Journal of Information Engineering and Applications
ISSN 2224-5782 (print) ISSN 2225-0506 (online)
Vol 2, No.9, 2012

An enhanced Least Significant Bit Steganographic Method for Information Hiding

Gabriel Macharia Kamau[1*] Stephen Kimani[2] Waweru Mwangi[2]

1. School of Computer Science and Information Technology, Kimathi University College of Technology, PO box 657-10100, Nyeri, Kenya
2. Institute of Computer Science and Information Technology, Jomo Kenyatta University of Agriculture and Technology, PO box 62000-00200, Nairobi, Kenya

* E-mail of the corresponding author: ngorosh2003@yahoo.com

Abstract

The least significant bit (LSB) insertion method is a simple steganographic algorithm that takes the least significant bit in some bytes of the cover medium and swaps them with a sequence of bytes containing the secret data in order to conceal the information in the cover medium. However its imperceptibility and hiding capacity are relatively low. This is as revealed by the statistical characteristics of its resultant stego images compared to the original cover images. To increase the level of imperceptibility and the hiding capacity in the LSB insertion method, this research proposes an enhanced LSB method that employs a selective and randomized approach in picking specific number of target image bits to swap with the secret data bits during the embedding process. To facilitate the selective picking of the target image bits, the standard minimal linear congruential number generator (LCG) is used. The message digest (digital signature) of a user supplied password is used to seed the LCG and to extract the message from the cover medium. In measuring the effectiveness of the proposed method, the study adopted an experimental research design where the statistical characteristics of the proposed method stego images were compared with those of the traditional LSB method in a comparative experiment designed to establish the levels of image distortion (noise) introduced in the original cover image when either of the methods is used under the same payload and image. The experiment results indicated improved levels of imperceptibility and hiding capacity in the proposed method.

Key Words: Steganography, Steganalysis, Stego image, payload, imperceptibility

1. Introduction

According to Mohammad and Abdallah (2008), Steganography "is the art and science of writing hidden messages inside innocent looking containers such as digital files, in such a way that no one apart from the sender and intended recipient realizes the existence of the hidden message". The secret message is normally embedded in a cover medium known as a stego file in a way that totally conceals the existence of any form of communication going on. Digital images are the most widely used cover files in the world of digital steganography. The reason for this is because the human visual system can hardly pick the difference between an original image and a stego image when embedding of secret information is properly done.

One of the most popular and commonly used steganographic algorithms for information hiding in digital images is the LSB insertion method. It is a simple algorithm that swaps the least significant bit in some bytes of the cover medium with a sequence of bytes containing the secret data to be hid. However, though the LSB algorithm hides data in the cover medium (image) in a way that is imperceptible to the human visual system (HVS), its imperceptibility to statistical steganalysis is relatively low. This is mainly because the significant bits of the secret message are hidden in the cover medium in a linear and deterministic pattern. Retrieval of secret data using steganalysis software tools therefore becomes relatively easy once the algorithm used is known. The following table shows the values of the statistical (perceptibility metrics values) of stego images produced through the use of the LSB insertion method. Same payload (a 12 kilo byte document) is hidden in each image.

Table 1. Perceptibility metrics data for the Traditional LSB Method.

IMAGE	A AD (db)	M S E (db)	SNR (db)	PSNR (db)	RS (db)	SPA (db)
Banana.jpg	0.1341	0.268	613629	2177125	19.47	19.54
Dancers.jpg	0.1514	0.304	538341	1903976	16.22	17.29
Graduation.jpg	0.6695	1.34	135300	436554	49.74	51.38
Office.jpg	0.151	0.302	696202	1931692	17.46	18.97
zhbackground.bmp	0.5011	1.003	75296	581631	39.33	39.51

An enhanced LSB method that employs a selective and randomized approach in picking the target image bits to swap with the secret data bits during the embedding process using the standard minimal linear congluential pseudo random number generator (LCG) is proposed. A hushed stego key (k) value is used as a seed to determine the set of selected numbers used for targeting specific image bits for data hiding.

