Document Security within Institutions Using Image Steganography Technique

Ahaiwe J.

[1]Department of information Management Technology,
Federal University of Technology, Owerri, Nigeria

Abstract: ***Steganography is the art or practice of concealing a message, image, or file within another message, image, or file. This paper looks at how documents within an institution can be concealed using images within the institutions network. This application developed using C programming language is low cost and stand alone and user friendly. The process of encryption is done on the user's computer before the document is sent over the institutions network to its intended recipient. This process of encrypting and decrypting images makes a hacker or casual user unable to read documents that is not originally intended for such user.***

Keywords: steganography, image, encrypting, decryption, image

1. Introduction

Due to advances in information and communication technology (ICT), most of information is electronically stored and transmitted. Consequently, the security of information has become an underlying concern as well as a threat to all and sundry. The rate of migration to this e-platform is at constant increase. The growing possibilities of modern communications need the special means of security especially on computer network. The network security is becoming more important as the number of data being exchanged on the internet increases. Therefore, the confidentiality and data integrity are required to protect against unauthorized access and use [1]. This has resulted in an explosive growth of the field of information security. The use of image to hide information is one of such security techniques which is known as image steganography. This work looks at how information within institutions can be secured as they seek to go electronic in administration. The steganography technique apart from hiding an information from a user, it disguises the information as an image thus distracting the intruder from his/her original intention.

2. Steganography Vs Cryptography

Various tools and ways have been adopted to protect and secure information especially on the internet. The most commonly use is cryptography mainly due to its simplicity as well as its muddled nature. This method however, is obviously inefficient due to its overt nature of announcing the so-called secured information to the intruders, thereby inviting the intruders to launch attacks on such confidential information. Also, manifold efficacious tools have been set sailed to unveil information locked up using this type of information security tool. To put an end to this unauthorized access of such confidential information, there is a dare need to employ one of the modern information security tools called steganography. According to Bender [2], steganography is a technique of hiding information in digital media. In contrast to cryptography, it is not to keep others from knowing the hidden information but it is to keep others from thinking that the information even exists.

3. Information Hiding

Information hiding is an emerging research area, which encompasses applications such as watermarking, fingerprinting, copyright protection for digital media, and steganography. In watermarking applications for instance, the message contains information such as owner identification and a digital time stamp, which usually applied for copyright protection [2]. "Regarding fingerprint, the owner of the data set embeds a serial number that uniquely identifies the user of the data set. This adds to copyright information to make it possible to trace any unauthorized use of the data set back to the user." [3].

Besides cryptography, steganography can be employed to secure information. In steganography, Moerland [4] has it that, the message or encrypted message is embedded in a digital host before passing it through the network, thus the existence of the message is unknown. Besides hiding data for confidentiality, this approach of information hiding can be extended to copyright protection for digital media: audio, video and images.

Steganography becomes more important as more people join the cyberspace revolution. Silman [5] puts, "steganography is the art of concealing information in ways that prevents the detection of hidden messages. Steganography includes an array of secret communication methods that hide the message from being seen or discovered."

Steganography secreted the confidential message within the host data set and presence subtle and is to be reliably communicated to a receiver. The host data set is purposely corrupted, but in a covert way, designed to be invisible to an information analysis.

4. Aim and Objective Of Research

The goal of steganography is covert communication. So, a fundamental requirement of this steganography system is that the hidden message carried by stego-media should not be sensible to human beings. The other goal of steganography is to avoid drawing suspicion to the existence

of a hidden message. This approach of information hiding technique has recently become important in a number of application areas. This research project has following specific objectives:

1) To provide security tools based on steganography techniques.
2) To explore techniques of hiding data using encryption module of this research.
3) To extract techniques of getting secret data using decryption module.
4) To create program this can be used to conceal vital business information at a low cost.
5) To design a user friendly and distinct security system that enhances information security.

