer and the research being put into image based steganography, this field will continue to grow at a very rapid pace[1, 3, 4, 5].

Steganography has a wide range of applications. The major application of steganography is for secret data communication. Covert channels in TCP/IP involve masking identification information in the TCP/IP headers to hide the true identity of one or more systems. Cryptography is also used for the same purpose but steganography is more widely used technique as it hides the existence of secret data. Another application of steganography is feature tagging. Captions, annotations, time stamps, and other descriptive elements can be embedded inside an image, such as the names of individuals in a photo or locations in a map [1, 2, 5, 6].

Steganography can be also used to combine explanatory information with an image (like doctor's notes accompanying an X-ray).Steganography is used by some modern printers, including HP and Xerox brand colour laser printers. Tiny yellow dots are added to each page. The dots are barely visible and contain encoded printer serial numbers, as well as date and time stamps. The application list of image steganography is very long [1, 6].

In this paper, LSB image steganographyis used to hide information by performing a proposed word-based treatment of message instead of character-based treatment. In addition to this treatment, a specific type of words dictionary, like a code box that is used in the cryptography techniques, is used at the sender and the receiver site to add more security to the secret message. The Steganography technique is the perfect supplement for encryption that allows a user to hide large amounts of information within an image. Thus, it is often used in conjunction with cryptography so that the information is doubly protected; first it is encrypted and then hidden so that an adversary has to first find the hidden information before decryption take place [7, 8, 9, 10]. The problem with cryptography is that the encrypted message is obvious. This means that anyone who observes an encrypted message in transit can reasonably assume that the sender of the message does not want it to be read by casual observers. This makes it possible to deduce the valuable information. Thus, if the sensitive information will be transmitted over

unsecured channel such as the internet, steganography technique can be used to provide an additional protection on a secret message [2].

A good technique of image steganography aims at three aspects. First one is capacity (the maximum data that can be stored inside cover image). Second one is the imperceptibility (the visual quality of stego image after data hiding) and the last is robustness [7].

II. RELATED WORKS

When hiding information inside images usually Least Significant Bit (LSB) method is used. In the LSB method the 8th bit of every byte of the carrier file is substituted by one bit of every bit of the secret information [11]. The LSB method usually does not increase the file size, but depending on the size of the information that is to be hidden inside the file, the file can become noticeably distorted.

Ross J. Anderson and Fabien A.P. Petitcolas argued that every steganographic approach will have its limitations; they proposed an information theoretic approach using Shannon's theory for perfect secrecy [12]. In the methods that are proposed by H. Motameni and his colleague's one can embed at the dark corners of an image [13]. One can also embed the secret information in frequency domain by using Discrete Wavelet Transform method [14. In this method the embedding should be done at high frequency coefficients. P. Mohan Kumar and D. Roopa suggested that one can apply block matching procedure to search the highest similarity block for each block of the secret image and embed in LSBs of the cover image [15]. Mohammed A.F. AlHusainy employed different strategy in image steganography art by mapping the pixels of image to English letters and special characters [16]. Lisa M Marvel and CharlesG Boncelet proposed to hide at the inherent noise places [17]. Ran-Zan Wang and Yeh-shun Chen also did the two way block matching for image in image steganography [18]. But this approach is suspicious to the hackers. Xinpeng Zhang and his colleagues proposed an approach called "multibit assignment steganography for palette images", in which each gregarious colour that possesses close neighbouring colour in the palette is exploited to represent several secret bits [19]. In reference [20] authors have discussed a double substitution algorithm for encrypting at sender and decrypting at receiver and the embedding process was at 7th and 8th bit positions alternatively. In [21] an image steganography with palette based images is suggested. The method is based on a palette modification scheme, which can iteratively embed one message bit into each pixel in a palette based image.