2. Related Work

2.1 The Optimal LSB Insertion Method

This insertion method improves the stego-image quality by finding an optimal pixel after performing an adjustment process. Three candidates are picked out for the pixel's value and compared to see which one has the closest value to the original pixel value with the secret data embedded in. The best candidate is then called the optimal pixel and used to conceal the secret data (Chan and Cheng, 2004). This however makes the hiding capacity of the carrier image very low.

2.2 The Pixel Value Differencing (PVD) Method

The pixel-value differencing (PVD) method is proposed by (Wu and Tsai, 2003). In this approach, the payload of each individual pixel is different, and the resultant stego-image quality is extremely fine with perfect modification and invisibility. The resultant stego-images quality that the method produces is better in terms of human visual perception. However steganalysis is easy as the hidden message is not well spread across the entire image.

2.3 Blind Hide algorithm

According to (Bailey, K. and Curran, K., 2006), this algorithm blindly hides the secret data in the image starting at the top left corner of the image and working its way across the image (then down - in scan lines) pixel by pixel changing the least significant bits of the pixel colors to match the message. To extract the hidden information, the least significant bits starting at the top left are read off. This embedding procedure is not very secure as it's really easy to read off the least significant bits starting from the top left corner of the image sequentially.

2.4 Algorithm Pixel Swap

This method is proposed by Lee *et al.* (2010). It works as follows

- Randomly select 2 pixels *x1* and *x2* from the cover image using a pseudo–random sequence.
- If the two pixels lie within a specified distance α (α=2 or 3 generally), they are suitable for embedding, otherwise generate another set of pixels.
- Take the specific message bit to hide. If the message bit is zero, check if x1 > x2 otherwise swap x1 and x2 and hide the bit in the LSB of the pixel. Do the reverse operation if the message bit is one.

- For extracting the hidden message, select the pixels using the same pseudo-random sequence. Check if the 2 pixels are within the pre-specified range α. If x1>x2, the message bit is zero (one) otherwise the message bit is one (zero).

This method does not add visible distortions to the cover image since only one bit is changed per pixel but its hiding capacity is highly limited.

3. The Proposed Method

The standard minimal linear Congruential Generator (LCG) method (Park and Miller., 1988) is used to generate the pseudo random numbers used to match the specific bits in the cover image where the secret data bits are hid. This is one of the most successful random number generators particularly with computer memory.
The formula is explained below.

$$\mathbf{X_{n+1} = (aX_n + c) \bmod m}$$

Where:
X_0 is the starting value , the seed ; $0 <= X_0 < m$
a is the multiplier; $a \geq 0$
c is the increment; $c \geq 0$
m is the modulus; $m > X_0$, $m > a$, $m > c$

The desired sequence of random numbers < Xn > is then obtained by setting
$X_{n+1} = (aX_n + c) \bmod m, \quad n \geq 0$

X_n is chosen to be in [0, m-1], $n \geq 0$

Given that the previous random number was X_i, the next random number X_{i+1} can be generated as follows.
$X_{i+1} = f(X_i, X_{i-1,\ldots,} X_{i-n+1}) (\text{mod } m) = (a_i x_i + a_2 \; x_{i-1} + \ldots + a_n x_{i-n+1} + c) (\text{mod } m)$

The communicating partners share a stego key (k) which in this case is the message digest of the user supplied password.

According to Hull & Dobell (1972), a linear congruential sequence defined by *m*, *a*, *c* and X_0 has full period if and only if the following three conditions hold:

- The only positive integer that exactly divides *m* and *c* is 1
- If *q* is a prime number that divides *m*, then *q* divides a -1
- If 4 divides *m*, then 4 divides *a* -1

Additionally, the value of *m* should be rather large since the period cannot have more than *m* elements. The value of *m* should also necessitate a fast computation of (aXn+c) i.e speed the generation of random numbers. Observing all these requirements, the parameters for the LCG used in this research are as follows:

3.1 Modulus (m)

The 48-bt computer word length was picked as the value of *m.* Any Pentium IV and above computer should have this word length or larger. This in essence provides the size of *m* to be 2^{48} which is equivalent to **281,474,976,710,656.** For the sake of this experiment and bearing in mind that the digital images used are a few kilobytes in size, this period is sufficient enough to set up the experiment.