5. Steganography Concepts

Although steganography is an ancient subject, the modern formulation of it is often given in terms of the *prisoner's problem* proposed by Simmons [6], where two inmates wish to communicate in secret to hatch an escape plan. All of their communication passes through a warden who will throw them in solitary confinement should she suspects any covert communication [7].

The warden, who is free to examine all communication exchanged between the inmates, can either be passive or active. A *passive* warden simply examines the communication to try and determine if it potentially contains secret information. If she suspects a communication to contain hidden information, a passive warden takes note of the detected covert communication, reports this to some outside party and lets the message through without blocking it. An *active* warden, on the other hand, will try to alter the communication with the suspected hidden information deliberately, in order to remove the information [8].

6. Applications of Steganography

Steganographic technologies are very important part of the future of Internet security and privacy on open systems such as the Internet. Steganographic research is primarily driven by the lack of strength in the cryptographic systems on their own and the desire to have complete secrecy in an open-systems environment. Many governments have created laws that either limit the strength of cryptosystems or prohibit them completely. This has been done primarily for fear by law enforcement not to be able to gain intelligence by wiretaps, etc. This unfortunately leaves the majority of the Internet community either with relatively weak and a lot of the times breakable encryption algorithms or none at all.

Civil liberties advocates fight this with the argument that "these limitations are an assault on privacy". This is where Steganography comes in. Steganography can be used to hide important data inside another file so that only the parties intended to get the message even knows a secret message exists. To add multiple layers of security and to help subside the "crypto versus law" problems previously mentioned, it is a good practice to use Cryptography and Steganography together. As mentioned earlier, neither Cryptography nor Steganography are considered "turnkey solutions" to open systems privacy, but using both technologies together can provide a very acceptable amount of privacy for anyone connecting to and communicating over these systems.

7. Kinds of Steganographic Systems

Almost all digital file formats can be used for steganography, but the formats that are more suitable are those with a high degree of redundancy. Redundancy can be defined as the bits of an object that provide accuracy far greater than necessary for the object's use and display [9]. The redundant bits of an object according to Anderson & Petitcolas [8] are those bits that can be altered without the alteration being detected easily. Image and audio files especially comply with this requirement, while research has also uncovered other file formats that can be used for information hiding. Four main categories of file formats that can be used for steganography: text, image, audio/video, protocol/network.

7.1. Text Steganography

Steganography can be applied to different types of media including text, audio, image, video, etc. However, text steganography is considered to be the most difficult kind of steganography due to the lack of redundancy in text as compared to image or audio. One method that could be used for text steganography is data compression. Data compression encodes information in one representation, into another representation. The new representation of data is smaller in size. One of the possible schemes to achieve data compression is Huffman coding. Huffman coding assigns smaller length codewords to more frequently occurring source symbols and longer length codewords to less frequently occurring source symbols. Wayner also stated that unicode steganography uses lookalike characters of the usual ASCII set to look normal, while really carrying extra bits of information. If the text is displayed correctly, there should be no visual difference from ordinary text. Some systems, however, may display the fonts differently, and the extra information would be easily spotted [8]

.

Hiding information in text is historically the most important method of steganography. An obvious method was to hide a secret message in every *nth* letter of every word of a text message. It is only since the beginning of the Internet and all the different digital file formats that is has decreased in importance [4]. Text steganography using digital files is not used very often since text files have a very small amount of redundant data.

7.2 Audio/Video Steganography

To hide information in audio files similar techniques are used as for image files. One different technique unique to audio steganography is masking, which exploits the properties of the human ear to hide information unnoticeably. A faint, but audible, sound becomes inaudible in the presence of another louder audible sound [4]. This property creates a channel in which to hide information. Although nearly equal to images in steganographic potential, the larger size of meaningful audio files makes them less popular to use than images [1].