In each iteration, both the cost of removing an entry colour in a palette and the benefit of generating a new one to replace it are calculated. If the maximal benefit exceeds the minimal cost, an entry colour is replaced. It is found that the fundamental statistics of natural images are altered by the hidden non-natural information [22]. But if we do not touch the bytes those carry the image features and embed in the other bytes then the problem can be solved. As LSB embedding is very common, many steganalysis tools are available for it [23]. So LSB embedding is no more secured now-a-days. So, new embedding techniques are to be welcomed to the steganographic world. Due to the large number of steganographic tools available over the internet, a particular threat exists when criminals use steganography to conceal their activities with in digital images in cyber space. Reference [24] presents two JPEG steganographic methods using Quantization Index Modulation (QIM) in the Discrete Cosine Transform (DCT) domain. The two methods approximately preserve the histogram of quantized DCT coefficients, aiming at secure steganography against histogram-based attacks.

III. CLASSIC LSB STEGANOGRAPHY TECHNIQUE

The Least Significant Bit (LSB) steganography technique works by representing each character (byte) of the secret message as a set of 8-bits (where 1byte ≡ 8bits). And then hide/substitute the bits of the characters in the least significant bit of the pixels in the stego-image. When the secret message has *n* characters, then LSB technique need at least (*n**8) pixels in the stego-image to hid the bits of the *n* characters.

By replacing LSB of each pixel in the stego-image with one bit (from the 8-bits) of each character in the secret message, this replacement operation will cause some distortion/noise in the stego-image. By using Human Visual System (HVS), the attackers may doubt that the stego-image contain a secret information in it. In general, whenever the length of the secret message (i.e., number of characters) increase, then the noise in the stego-image probably will increase as a result. This will make a restriction in hiding a very long message in a small stego-image. Therefore, we will tend to choose a short message to hide it in a large stego-image to minimize the noise that is happened in the pixels of the stego-image and to put aside the doubt about containing the stego-image any secret information.

Also, when any attacker success to know that the stego-image has a secret message, it is easy to get this message by reconstructed it from the LSB of the pixels in the stego-image.

IV. THE PROPOSED WORD-BASED LSB STEGANOGRAPHY TECHNIQUE

A new perspective of LSB image steganography technique is presented in this paper. The main idea behind this technique is treating the contents of the secret message as a set of words instead of as a set of characters. And using a specific words dictionary, at the sender and the receiver of the secret message, to represent each word in the secret message as a number and hid the bits of these numbers in the Least Significant Bit (LSB) of the pixels in the stego-image. This will give the following strong points to Word-Based LSB technique.

a. Add more security to the secret message.
b. Increase the capability of hiding very long secret message in a small stego-image.
c. Decrease the noise that is appearing in the stego-image because the change that may happen in the LSB of the pixels in the stego-image.
d. Minimize the time that is needed to hide and extract the secret message.

The following paragraphs give a detail explanation of the Word-Base LSB technique.

A. Words Dictionary:

Before starting to hide the secret message in the stego-image, both sender and receiver choose a specific small

dictionary of words; this dictionary might contain any word even though fake words. Each word in the dictionary has an index number that is calculated from the following formula:

$$WordIndex = \sum_{k=1}^{n} Order(k) \qquad (1)$$

Where *n* is the number of letters in the word, and the *Order* of the alphabetic English letters is shown in Table 1:

Table 1: Order of Alphabetic English Letters.

Order	Letter		Order	Letter
1	a		14	n
2	b		15	o
3	c		16	p
4	d		17	q
5	e		18	r
6	f		19	s
7	g		20	t
8	h		21	u
9	i		22	v
10	j		23	w
11	k		24	x
12	l		25	y
13	m		26	z

To clarify how the *WordIndex* calculated from the above formula, consider we have the English word **"message"**: the *WordIndex*of this word is calculated as:

***WordIndex*("message")= 13+5+19+19+1+7+5 =69**

Most English words, even the long one, will have *WordIndex* value <255, this means that we need only 1-byte (8-bits) to represent the *WordIndex* value of each word in the words dictionary. And the number of indices in the words dictionary table is not exceeds 255.