To ensure faster generation, *m* is recommended to be a power of *2* or close to a power of *2* and hence the choice of the word length. The AND operation also enhanced speed instead of the normal division operation which is considered slower.

3.2 The seed (X_0)

The first value of the seed (X_0) is supplied by the message digest of the user supplied password. This is done using a special form of encryption that uses a one-way algorithm which when provided with a variable length unique input (message) will always provide a unique fixed length output called hash, or message digest.

3.3 a (Multiplier) and c (Increment)

To ensure full period and in following with requirements identified above, the values of the multiplier and the increment are picked as follows.

a (Multiplier) = 25214903917
c (Increment) = 11

The values are used to initialize the random number generator used for the prototype.
The numbers generated by this PRNG determines the specific bits in the pixel bytes of the cover image where data bits of the secret data file are to be embedded. For example considering storing the number 200, whose binary representation is 11001000 in a grid of 3 pixels of a 24-bit image, utilizing a single LSB of each color channel the enhanced LSB algorithm will store the significant bits of the message randomly into the cover image bits as shown in figure 1 and figure 2 below.

0	0	1	0	1	1	0	1	0	0	0	1	1	1	0	0	1	1	0	1	1	1	0	0
1	0	1	0	0	1	1	0	1	1	0	0	0	1	0	0	0	0	0	0	1	1	0	0
1	1	0	1	0	0	1	0	1	0	1	0	1	1	0	1	0	1	1	0	0	0	1	1

Figure 1. Original Image Bits

0	0	1	0	1	1	1	0	0	1	0	1	1	1	0	1	1	1	0	1	1	1	0	0
0	0	1	0	0	1	1	0	1	1	0	0	1	1	0	1	0	0	0	0	1	0	0	0
1	1	0	0	0	0	1	1	1	0	1	0	1	1	0	0	0	1	1	0	0	0	1	0

Figure 2. Modified Image Bits

4. Research Method

The study adopted an experimental research method to measure the effect of using selected varied and random pixels during the embedding process on imperceptibility and hiding capacity. This method represents the standard practice applied in manipulating independent variables in order to statistically analyze the generated data to test research hypotheses. A notable advantage of experimental research is the fact that it enables other researchers to easily replicate the experiment and be able to validate the results. It is therefore considered an accurate method of research (Shuttleworth, 2008), as the researcher can effectively establish a causal relationship between variables by manipulating independent variable(s) to assess the effect upon dependent variable(s).
The effect of dispersing the significant bits of the hidden message across a cover image during the embedding process represented the variable that we weree to understand and study. Thus an experiment was carried out to test the relationship between the specific embedding process (i.e. Proposed method) and the outcome (ie

imperceptibility level). Essentially the output of traditional least significant bit steganography method was used to evaluate the performance and effectiveness of the proposed method's output by comparing the stego images generated by the proposed method with those generated by the traditional least significant bit steganography method. This is commonly referred to as comparative experiment (Hinkelmann and Kempthorne, 2008).