7.3 Protocol/Network Steganography

Ahsan & Kundur [10] puts that, "the term protocol steganography refers to the technique of embedding information within messages and network control protocols used in network transmission. In the layers of the OSI network model there exist covert channels where steganography can be used [11]. An example of where information can be hidden is in the header of a TCP/IP packet in some fields that are either optional or are never used. Typical network steganography methods involve modification of the properties of a single network protocol. Such modification can be applied to the PDU (Protocol Data Unit), to the time relations between the exchanged PDUs, or both (hybrid methods). Moreover, it is feasible to utilize the relation between two or more different network protocols to enable secret communication. These applications fall under the term inter-protocol steganography.

7.4 Image Steganography

Given the proliferation of digital images, especially on the Internet, and given the large amount of redundant bits present in the digital representation of an image, images are the most popular cover objects for steganography. In the domain of digital images many different image file formats exist, most of them for specific applications. For these different image file formats, different steganographic algorithms exist.

Image steganography techniques can be divided into two groups: those in the Image Domain and those in the Transform Domain [12]. Image - also known as *spatial* - domain techniques embed messages in the intensity of the pixels directly, while for transform - also known as *frequency* - domain, images are first transformed and then the message is embedded in the image [13].

8. Steganography Imperceptible property

All steganographic algorithms have to comply with a few basic requirements. The most important requirement is that a steganographic algorithm has to be imperceptible. The authors propose a set of criteria to further define the imperceptibility of an algorithm. These requirements are as follows:

8.1 Invisibility

The invisibility of a steganographic algorithm is the first and foremost requirement, since the strength of steganography lies in its ability to be unnoticed by the human eye. The moment that one can see that an image has been tampered with, the algorithm is compromised

8.2 Payload capacity

Unlike watermarking, which needs to embed only a small amount of copyright information, steganography aims at hidden communication and therefore requires sufficient embedding capacity.

8.3 Robustness against statistical attacks

Statistical steganalysis is the practice of detecting hidden information through applying statistical tests on image data. Many steganographic algorithms leave a 'signature' when embedding information that can be easily detected through statistical analysis. To be able to pass by a warden without being detected, a steganographic algorithm must not leave such a mark in the image as be statistically significant.

8.4 Robustness against image manipulation

In the communication of a stego image by trusted systems, the image may undergo changes by an active warden in an attempt to remove hidden information. Image manipulation, such as cropping or rotating, can be performed on the image before it reaches its destination. Depending on the manner in which the message is embedded, these manipulations may destroy the hidden message. It is preferable for steganographic algorithms to be robust against either malicious or unintentional changes to the image

8.5 Independent of file format

With many different image file formats used on the Internet, it might seem suspicious that only one type of file format is continuously communicated between two parties. The most powerful steganographic algorithms thus possess the ability to embed information in any type of file. This also solves the problem of not always being able to find a suitable image at the right moment, in the right format to use as a cover image.

8.6 Unsuspicious files

This requirement includes all characteristics of a steganographic algorithm that may result in images that are not used normally and may cause suspicion. Abnormal file size, for example, is one property of an image that can result in further investigation of the image by a warden. The following table compares least significant bit (LSB) insertion in BMP and in GIF files, JPEG compression steganography, and the patchwork approach and spread spectrum techniques.

Table 1: Comparison of image Steganography Algorithm

Property	*A*	*B*	*C*	*D*	*E*
Invisibility	H*	M*	H	H	H
Payload capacity	H	M	M	L	M
Robustness against statistical attacks	L	L	M	H	H
Robustness against image manipulation	L	L	M	H	M
Independent of file format	L	L	L	H	H
Unsuspicious files	L	L	H	H	H

* Depends on cover image used
H: High, M: Medium, L: Low
Table Headings
A: LSB in BMP
B: LSB in GIF
C: JPEG compression
D: Patchwork
E: Spread spectrum

The levels at which the algorithms satisfy the requirements are defined as high, medium and low. A high level means

that the algorithm completely satisfies the requirement, while a low level indicates that the algorithm has a weakness in this requirement. A medium level indicates that the requirement depends on outside influences, for example the cover image used. LSB in GIF images has the potential of hiding a large message, but only when the most suitable cover image has been chosen.

