International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 1 | ISSN : 2456-3307

Steganography Method for Hiding Data In The Name_Field of A List of Names Created by Microsoft Word

Amosa Babalola*, Longe Oluwaseun, Onyeka Ndidi and Oluwatosin Adefunke
Department of Computer Science, Federal Polytechnic, Ede, Nigeria

ABSTRACT

This work presents a new algorithm to hide data inside data inside the name_field of a list of names created by Microsoft Word using steganography technique. The algorithm was designed to hide the data inputted within a record of a list of names created by Microsoft word so as to protect the privacy of the data. The system uses three algorithms namely: Algorithm-1 to search for the unused block/spaces in a record of list of names. Algorithm-2: to hide data in the unused blocks/spaces and Algorithm-3: to extract the hidden data. When the algorithms were tested with a few list of names created by Microsoft Word, it was discovered that the records with the hidden data in Microsoft Word do not have a noticeable distortion on it (as seen by the naked eyes). It was also tested using PSNR value, based on these, the PSNR value of each of the list of names created by Microsoft word has a higher PSNR value.

Keywords : Steganography, Unused Blocks, Retrieval, Microsoft Word.

I. INTRODUCTION

The term steganography is the technique of embedding secret information in a communication channel in such a manner that the very existence of the information is concealed. Steganography techniques have been successfully applied to text files, images, audio and video files [1]. Hiding data is the process of embedding information into digital content without causing perceptual degradation [2]. In data hiding, three famous techniques can be used. They are watermarking, steganography and cryptography. Steganography is defined as covering writing in Greek. It includes any process that deals with data or information within other data. According to [3], steganography is hiding the existence of a message by hiding information into various carriers. The major intent is to prevent the detection of hidden information.

Research in steganography technique has been done back in ancient Greek where during that time the ancient Greek practice of tattooing a secret message on the shaved head of a messenger and letting his hair grow back before sending him through enemy territory where the latency of this communications system was measured in months. The most famous method of traditional steganography technique around 440 B.C. is marking the document with invisible secret ink, like the juice of a lemon to hide information. Another method is to mark selected characters within a document by pinholes and to generate a pattern or signature [4]. Data security is the practice of keeping data protected from corruption and unauthorized access.

The focus on data security is to ensure privacy while protecting personal or corporate data. Privacy, on the other hand, is the ability of an individual or group to seclude them or information about themselves and thereby reveal them selectively. Data privacy or information privacy is the relationship between collection and dissemination of data, technology, the public expectation of privacy, and the legal issues.

Data privacy issues can arise from a wide range of sources such as healthcare records, criminal justice investigations and proceedings, financial institutions and transactions, biological traits, residence and geographic records and ethnicity. Data security or data privacy has become increasingly important as more and more systems are connected to the Internet. There are information privacy laws that cover the protection of data or information on private individuals from intentional or unintentional disclosure or misuse. Thus, hiding the data in a kind of form such as within a list of names created by Microsoft word is paramount in order to make sure that security or privacy of the important data is protected. The objective of this work is to design a steganography model for hiding information in Microsoft word document. A more secure Steganographic method capable of surviving Steganalysis attack and that will reduce the probability of detecting the hidden data in Microsoft word is very essential. However in this work, the system uses three algorithms namely: Algorithm-1 to search for the unused block/spaces in a record of list of names. Algorithm-2: to hide data in the unused blocks/spaces and Algorithm-3: to extract the hidden data. The algorithms were tested with a few list of names created by Microsoft word. It was also tested using PSNR value.

II. RELEVANT LITERATURE

This section presents a review on some of the most recent and related work. But, before we proceed, let us describe the basic components of steganography system and basic steganography approaches. A steganography system usually consists of three main components, namely, secret, cover media, and stego media [5]. For a secure steganography system, a forth component is required, which is the key or the password. In general, the secret and the cover media can have the form of text file, audio, video, image or other media file. Figure 1 presents a schematic showing the basic components of a secure steganography system. Steganography has four basic broad approaches.

a. *Lease Significant Bit (LSB) approach*. In this approach, the LSB of each byte of the cover file is replaced with bits from the message.

b. *Injection approach*. In this approach, the source message is hidden in sections of the cover file that are ignored by the processing application. Therefore, avoid modifying those file bits that are relevant to an end-user leaving the cover file perfectly usable.

c. *Substitution approach*. In this approach, the least significant meaningful content of the cover file is replaced with the source message in a way that causes the least amount of distortion to the cover file.

d. *Generation approach*. It is unlike injection and substitution; it does not

require an existing cover file but generates a cover file for the sole purpose of hiding the message.

A large number of steganography algorithms have been developed utilizing the above approaches. In this project, we concern with the first steganography approach, namely, the LSB approach for hiding data in Microsoft word to word document, therefore, in what follow we shall only review some of the most recent researches on Microsoft word to word document steganography.