5. Algorithm Design Framework

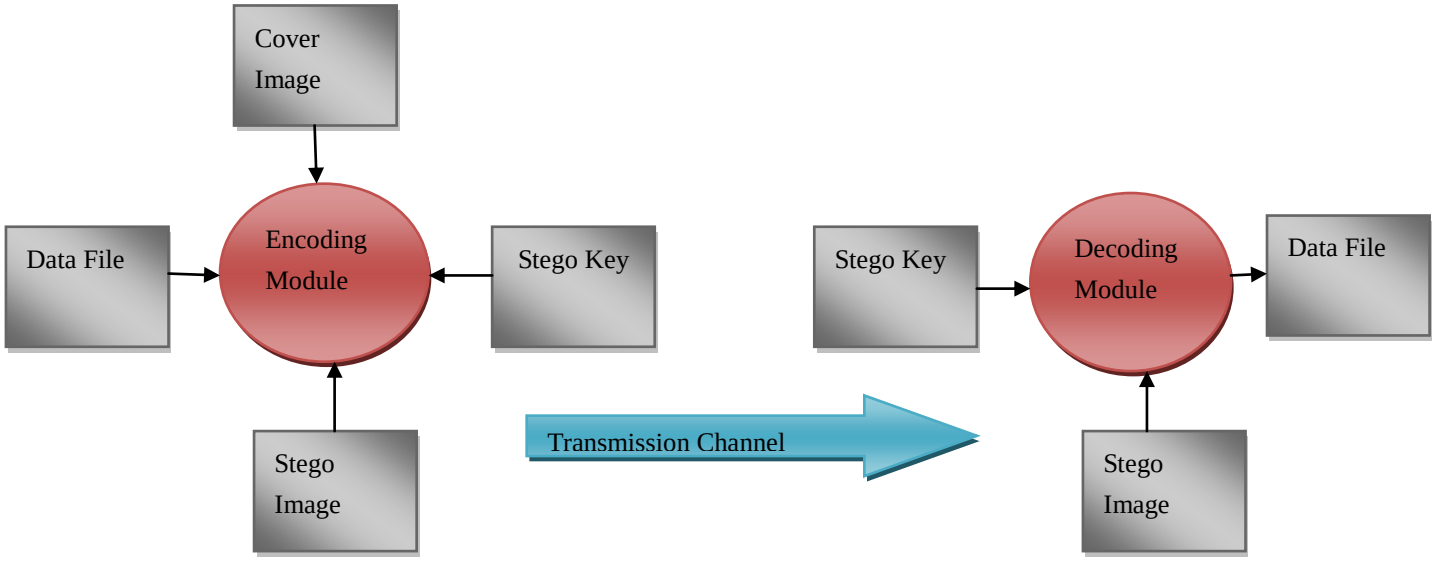

Figure 3. Framework for the proposed system(Pfitzmann, 1996)

5.1 Embedding and Extraction Algorithms

Input : Cover Image, Secret file (Payload)
Output : Stego image (image containing hidden file)
16. Use LCG to
Select a random pixel
Select a random pixel color channel
Select a random color channel bit
2. Let bitToWrite [x][y][channel][bit] denote the selected bit in a specific color channel for writing
3. Let m_i denote the message bit embedded in a color channel bit, bitToWrite[x][y][channel][bit]
4. For all image color channels do the following
5. If LSB (bitToWrite[x][y][channel][bit]) = m_i then
6. do nothing
7. If LSB(bitToWrite[x][y][channel][bit]) not equal to m_i then
8. bitToWrite[x][y][channel][bit] = m_i
9. while secret file length; Repeat step 16 to 23 to embed the entire message
10. Close stream.

Figure 4. Embedding Algorithm

Input : Stego Image, Password message digest
Output : Secret file

1. Use LCG to
   Select a random pixel
   Select a random pixel color channel
   Select a random color channel bit
2. Let bitToRead([x][y][channel][bit])denote the selected bit in a specific color channel for reading
3. Let m_i denote the message bit read in a color channel bit bitToRead([x][y][channel][bit])
4. For all image color channels do the following
5. If LSB(bitToRead([x][y][channel][bit]) not equal to m_i then
6. do nothing
7. If LSB (bitToRead([x][y][channel][bit]) = m_i then
8. bitToRead ([x][y][channel][bit])= m_i
9. Pack bit in bitSet
10. While secret file length; Repeat step 11 to 20 to read the entire file
11. Close stream.
12. Obtain the entire message stream and convert it back into ASCII format

Figure 5. Extraction Algorithm

6. Experimental Design and Testing

An analysis to examine the statistical properties of the stego images produced by the proposed method and the traditional LSB method was carried out. Statistical attacks are more powerful than visual attacks as they are able to reveal the tiniest modifications in the statistical properties of an image (Artz, 2001).
The following image quality metrics were employed for this purpose:

6.1 Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE)
Both of these metrics are the most common and widely used full reference metrics for objective image quality evaluation. In particular, PSNR is used in many image processing applications and considered as a reference model to evaluate the efficiency of other objective image quality evaluation methods (Wang *et al*., 2002b). PSNR as a metric computes the peak signal-to-noise ratio, in decibels, between two images. It is used in steganography to measure the peak signal-to-noise ratio in the original image and the stego image after embedding the hidden data. In the literature, PSNR has shown the best advantage almost over all other objective image quality metrics under different image distortion environments and strict testing conditions (Wang *et al*., 2002a).