9. Data Flow Diagram

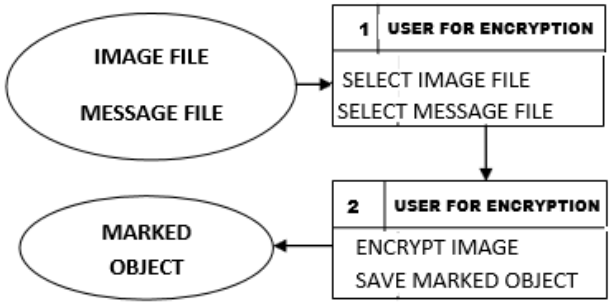

Figure 1: Data flow diagram (encryption)

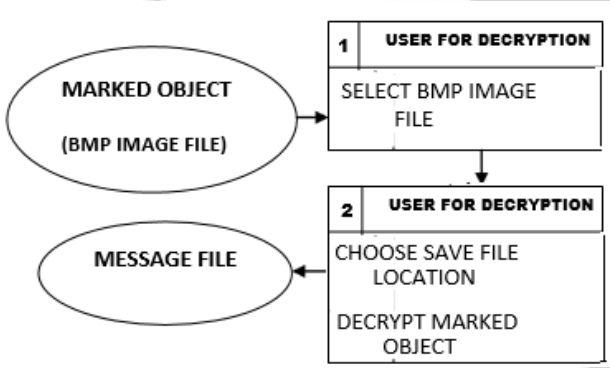

Figure 2: Data flow diagram (decryption)

From the above diagrams, the steganography system is made up of two modules: encryption and decryption. For concealing the information, the encryption module is selected which prompts the user to choose both the cover image and the embedded data. After successful encryption, the information is hidden under the image to produce BMP image file. On recovering the encrypted message, the decryption module is selected. This prompts the user to select the marked object and the location for saving the uncovered message file. The decrypted message file maintains its original file name used during the encryption process

10. Steganography System Images

Steganography system requires any type of image file and the information or message that is to be hidden. It has two modules encrypt and decrypt. Microsoft.Net framework prepares a huge amount of tool and option which helps in simplifying the programming. One of such .Net tools for pictures and images is "auto-converting most types of pictures to BMP format". This tool is used in this software called "Steganography" that is written in C#.Net language. Hence, it is quite easy to use this software to hide in information in any format of pictures without any need of converting its format to BMP. In other words, the software converts the picture on itself.

The basic inputs in the current stegosystem consist of the following:

- Image file and
- Information file

The format specification of the image file is bitmap while the information file can be any type, viz: .doc, .docx, .pdf, .xls etc. The size of the image file determines the size of information file to hide in the image. The formula connecting these two parameters is given below:

$$S = (8.0 * (height * (width / 3) * 3) / 3 - 1)) / 1024$$

Note that: Width = width of image file,
Height = height of image file
S = maximum size of information that can be embedded by the image

11. Algorithm used

The algorithm used for Encryption and Decryption in this application is provided using several layers in place of using only LSB layer of image. Writing data starts from last layer (8st or LSB layer); because significant of this layer is least and every upper layer has doubled significant from its down layer. So every step we go to upper layer image quality decreases and image retouching transpires.