But really, when we are using the above formula to calculate the *WordIndex* of many, more than one word in the words dictionary might have the same value of *WordIndex*. This means that a ***collision*** might occur in the indices of the words dictionary. For example:

***WordIndex*("message") = 13+5+19+19+1+7+5 = 69**
***WordIndex*("financial") = 6+9+14+1+14+3+9+1+12=69**

It's an easy solved problem, by using a specific small words dictionary at sender and receiver sides, the number of collision indices can be reduced. And for the remained few collision words in the words dictionary, it's not a hard job that the receiver can distinguish between the correct word and the wrong words from the context of each sentence in the secret message. For the above two collision words, it's easy to choose the correct word that is suitable to the context of the following sentences.

"the problem with cryptography is that the encrypted (message/~~financial~~) is obvious. this means that anyone who observes an encrypted (message/~~financial~~) in transit can reasonably assume that the sender of the (message/~~financial~~) does not want it to be read by casual observers"

When the sender and the receiver are using the mentioned specific small words dictionary (and this dictionary may contains many fake words), this will put additional difficulty in front of attackers to know what words that are really contained in the secret message. Using this type of a specific words dictionary, in the proposed Word-Based LSB technique, works as a substitution box that is usually used in many encryption techniques.

B. Word Indices Instead of Message Characters:

After choosing a specific small words dictionary by sender and receiver, the sender convert the secret message to be represent as a set of indices (bytes) of all the words in this message. The length (number of bytes) in the new secret message equals the number of words in the source secret message. As a result, the number of bytes of the new secret message is too less than the characters (bytes) of the source secret message. For example, the number of characters (bytes) of the following sentence equals (100 bytes), but when we represent each word in the sentence as its *WordIndex* value, the number of bytes that is needs to represent this sentence equals (16 bytes):

"steganography is the art of hiding information in ways that prevent the detection of hidden messages"

It's a strong point of the new Word-Based LSB technique, because it's decrease the number of bytes that is needed to hide in the stego-image. This is lead to decrease the distortion/noise that is occurred in the stego-image and minimize the time that is needed to hide and extract the secret message.

C. Embedded Word-Indices in the LSBs of the Stego-image Pixels :

From the section 4.2., we get a set of indices (bytes) for all the words in the secret message. Now, implement the classic LSB to hide (substitute) the bits of the indices (bytes) in the LSB of the pixels in the stego-image (as explained in section 3 above).

D. Extract the Hidden Word-Indices of the Secret Message:

When the receiver receives the stego-image that is containing the hidden indices of the words of the secret message, he/she re-collect the bits of the indices from the LSB of the pixels in the stego-image. Then represent these bits as a set of bytes (such that, these bytes represent the indices (in the receiver words dictionary) for each word in the source secret message. After excluding any collision might be occurred in some words, the receiver will recover the source secret message that is hided from the sender in the stego-image.

V. EXPERIMENTS AND DISCUSSIONS

To demonstrate the performance of the Word-Based LSB technique, the technique is implemented to hide some messages into different stego-images. The set of (.bmp) images that are used in the experiments is listed in Table 2.

Table (2): Bitmap Stego-Images (Width × Height × Palette)

 Ice (600×479×3)	 Boat (64×64×3)


Sunset (600×450×3)


Butterfly (128×128×3)

To give the reader of this paper an ability to make a comparison between the classic LSB and the proposed Word-Based LSB technique. The two techniques are used to hide the same messages in the same stego-images; the Signal to Noise Ratio (SNR), the size of the embedded information, and the time that is needed to hide the secret information in the stego-images are recoded in Table 3.

Table (3): Performance Comparison between Classic LSB &Word-Based LSB Techniques

Message Length (Character)		Ice	Boat	Sunset	Butterfly
		6000	2000	3500	1000
Classic LSB	Size of Embedded Information (Byte)	6000	2000	3500	1000
	Time of Embedding Process (Second)	0.292	0.069	0.121	0.021
	SNR (db)	59.67	48.29	43.31	43.59
Word Based LSB	Size of Embedded Information (Byte)	942	314	553	157
	Time of Embedding Process (Second)	0.153	0.014	0.020	0.008
	SNR (db)	67.34	54.85	50.02	47.65