On the other hand, MSE measures the statistical difference in the pixel values between the original and the reconstructed image (Stoica *et al*., 2003; Wang *etal*., 2003). The mean square error represents the cumulative squared error between the original image and the stego-image.

A lower MSE value means a better image quality ie lesser distortion in the cover image while the higher the PSNR value the better the quality of the image. (Mei Jiansheng *et al* .2009). PSNR and MSE are defined as shown in equations (1) and (2) below respectively (Stoica *et al.*, 2003; Wang *et al.*, 2003):

$$PSNR = 10.Log_{10} \; I^2/MSE \quad db \tag{1}$$

$$MSE = (1/MN) \sum_{i=1}^{M} \sum_{j=1}^{N} (X_{ij} - \overline{X_{ij}})^2 \tag{2}$$

Where:

X_{ij} is the i^{th} row and the j^{th} column pixel in the original (cover) image,

X_{ij} is the i^{th} row and the j^{th} column pixel in the reconstructed (stego) image,

M and *N* are the height and the width of the image,

I is the dynamic range of pixel values, or the maximum value that a pixel can take, for 8-bit images: I=255.

6.2 Reed-Solomon (RS) analysis

This is a method proposed by Fridrich for detecting the use of LSB steganography (Fridrich *et al,*.2001). RS measures the smoothness of the changes among pixels of an image (the lower the value, the smoother the changes among them, or the lesser the noise of the image). RS is one of the most reliable quantitative steganalysis methods (Fridrich *et al,*.2001).

In order to evaluate the performance of the proposed method an experimental design was set up. Stego images from both the traditional LSB method and the proposed method were compared using the testing metrics discussed above. All the experiments were implemented and run on a PC Pentium IV Duo core, 2.1 GHz with 2GB of RAM under the Windows 7 Home Edition operating system. The following constants were ensured.

- Same images were used on both the methods
- Same information was embedded in each image ie equal payload
- Same evaluation metrics were used for each image
- Five digital images were used as test data files (cover images). Table 2 shows the list of these digital images.

Table 2. Test data Images

FILE NAME	DIMENSIONS	FILE SIZE
Banana.jpg	685 x 514 Pixels	157 Kilo Bytes
Dancers.jpg	685 x 457 Pixels	224 Kilo Bytes
Graduation.jpg	685 x 457 Pixels	161 Kilo Bytes
Office.jpg	685 x 457 Pixels	163 Kilo Bytes
zhbackground.bmp	685 x 610 Pixels	1.19 MB

The specific data hiding steganogrtaphic method used was taken to be the independent variable (in this case the traditional LSB method and the proposed enhanced LSB method). In order to evaluate the efficiency of the proposed steganography method, the evaluation dependent variables all of which measure the image distortion levels were considered. Accordingly, for each steganography method (the traditional LSB method and the proposed enhanced LSB method) and for each cover image the value of each dependent variable was measured. The values of the dependent variables for both embedding methods were then compared.

7. Experimental Results

7.1 Peak Signal to Noise Ratio (PSNR)

Figure 5 below shows the comparison of the PSNR of the five stego images for both the traditional LSB method and the enhanced LSB method.