The rationale behind the use of the Least Significant Bit (LSB) algorithm is because insertion is easy and requires simple approach to embedding information in a cover image. In other words, by using LSB algorithm, storing 3 bits in each pixel of the image is made possible. An 800 × 600 pixel image, can thus store a total amount of 1,440,000 bits or 180,000 bytes of embedded data. For example a grid for 3 pixels of a 24-bit image can be as follows:

(00101101 00011100 11011100)
(10100110 11000100 00001100)
(11010010 10101101 01100011)

When the number 200, which binary representation is 11001000, is embedded into the least significant bits of this part of the image, the resulting grid is as follows:

(00101101 00011101 11011100)
(10100110 11000101 00001100)
(11010010 10101100 01100011)

Although the number was embedded into the first 8 bytes of the grid, only the 3 underlined bits needed to be changed according to the embedded message. On average, only half of the bits in an image will need to be modified to hide a secret message using the maximum cover size. Since there are 256 possible intensities of each primary colour, changing the LSB of a pixel results in small changes in the intensity of the colours. These changes cannot be perceived by the human eye - thus the message is successfully hidden. With a well-chosen image, one can even hide the message in the least as well as second to least significant bit and still not see

the difference

12. Justification of Programming Language

C# is a multi-paradigm programming language encompassing strong typing, imperative, declarative, functional, generic, object-oriented (class-based), and component-oriented programming disciplines. It was developed by Microsoft within its .NET initiative and later approved as a standard by Ecma (ECMA-334) and ISO (ISO/IEC 23270:2006). C# is one of the programming languages designed for the Common Language Infrastructure (CLI).

The rationale behind the choice of C# in the design of this system is conspicuously highlighted below:

- C# language is a simple, modern, general-purpose, object-oriented programming language.
- C# attempts to simplify the syntax to be more consistent and more logical in the design of the system.
- It provides software robustness, durability, and makes programmer productivity easy.
- It enhances source code portability, as is programmer portability.
- C# is suitable for writing applications for both hosted and embedded systems, ranging from the very large that use sophisticated operating systems, down to the very small having dedicated functions.
- C# applications are more economical with regard to memory and processing power requirements. C# removes memory management issues from the developer by using .NET's garbage collection scheme. Items no longer referenced are marked for garbage collection, and the Framework can reclaim this memory as needed.

13. The Steganography System Block Diagram and Flow Chart

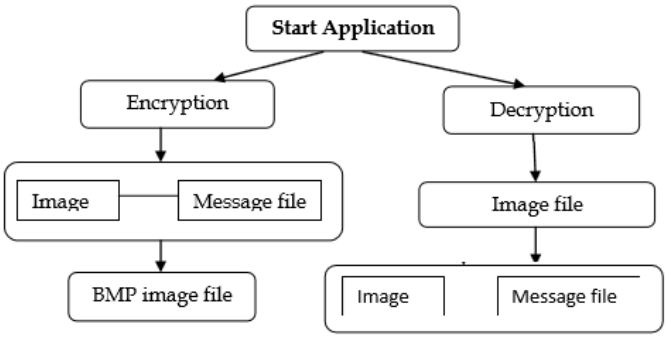

Figure 3: System block diagram

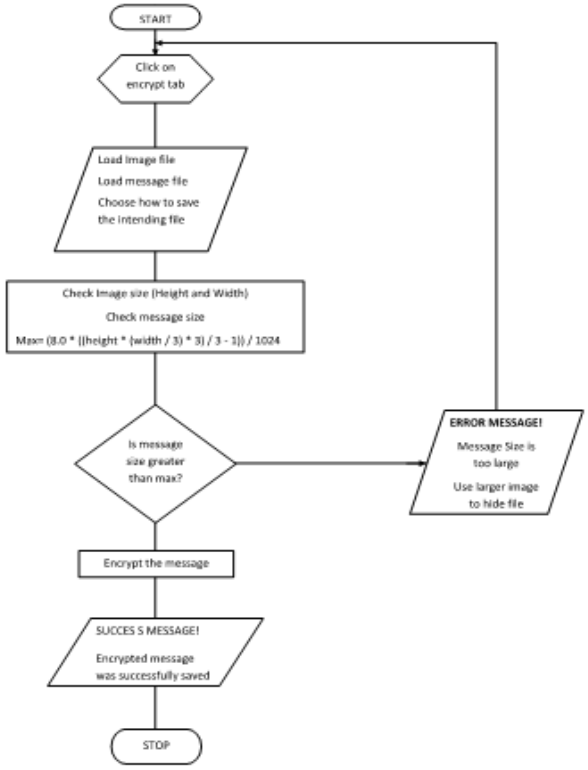

Figure 4: System flow chart.