A steganography algorithm using non-uniform adaptive Microsoft word to word document segmentation with an intelligent computing technique to conceal a large amount of confidential data into color Microsoft word to word document was presented in [6]. Their results demonstrate that the algorithm can efficiently hide a secret data with a capacity of up to four bits per byte while maintaining Microsoft word to word document visual quality. A steganography method that divides the cover Microsoft word to word document into blocks and embeds the corresponding secret data bits into each block without any permutation was presented by [7]. Instead, for a given secret bit sequence, it performs a search on the rows and columns of the layers for finding the most similar row or column. Experimental results showed that the method introduces a very low Microsoft word name-field distortion in comparison to other existing steganography techniques.

A Novel Genetic Algorithm (GA) evolutionary process to secure steganography encoding on text Microsoft word to word document was developed by [8] while [1] introduced two methods wherein cryptography and steganography are combined to encrypt the secret data as well as to hide the encrypted data in another medium. Although such methods can provide higher resistance to steganalysis, it takes long processing time.

A steganography algorithm for hiding text message using the LSB method along with the concept of non-linear dynamic system (chaos) theory was presented in [5]. The algorithm provides security and maintains the secrecy of the secret and provides more randomness. The performance of the algorithm is analyzed and the Peak Signal to Noise Ratio (PSNR) value is also being calculated to indicate the effect of the proposed algorithm on the Microsoft word to word document quality. Also [9] developed a general LSB substitution model called the transforming LSB substitution model to embed secret data in LSBs of pixels in a cover Microsoft word to word document.

III. TYPES OF STEGANOGRAPHY

There are many types of steganography methods identified by [10] are:

A. Text Steganography

Text steganography can be achieved by altering the text formatting, or by altering certain characteristics of textual elements (e.g., characters). The goal in the design of coding methods is to develop alterations that are reliably decodable (even in the presence of noise) yet largely indiscernible to the reader. These criteria, reliable decoding, and minimum visible change are somewhat conflicting; herein lies the challenge in designing document marking techniques. The document format file is a computer file describing the document content

and page layout (or formatting), using standard format description languages such as PostScript2, TeX, @off, etc. It is from this format file that the image - what the reader sees - is generated.

B. Microsoft word Steganography

Hiding information inside Microsoft word to word document is a popular technique nowadays. A Microsoft word to a word document with a secret message inside can easily be spread over the World Wide Web or in newsgroups. The use of steganography in newsgroups has been researched by German steganographic expert Niels Provos, who created a scanning cluster which detects the presence of hidden messages inside Microsoft word to a word document that was posted on the net. However, after checking one million Microsoft word to word document, no hidden messages were found, so the practical use of steganography still seems to be limited.

To hide a message inside a Microsoft word to word document without changing its visible properties, the cover source can be altered in ”noisy” areas with many color variations, so less attention will be drawn to the modifications.

C. Audio Steganography

In audio steganography, the secret message is embedded into digitized audio signal which results in a slight altering of the binary sequence of the corresponding audio file. There are several methods are available for audio steganography. More approaches were presented in [11], [12], [13], [14], [15], and [16].

IV. STEGANOGRAPHY METHOD FOR HIDING DATA IN THE NAME_FIELD OF A LIST OF NAMES CREATED BY MICROSOFT WORD

In this method for hiding data in the name_field of a list of names created by Microsoft Word, three algorithms will be used namely;

a. Algorithm-1 to search for the unused block/spaces in a record of list of names
b. Algorithm-2: to hide data in the unused blocks/spaces
c. Algorithm-3: to extract the hidden data

Algorithm-1 Search for the unused block/spaces
Input: Document of Microsoft Compound Document Binary
File Format (MCDWD).
Output: Unused Block Location.
Steps:
Step1: Loading Compound Document header of MCDWD file.
Step2: Extracting information and offset from header like
(Microsoft signature, Block size, Block index of the first block of the property table (first Directory), byte ordering, Block Allocation Table (BAT) ID, minimum size of a stream).
Step3: Go to the first Directory (Root) Address.
Step4: Extract index of first block in file (starting Block).
Step5: Go to the Block Allocation Table (BAT) Address.
Step6: Loading Block Allocation Table (BAT).
Step7: Accessing from index of the first block in file to all other blocks.
Step8: if Block index = -1
- Calculate the Address of block index in file.
- Record the block as unused block.
Step 9: Else if (Not End of BAT) Go to step7.
Step 10: End.

Algorithm-2: Hiding Data
Input: Document of Microsoft Compound Document File

Format(MCDWD)
Output: Stegodocument
Steps:
Step1: Open MCDWD file.
Step2: Read secret data from user.
Step3: Encode secret data with Huffman Coding.
Step4: Search for unused block in MCDWD file.
Step 5: insert secret data into unused block of MCDWD file.
Step 6: Save the content of document file format.
Step7: End.

Algorithm-3: Extracting hidden data
Input: Stegodocument
Output: Hidden data
Steps:
Step1: Open stegodocument
Step2: Search for unused block in stegodocument
Step3: Extract secret e from unused block of stegodocument.
Step4: Decode secret data.
Step5: End.

V. CONCLUSION

In this work, a steganography method for hiding data in the name_field of a list of names created by Microsoft word has been presented. The work uses a new steganography algorithm with 2 layers of security. It was tested with a few list of names created by Microsoft word. With the algorithm, it was discovered that the records with the hidden data in Microsoft word does not have a noticeable distortion on it (as seen by the naked eyes). It was also tested using PSNR value, based on these, the PSNR value of each of the list of names created by Microsoft word has a higher PSNR value. Hence this new steganography algorithm is very efficient to hide the data inside a list of names created by Microsoft word.

In resume, the use of a steganography model for hiding information in the_field of a list of names created by the Microsoft word **c**annot be overemphasized, considering the various steps that were taken and which are necessary for the development of this system.

VI. REFERENCES

[1]. Shaff.k, sankar, narayanan, prashanth," A Novel Audio Steganography Scheme using Amplitude Differencing ", IEEE Xplore, ISBN: 978-1-4244-9008- 0 vol.10, pp: 163-167, 2011.

[2]. M. Chen, N. Memon, E.K. Wong, Data hiding in document images, in: H. Nemati (Ed.). Premier Reference Source–Information Security and Ethics: Concepts, Methodologies, Tools and Applications, New York: Information Science Reference, 2008, pp. 438-450.

[3]. D.C. Lou, J.L. Liu, H.K. Tso, Evolution of information – hiding technology, in H. Nemati (Ed.), Premier Reference Source–Information Security and Ethics: Concepts, Methodologies, Tools and Applications, New York: Information Science Reference, 2008, pp. 438-450.

[4]. Schneider, Secrets & Lies, Indiana:Wiley Publishing, 2000.

[5]. S. Bhavana and K. L. Sudha. Text Steganography Using LSB Insertion Method Along with Chaos Theory. International Journal of Computer Science, Engineering and Applications (IJCSEA), Vol.2, No.2, pp. 145-149, April 2012.

[6]. Nameer N. El-Emam, Rasheed Abdul Shaheed Al-Zubidy. New Steganography Algorithm to Conceal a Large Amount of Secret Message Using Hybrid Adaptive Neural Networks with Modified Adaptive Genetic Algorithm. Journal

of Systems and Software, Vol. 86, Issue 6, pp. 1465-1481, June 2013.

[7]. O. Kurtuldu and N Arica. A New Steganography Method Using Image Layers. Proceedings of the 23rd International Symposium on Computer and Information Sciences (ISCIS '08), pp. 1-4, Istanbul, Turkey, 27-29 October 2008.

[8]. A. M. Fard, M. M. R. Akbarzadeh-T, and F. Varasteh-A. A New Genetic Algorithm Approach for Secure JPEG Steganography. Proceedings of the IEEE International Conference on Engineering of Intelligent Systems, pp. 1-6, Islamabad, Pakistan, 2006.

[9]. G. R. Xuan, Y. Q. Shi, J. J. Gao, D. Zou, C. Yang, Z. Zhang, P. Chai, C. Chen, and W. Chenl. Steganalysis based on Multiple Features Formed by Statistical Moments of Wavelet Characteristic Functions. Proceedings of the 7th International Information Hiding Workshop, LNCS, Vol. 3727, pp. 262-277, Springer-Verlag, 2010

[10]. Souvik Bhattacharyya, Indradip Banerjee and Gautam Sanyal, Data Hiding Through Multi Level Steganography and SSCE, Journal of Global Sujay Narayana and Gaurav Prasad. Two New Approaches for Secured Image Steganography Using Cryptographic Techniques and Type Conversions. The International Journal of Signal & Image Processing (SIPIJ), Vol.1, No.2, pp. 60-73, December 2011.

[11]. M. H. S. Shahreza, and M. S. Shahreza, "A new approach to Persian/Arabic text steganography," In Proceedings of 5th IEEE/ACIS Int. Conf. on Computer and Information Science and 1st IEEE/ACIS Int. Workshop on Component-Based Software Engineering, Software Architecture and Reuse, 2006, pp. 310-315.

[12]. S. H. Low, N. F. Maxemchuk, J. T. Brassil, and L. O. Gorman, "Document marking and identification using both line and word shifting," INFOCOM'95 Proceedings of the Fourteenth Annual Joint Conf. of the IEEE Computer and Communication Societies, 1995, pp. 853-860.

[13]. J. Cummins, P. Diskin, S. Lau, and R. Parlett, "Steganography and digital watermarking," School of Computer Science, 2004, pp.1-24.

[14]. W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," IBM Systems Journal, vol.35, pp. 313-336, 1996.

[15]. I. Banerjee, S. Bhattacharyya, and G. Sanyal, "Novel text steganography through special code generation," Int. Conf. on Systemics, Cybernetics and Informatics, 2011, pp. 298-303. Information," in Proc. IEEE Information Technology Conf., Syracuse, NY, Sep.1998, pp.113-116.

[16]. T. Y. Liu, and W. H. Tsai, "A new steganographic method for data hiding in Microsoft word documents by a change tracking technique," IEEE Transactions on Information Forensics and Security, vol.2, no.1, pp. 24-30, 2007.