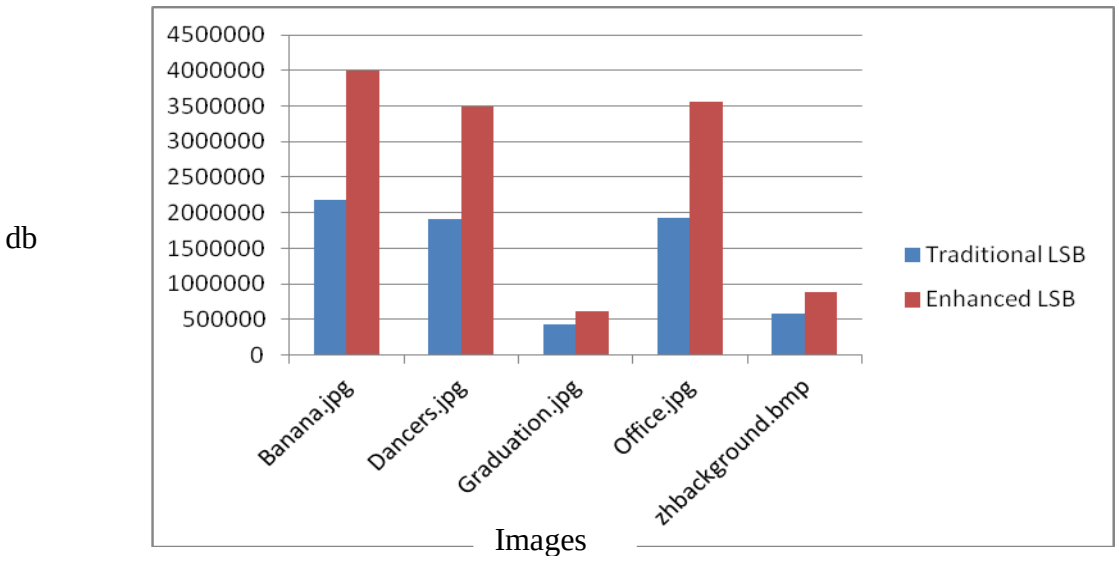

Figure 5: The PSNR (db) of stego images hiding in (Traditional LSB) vs hiding in (Enhanced LSB)

Every image tested registered a higher PSNR for enhanced LSB method as compared to the Traditional LSB method showing that the enhanced LSB embedding method distorts the image less improving on imperceptibility of the hidden data since a higher Peak Signal to Noise Ratio (PSNR) indicates less distortion (Mei Jiansheng *et al*,.2009).

7.2 Mean Square Error (MSE)

Figure 6 below shows a summary of the comparison of the MSE of five stego images for both the traditional LSB method and the enhanced LSB method.

For each stego image, a lower MSE was recorded with the enhanced LSB method as compared to the traditional LSB method. A lower MSE value means a better image quality ie lesser distortion in the cover image (Mei Jiansheng *et al*,.2009). This means that stego images generated by the enhanced LSB method have lesser distortions compared to those generated by the traditional LSB method and hence improved imperceptibility.