14. The System Model and Output Screens and Implementation

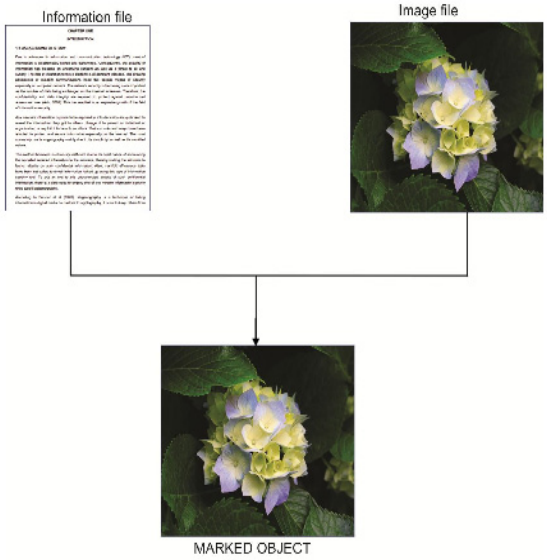

Figure 5: Encryption process model

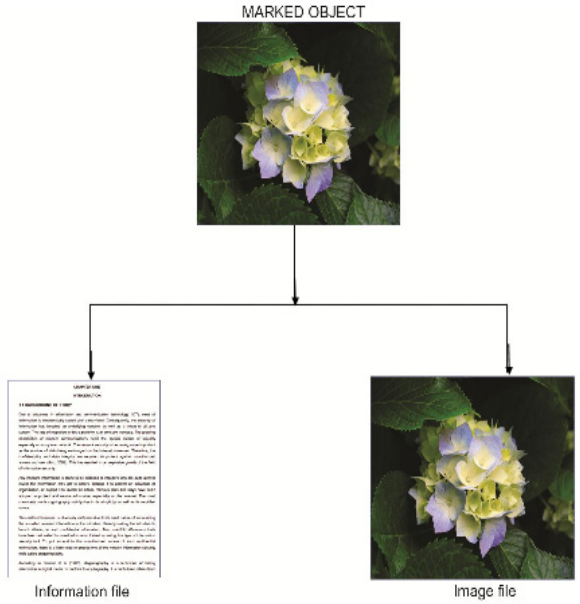

Figure 6: Decryption process model

In figure 6, a document is hidden i.e. encrypted inside an image to form a "marked object", while the marked object in figure 7 is decrypted to generate the original document encrypted. This process describes the steganography process described in the paper.