From the nature of the operations that are doing in the proposed Word-Based LSB and the result numbers from the experiments in Table 3, we can record the following strong points about the proposed Word-Based LSB technique:

a. Using specific small words dictionary at the sender and receiver side will add more difficulties in front of attackers, because any attacker must check all possible word might have the same indices. Even if the attacker did that, still there is a possibility that the sender and the receiver used a set of fake words in their dictionary. This strategy is look like the strategies that are using in the encryption/cryptography techniques when they are employing a code table/book to encrypt the secret information.
b. The size of the embedded information (in byte) in the Word-Based LSB is too less when it is comparing with the embedded information in the classic LSB.
c. Because the Word-Based LSB depend in its operations on the numbers of word in the secret message instead of the number of characters in the secret message (as its doing in the classic LSB).This strategy give a capability to the Word-Based LSB in hiding a very long message in a small stego-image while the classic LSB suffer from this point.
d. Because the size of the embedded information is small, this means that the number of least significant bit of the pixels in the stego-image that are required to hide the secret information become fewer than in the classic LSB. This will lead to decrease the distortion (Signal to Noise Ratio (SNR)) that is occurred in the stego-image as it shown in Table 3.
e. Also, whenever the size of the embedded information becomes small, the time that is spent in the hiding and extracting information in/from the stego-image is minimize.

VI. CONCLUSION

A Word-Based LSB image steganography technique was proposed in this paper. The new perspective here is that the propose LSB technique treat the secret message as a set of words rather than a set of characters, and using a specific words dictionary in the sender and the receiver sides. When testing this technique in hiding secret messages in different stego-images. The performance and the recorded results shown that the strategy that is using in the Word-Base LSB technique gave the technique a set of powerful points: (1) Add more security to the secret message, (2) Increase the capability to hide a very long message in a small stego-image, (3) Decrease the distortion/SNR that is occurred in the pixels of the stego-image, (4) Minimize the time that is required to hide/extract the secret information in/from the stego-image. From the above points, we can conclude that the Word-Based LSB technique promise to use it as a one of effective steganography technique.

VII. REFERENCES

[1] Cheddad, J. Condell, K. Curran, &P. Kevitt, (2010). Digital image steganography- survey and analysis of current methods.Signal Processing, 90, 727–752.doi:http://10.1016/j.sigpro.2009.08.010

[2] Adnan Gutub, Ayed Al-Qahtani, &AbdulazizTabakh. (2009). Triple-A: secure RGB image steganography based on randomization. AICCSA, IEEE/ACS International Conference on Computer Systems and Applications, Rabat, Morocco, 400-403,doi: http://doi.ieeecomputersociety.org/10.1109/AICCSA.2009.5069356

[3] Kaur, R. Dhir, &G. Sikka. (2009). A new image steganography based on first component alteration technique. International Journal of Computer Science and Information Security (IJCSIS),6, 53-56.http://arxiv.org/ftp/arxiv/papers/1001/1001.1972.pdf

[4] Alvaro Martin, Guillermo Sapiro,&GadielSeroussi. (2005). Is Steganography Natural. IEEE Transactions on Image Processing,14(12), 2040-2050.doi: 10.1109/TIP.2005.859370

[5] Bhattacharyya, A. Roy, P. Roy, &T. Kim. (2009). Receiver compatible data hiding in color image. International Journal of Advanced Science and Technology, 6, 15-24.http://www.sersc.org/journals/IJAST/vol6/2.pdf

[6] EE. Kisik Chang, J. Changho, &L. Sangjin. (2004). High Quality Perceptual Steganographic Techniques. Springer.2939, 518-531.doi:10.1007/978-3-540-24624-4_42, http://www.springerlink.com/content/c6guuj5xnyy4wj3c/

[7] C. Kessler. (2001). Steganography: Hiding Data Within Data. An edited version of this paper with the title "Hiding Data in Data". Windows & .NET Magazine. [Online] Available: http://www.garykessler.net/library/steganography.html (October 4,2011)

[8] GandharbaSwain,&S.K.lenka. (2010). Steganography-Using a Double Substitution Cipher. International Journal of Wireless Communications and Networking. 2(1), 35-39. ISSN: 0975-7163.http://www.serialspublications.com/journals1.asp?jid=436&jtype

[9] Hideki Noda, MichiharuNimi, &Eiji Kawaguchi. (2006). High- performance JPEG steganography using Quantization index modulation in DCT domain. Pattern Recognition Letters,27, 455-

46.http://ds.lib.kyutech.ac.jp/dspace/bitstream/10228/450/1/repository6.pdf

[10] Kathryn (2005). A Java Steganography Tool.http://diit.sourceforge.net/files/Proposal.pdf

[11] Motameni, M.Norouzi, M.Jahandar,& A. Hatami. (2007). Labeling method in Steganography. Proceedings of world academy of science, engineering and technology, 24, 349-354. ISSN 1307-6884.http://www.waset.org/journals/waset/v30/v30-66.pdf

[12] Zhang,&H. Tang. (2007). A novel image steganography algorithm against statistical analysis. Proceeding of the IEEE,19, 3884-3888.doi: 10.1109/ICMLC.2007.4370824

[13] Lisa M. Marvel,&Charles G. Boncelet. (1999). Spread Spectrum Image Steganography. IEEE Transactions on Image Processing,8(8), 1075-1083.doi: 10.1109/83.777088

[14] Mei-Yi Wu, Yu-Kun Ho, &Jia-Hong Lee. (2004). An iterative method of palette-based image steganography. Pattern Recognition Letters,25, 301-309.doi: 10.1016/j.patrec.2003.10.013

[15] Mohammed A.F Al Husainy. (2009). Image Steganography by mapping Pixels to letters. Journal of Computer Science,5(1), 33-38. ISSN 1549-3636.doi: 10.3844/jcssp.2009.33.38, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.165.7818&rep=rep1&type=pdf

[16] Mohammad Ali BaniYounes, &AmanJantan. (2008). A New Steganography Approach for Image Encryption Exchange by using the LSB insertion. IJCSNS International Journal of Computer Science and Network Security,8(6), 247-254.http://paper.ijcsns.org/07_book/200806/20080634.pdf

[17] M.T. Parvez ,&A. Gutub. (2008). RGB intensity based variable-bits image steganography. APSCC 2008 – Proceedings of 3rd IEEE Asia-Pacific Services Computing Conference, Yilan, Taiwan, doi: 10.1109/APSCC.2008.105

[18] N.F. Johnson, &J. Suhil.(2006). Exploring Steganography:Seeing the Unseen. Computing Practices.http://www.jjtc.com/pub/r2026.pdf

[19] P.MohanKumar,&D.Roopa (2007). An Image Steganography Framework with Improved Tamper Proofing. Asian Journal of Information Technology,6(10), 1023-1029. ISSN: 1682-3915.http://medwelljournals.com/abstract/?doi=ajit.2007.1023.1029

[20] Po Yuch Chen, & Hung Ju Lin.(2006). A DWT Based Approach for Image Steganography.International journal of Applied Science and Engineering, 4(3), 275-290.http://www.cyut.edu.tw/~ijase/2006/4-3(Microsoft%20Word%20-%20010-009-6).pdf

[21] Ran-Zan Wang, &Yeh-Shun Chen. (2006). High Payload ImageSteganography Using Two-Way Block Matching. IEEE Signal Processing letters,13(3), 161-164.doi: 10.1109/LSP.2005.862603

[22] Ross J. Anderson,&Fabian A.P. Petitcolas. (1998). On The Limits of steganography. IEEE Journal of selected Areas in communication,16(4), 474-481. Special Issue on Copyright and Privacy protection. ISSN 0733-8716.http://www.cl.cam.ac.uk/~rja14/Papers/jsac98-limsteg.pdf

[23] SorinaDumitrescu,&Xiaolin (2005). A New Framework of LSB Steganalysis of Digital Media.IEEE Transactions on Signal Processing,53(10), 3936-3947.doi: 10.1109/TSP.2005.855078

[24] Xinpeng Zhang, Shuozhong Wang,&Zhenyu Zhou. (2008). Multibit Assignment Steganography in Palette Images. IEEE Signal Processing Transactions,15, 553-556.doi:10.1109/LSP.2008.2001117