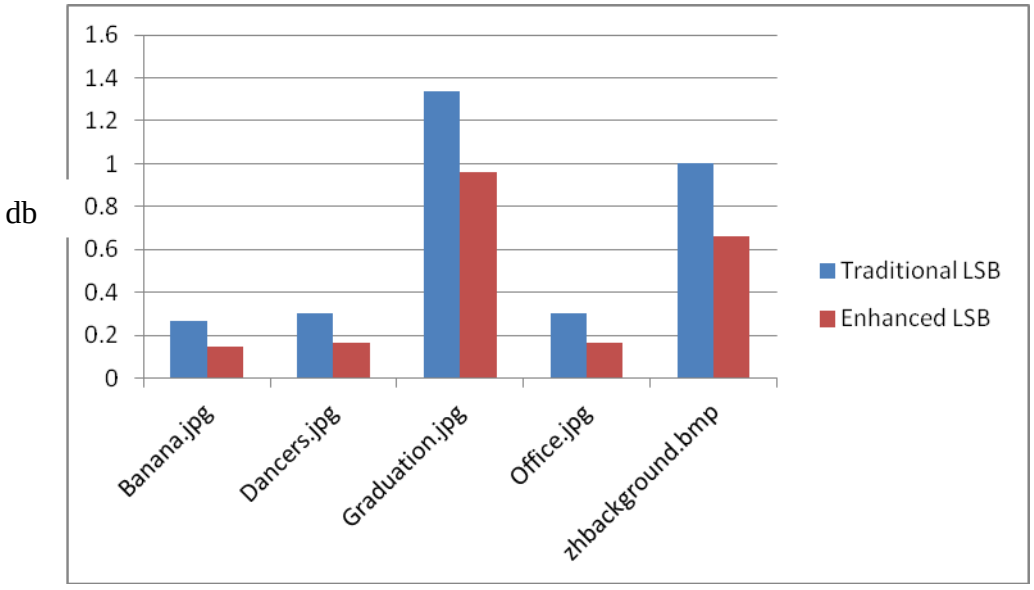

Figure 6: The MSE of stego images Hiding in (Traditional LSB) vs Hiding in (Enhanced LSB)

7.3 Reed-Solomon (RS) analysis

Stego images generated by the enhanced LSB method all recorded lower values of RS compared to those generated by the traditional LSB method as shown in figure 7. Therefore these images have lesser noise and the information hidden in them more imperceptible (Fridrich *et al*,.2001) .

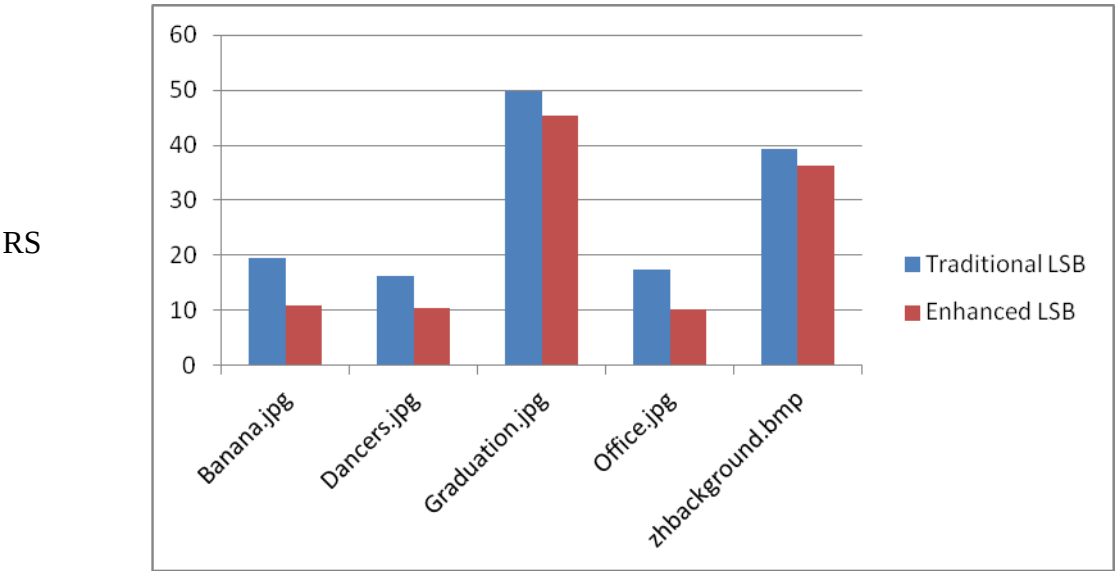

Figure 7: The RS of stego images hiding in (Traditional LSB) vs hiding in (Enhanced LSB)

9. CONCLUSION AND FUTURE WORK

In comparison to the traditional least significant bit algorithm, the data hiding steganographic method presented in this paper was found to demonstrate increased imperceptibility to statistical steganalysis attacks on the cover image. The hiding capacity can also be increased by varying the number of bits used per color channel. However this method is best suited for the purposes of communication and communication applications as more permanent aspects of steganography like watermarking are not included.

As with other steganographic applications, the cover images used should be high quality original photographs. The recommended mode of transmission of the stego images is through web postings or email attachments.

Digital steganography is a rapidly growing and increasingly interesting field of research for information hiding and data security. It is currently playing a vitally important role in defense and civil applications. In future, we are bound to see more of data security applications based on this technology.