Figure 7: The steganography system image select screen

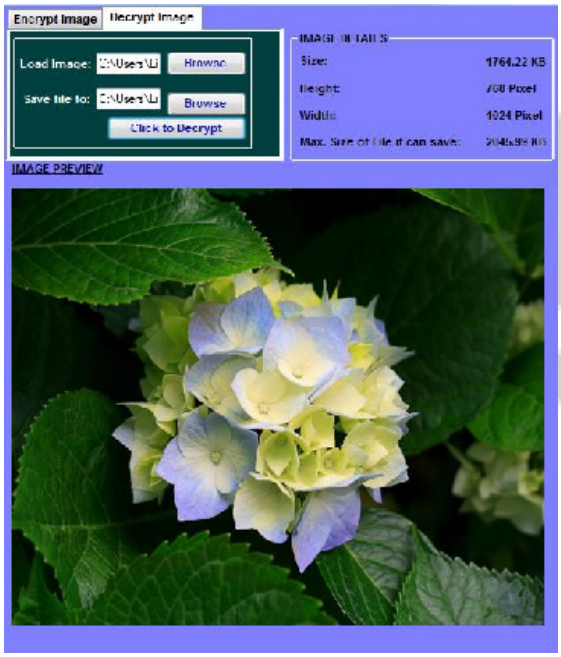

Figure 8: The steganography system after selecting the image and file

The system is made up of two modules: encrypt and decrypt modules. The encrypt module is used to hide information into the image; no one can see that information or file. This module requires any type of image and message and gives the only one image file in destination. The decrypt module is used to get the hidden information in an image file. It takes the image file as an output, and gives two files at destination folder, one is the same image file and the other is the message file that is hidden in it.

Before encrypting file inside image we must save name and size of file in a definite place of image. We could save file name before file information in LSB layer and save file size and file name size in most right-down pixels of image. Writing this information is needed to retrieve file from encrypted image in decryption state. The software which was developed using Microsoft C#.Net 2008 is compiled, packaged, developed and published to make it an executable stand alone program. The software can now be installed and run on other computer systems.

To use the software the user needs to run the application. The user has two tab options -encrypt and decrypt which appear as adjacent tabs at the top left-hand corner of the program. If user select encrypt, application give the screen to select image file, information file and option to save the image file. If user select decrypt, application gives the screen to select only image file and ask path where user want to save the secrete file. For encryption operation, the user first clicks on the encrypt tab to select the module. He then clicks on the image browse button to surf for the location of the image and select the image. At the top right hand corner of the system, the image information (size and height) as well as the maximum size of the message it can hide in Megabytes is well displayed. After this, the user then proceeds to the "information file" browse button to choose the information file to be hidden. Finally, the user clicks on the encrypt button and the system prompts him to save the message by choosing the location and the file name. This is sequentially followed by waiting information by the system to show the processing status of the system. On the other hand, to decrypt an already encrypted message using the system, the user first selects the decrypt tab followed by the image browse button to select the message to be decrypted. At the tail end of the process, the user selects the "decrypt button" and where to save the decrypted message.

15. System Documentation

15.1 Hardware Requirement

This has to do with the basic hardware the system needs to possess for optimum performance and includes system unit with the following configuration:

Table 2: System requirement

Parameter	*Minimum Requirement*
Processor	1GHz
Hard disk	1GB
RAM	512MB
Operating System	Windows xp or latter

15.2 Software Requirement

For effective functioning of this software, there is a software platform required to run on the system. This software tool acts as a platform for the new system to work. Here, I am referring to .NET Framework of at least version 3.5

15.3 Manpower or Operational Requirement

This deals with the skill and personal energy which is a necessary and/or otherwise the prerequisite for the functioning and manipulation of the system. The system however is easy to operate and manipulate even as little or no training is required for its operations.

15.4 Environmental Requirement

This has to do with the overall requirement of the environment in which the system operates. For this system to work well, it needs a dust free environment Air conditional

15.4 System Maintenance

System maintenance ensures that errors which appear during operation of the system are eliminated. It also implements system changes and expansions.

Therefore, this section is dedicated to outline the ways necessary for maintaining the new system, which are:

- Proper handling of the system: This entails starting up (booting) and rebooting the system rightly to prevent file corruption or system "halting".
- Scanning regularly the hard disks and floppy disks used in the system to avoid the invasion of viruses, horses and worms.
- To avoid sudden breakdown of the system, the computer hardware should be serviced regularly.

16. Summary and Recommendation

16.1 Summary

Although only some of the main image steganographic techniques were discussed in this paper, one can see that there exists a large selection of approaches to hiding information in images. All the major image file formats have different methods of hiding messages, with different strong and weak points respectively. Where one technique lacks in payload capacity, the other lacks in robustness. For example, the patchwork approach has a very high level of robustness against most type of attacks, but can hide only a very small amount of information. Least significant bit (LSB) in both BMP and GIF makes up for this, but both approaches result in suspicious files that increase the probability of detection when in the presence of a warden.

Thus for an agent to decide on which steganographic algorithm to use, he would have to decide on the type of application he want to use the algorithm for and if he is willing to compromise on some features to ensure the security of others.

In addition, it has been proved beyond reasonable doubt that, stegosystem has quite a multitudinous applications both in individual transactions and business dealings. It complements the existing system to leave up legal band and to reinforce the present level of security.

16.2 Recommendations

Steganography is seen as a high-level type of encryption; hence, I recommend it to be used in information security within institutions as its use will results in a mechanism to implement two of the five key pillars of information security, namely confidentiality and integrity. Here, the confidentiality of the hidden message is protected due to it being unrecognisable in its hidden and encrypted form both in the place of storage and during transmission while the encrypting of the concealed message protects the integrity of the data.

Besides this, I also recommend more research to be done in the area of stego-key provision to ensure absolute lock of the information being transmitted in this stego-system. Again, applications of steganography are wide ranging, and are indeed valuable if used in the correct manner. Therefore, I would like to recommend the use of this new technology in business and individual transactions to reduce cost and enhance the net revenue of such business or individual as the case may be. Since neither Cryptography nor Steganography are considered "turnkey solutions" to open systems privacy, I recommend the use of both technologies together to provide a very acceptable amount of privacy for anyone connecting to and communicating over these systems. Finally, since this security technology may be misused and abused, resulting in disastrous consequences, I recommend the use of official instrument to control the transmission of embedded information through Steganography to check the use of such tool in concocting criminal and inhumane plots in the society.

References

[1] Artz, D. (2001). Digital Steganography: Hiding Data within Data. *IEEE Internet Computing Journal,* (June).
[2] Bender, W., Gruhl, D., Morimoto, N. & Lu, A., (1996). Techniques for data hiding. *IBM Systems Journal,* 35(2).
[3] Dunbar, B. (2002). Steganographic techniques and their use in an Open-Systems environment. SANS Institute, January.
[4] Moerland, T.(2001). Steganography and Steganalysis. *Leiden Institute of Advanced Computing Science.* Accessed September 12, 2012. Available from www.liacs.nl/home/ tmoerl/privtech.pdf
[5] Silman, J.,(2001). Steganography and Steganalysis: An Overview. *SANS Institute.*
[6] Simmons, G., (1983). The prisoners problem and the subliminal channel. *CRYPTO.*
[7] Chandramouli, R., Kharrazi, M. & Memon, N.(2003). Image steganography and steganalysis: Concepts and Practice. *Proceedings of the 2nd International Workshop on Digital Watermarking,* October.

[8] Anderson, R.J. & Petitcolas, F.A.(1998). On the limits of steganography. *IEEE Journal of selected Areas in Communications,* (May): 22.
[9] Currie, D.L. & Irvine, C.E., (1996). Surmounting the effects of lossy compression on Steganography. *19th National Information Systems Security Conference*.
[10] Ahsan, K. & Kundur, D.(2002). Practical Data Hiding in TCP/IP. *Proceedings of the Workshop on Multimedia Security at ACM Multimedia.*
[11] Handel, T. & Sandford, M.(1996). Hiding data in the OSI network model. *Proceedings of the 1st International Workshop on Information Hiding,* June.
[12] Silman, J.,(2001). Steganography and Steganalysis: An Overview. *SANS Institute.*
[13] Lee, Y.K. & Chen, L.H. (2000). High capacity image steganographic model. *Visual Image Signal Processing,* 147(03), June.

Author Profile

Dr. Ahaiwe J., received the B.S. in Business with a minor in mathematics at the Agriculture and Mechanical university Alabama between 1977 and 1980. He has a Master of Science degree in Computer Science Technology in the same university between 1980 and 1983. Dr Ahaiwe J. got a PHD degree in Project Management Technology at the Federal University of Technology, Owerri where is currently a Senior Lecturer.