In relation to this research, future work should centre on development of stronger embedding algorithms whose output can survive image manipulations and those that can make use of more permanent embedding procedures. This will facilitate the use of steganography in more sensitive application areas like in computer digital forensics and in enhancement of security in electronic commerce and trading applications. Research along these lines will also help in ensuring a permanent solution to the issues of plagiarism of copy write digital content materials.

References

Artz D. (2001) "Digital steganography: hiding data within data", *Internet Computing, IEEE*, vol. 5, Issue: 3, pp. 75-80

Bailey, K. and Curran, K. (2006) An Evaluation of Image Based Steganography Methods Using Visual Inspection and Automated Detection Techniques. *Multimedia Tools and Applications*, 31, 55-88.

Chan, and Cheng, L.M. (2004). Hiding data in images by simple LSB substitution. *Computer Journal of Pattern Recognition Letters*, vol. 37, no. 3, pp. 469-474

Fridrich, J., Goljan, M. & Hogea, D. (2002) Attacking the OutGuess. *The ACM Workshop on Multimedia and Security.*

Lee, Y.K., Bell, G., Huang, S.Y., Wang, R.Z. and Shyu, S.J. (2010). An Advanced Least-Significant-Bit Embedding Scheme for Steganographic Encoding. Advances in Image and Video Technology. Berlin / Heidelberg: Springer, 349-360.

Hinkelmann, K. & Kempthorne, O. (2008) Design *and Analysis of Experiments: Introduction to Experimental Design*, John Wiley & Sons, Inc., Hoboken,New Jersey.

Mei Jiansheng, Li Sukang and Tan Xiaomei,(2009) "A Digital Watermarking Algorithm Based on DCT and DWT", *in Proceedings of the 2009 International Symposium on Web Information Systems and Applications (WISA'09) Nanchang*, P. R. China, pp. 104-107.

Mohammad Fahmi, Alalem Abdallah , Muhanah Manasrah,(2008) "*A Steganographic Data Security Algorithm with Reduced Steganalysis Threa*t," Birzeit University, Birzeit.

Park,S.K, and Miller, K.W. (1988). "Random Number Generators: Good Ones Are Hard To Find". Communications of the ACM 31 (10): 1192–1201.

Pfitzmann, B. (1996). 'Information Hiding Terminology', In:Information Hiding*: First International Workshop* (R Anderson, ed), *Lecture Notes in Computer Science* 1174, pp 347-350, Berlin: Springer-Verlag.

Wu, D.C. and Tsai, W.H. (2003). A steganographic method for images by pixel value differencing. *Pattern Recognition Letters*. Vol. 24 (9-10), 1613-1626.

Shuttleworth, M. (2008) Experiment Resources. Accessed: September 25, 2011. [Online] Available: http://www.experiment-resources.com. (July 10, 2012)

Stoica, A., Vertan, C. and Fernandez-Maloigne, C. (2003) Objective and subjective color image quality evaluation for JPEG 2000 compressed images. *International Symposium on Signals, Circuits and Systems, SCS 2003*, 1,137-140.

Wang, Z., Bovik, A. C. and Lu, L. (2002) Why is image quality assessment so difficult? *IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '02)*, 4, 3313-3316.

Wang, Z., Sheikh, H. R. and Bovik, A. C. (2002b) No-reference perceptual quality assessment of JPEG compressed images. *Proceedings of the International Conference on Image Processing*, 1, 477-480.

This academic article was published by The International Institute for Science, Technology and Education (IISTE). The IISTE is a pioneer in the Open Access Publishing service based in the U.S. and Europe. The aim of the institute is Accelerating Global Knowledge Sharing.

More information about the publisher can be found in the IISTE's homepage: http://www.iiste.org

CALL FOR PAPERS

The IISTE is currently hosting more than 30 peer-reviewed academic journals and collaborating with academic institutions around the world. There's no deadline for submission. **Prospective authors of IISTE journals can find the submission instruction on the following page:** http://www.iiste.org/Journals/

The IISTE editorial team promises to the review and publish all the qualified submissions in a **fast** manner. All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Printed version of the journals is also available upon request of readers and authors.

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar