Implementation of Digital Steganography Using Image Files-A Computational Approach

N.Rajender
ASSISTANT PROFESSOR, DEPT OF CSE, NALLA NARASIMHA REDDY EDUCATIONAL SOCIETY'S GROUP OF INSTITUTIONS, Hyderabad, Telangana, INDIA

P.Ajay Kumar
Lecturer, Dept of CSE, JNTU-CEH,Hyderabad,Telengana,India

V.Raju
Assistant Professor, Dept of CSE, NALLA NARASIMHA REDDY EDUCATIONAL SOCIETY'S GROUP OF INSTITUTIONS, Hyderabad, Telangana, INDIA

S.Rajesh
Assistant Professor, Dept of CSE, NALLA NARASIMHA REDDY EDUCATIONAL SOCIETY'S GROUP OF INSTITUTIONS, Hyderabad, Telangana, INDIA

Abstract— The main aim of this paper is to hide the message along with the existence of communication so that the attacker can't even get the encrypted message. The procedure follows as hiding the encrypted message in image files by embedding the message bits into least significant bit positions of each byte of image file. In image file each pixel is represented with 24 bits which is a combination of RED, BLUE, GREEN color proportions for that pixel.LSB has least importance in image files so it's modification can't results in a human detectable change. Here Data Encryption Standard (DES) algorithm is used to encrypt the message and MD5 algorithm is used to compute the message digest which is used to check the integrity of message. After performing Cryptographic functions message is compressed with GZip compression technique and then it is embedded into image file. This process results in a stego file (output image file) which is exactly similar to cover file (input image file).

Keywords- Encryption, Decryption, DES,MD5,LSB

I. INTRODUCTION

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 amount of data being exchanged on the Internet is increasing. Security requirements are necessary both at the final user level and at the enterprise level, especially since the massive utilization of personal computers, networks, and the Internet with its global availability. Throughout time computational security needs have been focused on different features: secrecy or confidentiality, identification, verification, non repudiation, integrity control and availability. This has resulted in an explosive growth of publishing and broadcasting technology also requires an alternative solution in hiding information. The copyright of digital media such as audio, video and other media available in the digital form may lead to large scale unauthorized copying. The problem of unauthorized copying is of great concern especially to the music, film, book and software publishing industries. To overcome this problem, some invisible information can be embedded in the digital media in such a way that it could not be easily extracted without a specialized technique[1-25].

There are a number of ways for securing data. One is cryptography, where the sender uses an encryption key to scramble the message, this scrambled message is transmitted through the insecure public channel, and the reconstruction of the original, unencrypted message is possible only if the receiver has the appropriate decryption key. The second method is image Steganography, where the secret message is embedded in an image, thus the existence of message is unknown[1-10].

Steganography derived from Greek word literally means covered writing. It includes vast array of secret communication method that conceals message's very existence. Computer based Steganography allows changes to be made to what are known as digital carriers such as images or sounds. Digital images, videos, sound files, and other computer files that contain perceptually irrelevant or redundant information can be used as covers or cover-image, a so-called Stego image is obtained. The basic model of image Steganography consists of cover file,

Message, Embedding algorithm and the Message file. The system deals with security during transmission of data. Commonly used technologies are cryptography. The system deals with implementing security using steganography. In this the end user identifies an image which is going to act as the carrier of data. The data file is also selected and then to achieve greater speed of transmission the data file and the image file are compressed and sent. Prior to this the data is embedded into the image and then sent. The image if hacked or interpreted by a third party user will open up in any image previewed but not displaying the data. This protects the data from being invisible and hence be secure during transmission. The user in the receiving end uses another piece of code to retrieve the data from the image[1-25].

II. BACKGROUND

A. significant terms of steganography

In general, terminology analogous to (and consistent with) more conventional radio and communications technology is used; however, a brief description of some terms which show up in software specifically, and are easily confused, is appropriate. These are most relevant to digital steganography systems[1-25].

The payload is the data to be covertly communicated. The carrier is the signal, stream, or data file into which the payload is hidden; which differs from the "channel" (typically used to refer to the type of input, such as "a JPEG image"). The resulting signal, stream, or data file which has the payload encoded into it is sometimes referred to as the package, stego file, or covert message. The percentage of bytes, samples, or other signal elements which are modified to encode the payload is referred to as the encoding density and is typically expressed as a number between 0 and 1[1-25].

B. Cover and Stego files

Cover-object embeds the message and serves to hide its presence. The suitable carriers that can be used as cover-object are network protocols such as TCP, IP and UDP. Image files such as bmp, gif and jpg, where they can be both color and grayscale. Message is the data that the sender wishes to remain it confidential. It can be plain text, cipher text, other image, or anything that can be embedded in a hit stream such as a copyright mark, a covert communication, or a serial number. Password is known as a stegokey which ensures that only the recipient who knows the message from a cover-object. The cover-object with the secretly embedded message is then called the stego-object. This stego-object is then transferred to the end, there we have detector algorithm which extract the message from cover object[1-25].

C. Digital Steganography

Development following that was slow, but has since taken off, going by the number of "stego" programs available: Over 800 digital steganography applications have been identified by the Steganography Analysis and Research Center. Digital steganography techniques include[1-25]:

- Concealing messages within the lowest bits of noisy images or sound files.
- Concealing data within encrypted data or within random data. The data to be concealed is first encrypted before being used to overwrite part of a much larger block of encrypted data or a block of random data (an unbreakable cipher like the one-time pad generates ciphertexts that look perfectly random if you don't have the private key).

D. LSB insertion method

The least significant bit insertion method is probably the most well known image Steganography technique. It is a common, simple approach to embed information in a graphical image file. This is the most common method used in this data to be hidden is inserted into the least significant bits of the pixel information. In digital, images are represented with the numerical values of each pixel where the value represents the color and intensity of the pixel. In 24 bit image we can embed 3 bits in each pixel while in 8-bit we can embed only 1 bit in each pixel. To hide an image in the LSBs of each byte of the 24-bit image, one can store 3 bits in each pixel. A 1024 X 768 image has the potential to hide a total of 2,359,296 bits of information[1-25].

E. Compression method

Compression, this technique maps arbitrary input into printable character output. The form of encoding has the following relevant characteristics. The range of the function is a character set that is universally re-presentable at all sites, not a specific binary encoding of that character set. Thus, the characters themselves can be encoded into whatever form is needed by a specific system. For instance, the character ‘E’ is represented in ASCII system as a hexadecimal 45 and in EDCDIC-based system as hexadecimal C5[1-25].

The character set consists of 65 printable characters, one of which is used for padding. With 2^6=64 available characters, each character can be used to represent 6 bits of input. No control characters are included in the set. Thus, the message encoded in Radix-64 can traverse mail-handling system. That scans the data stream for control characters. The hyphen character "-"is not included[1-25].

The system deals with security during transmission of data. Commonly used technologies are cryptography. The system deals with implementing security using steganography. In this the end user identifies an image which is going to act as the carrier of data. The data file is also selected and then to achieve greater speed of transmission the data file and the image file are compressed and sent. Prior to this the data is embedded into the image and then sent. The image if hacked or interpreted by a third party user will open up in any image previewed but not display ing the data. This protects the data from being invisible and hence be secure during transmission. The user in the receiving end uses another piece of code to retrieve the data from the image[1-25].

F. Gzip compression method

Gzip is any of several software applications used for file compression and decompression. The term usually refers to the GNU Project's implementation, "gzip" standing for GNU zip. It is based on the DEFLATE algorithm, which is a combination of Lempel-Ziv (LZ77) and Huffman coding. Gzip is based on the DEFLATE algorithm, which is a combination of LZ77 and Huffman coding. DEFLATE was intended as a replacement for other patent-encumbered data compression algorithms, which, at the time, limited the usability of compress and other popular archive's. "Gzip" is often also used to refer to the gzip file format, which is:

- a 10-byte header, containing a magic number, a version number and a time stamp
- optional extra headers, such as the original file name,
- a body, containing a DEFLATE-compressed payload
- an 8-byte footer, containing a CRC-32 checksum and the length of the original uncompressed data

Although its file format also allows for multiple such streams to be concatenated (zipped files are simply decompressed concatenated as if they were originally one file), gzip is normally used to compress just single file.

III. CURRENT SYSTEM ANALYSIS AND PROPOSED SYSTEM

A. Current System

In today's dynamic and information rich environment, information systems have become vital for any organization to survive. With the increase in the dependence of the organization on the information system, there exists an opportunity for the competitive organizations and disruptive forces to gain access to other organizations information system. This hostile environment makes information systems security issues critical to an organization. Current information security literature either focuses on anecdotal information by describing the information security s\attacks taking place in the world or it comprises of the technical literature describing the types of security threats and the possible security systems. In order to secure the transmission of data, Cryptography has to be implemented. Cryptography is the science of devising methods that allow information to sent in a secure form in such a way that the only person able to retrieve this information is intended recipient.

The basic principle is this: A message being sent is known as plain text . The message is then coded using a cryptographic algorithm. This process is called encryption, an encrypted message is known as cipher text, and is turned back into plain text by the process of decryption. Encryption can be done using symmetric or asymmetric algorithms. In symmetric algorithms only one key is used both to encrypt a message is different from the key that is used to decrypt the message. There are several algorithms present in the market for Cryptography. Some of the commonly used once are IDEA, RSA and BLOWFISH that involves asymmetric or symmetric methods and also involves private and public keys.

B. Proposed System

In this proposed system we have the software for data encryption and then embed the cipher text in an image with help of stego key. This algorithm combines the effect of these two methods to enhance the security of the data.

The proposed algorithm encrypts the data with a crypto algorithm and then embeds the encrypted text in an image file. This algorithm improves the security of the data by embedding the encrypted text and not the plain text in an image. It also calculates message digest of cover file to check integrity of message. This system will satisfy four requirements which we must require for secure data transmission. These are Confidentiality, or Message Content Security, Integrity of Message Content, Authentication, Security in an open system.

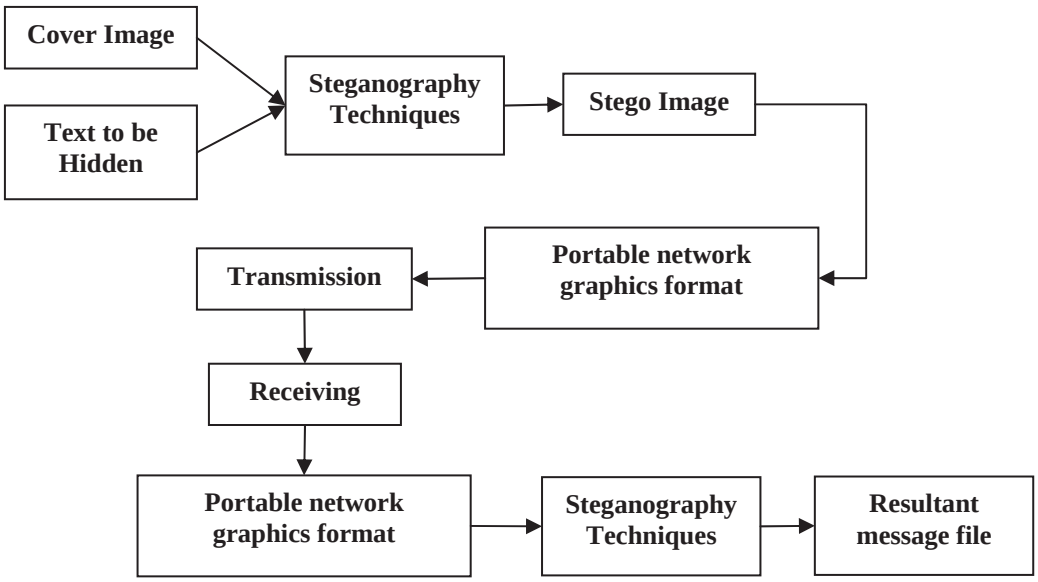

Fig 1 Steganography System Architecture

C. Processing

The end user identifies an image which is going to act as the carrier of data. The data file is also selected and then to achieve greater speed of transmission the data file and the image file are compressed and sent. Prior to this the data is embedded into the image and then sent. The image if hacked or interpreted by a third party user will open up in any image previewed but not displaying the data. This protects the data from being visible and hence be secure during transmission. The user in the receiving end uses another piece of code to retrieve the data from the image. The two primitive operations to be performed on that message file are encryption and compression, this process is shown in Fig 1.

Password-based-encryption technique ensures that the symmetric keys are protected from unauthorized access. This is a password based technique for encrypting a session key. In this we first encrypt the plain text message with the symmetric key and then encrypt the symmetric key with a Key Encryption Key (KEK).To prevent attacks such as dictionary attack, apart from password, two additional pieces of information are used in the key generation mechanism. They are: salt and iteration count. Salt is simply a bit string, which is combined with the password to produce the KEK. Iteration count specifies the number of operations that must be performed on the combination of the password and the salt to generate the KEK.

Data compression is the process of encoding information using fewer bits (or other information-bearing units) than an un-encoded representation would use through use of specific encoding schemes. As with any communication, compressed data communication only works when both the sender and receiver of the information understand the encoding scheme.

D. Algorithm Explanation:.The explanation of the algorithm is shown in Table I

TABLE –I: THE EXPLANATION OF DIFFERENT ALGORITHMS

Function	Algorithm	Description
Compression	GZip	(GNU zip)This is used to compress the data or the message file before it is inserted into the image.
Key generation	MD5	(Message Digest 5) The message digest of the text is obtained and is added at the end of the actual message.
Encryption	DES	(Data Encryption Standard) The password entered is the key and the message file is the plain text.DES generates the corresponding cipher text.
Embedding	LSB	(Least Significant Bit insertion technique) The plain/cipher text is embedded into the image file by using the LSB insertion technique. With this 3 bits per pixel can store our data.

E. Experimental Results

This experiment takes text, image, audio, video files as input message files and image files such as JPEG, BMP, GIF, JPG, PNG, TIFF etc as Cover files and confirm on a password then automatically the output stego file will be generated which looks just like the cover file and attacker can't even suspect that there is a message in the image file. We take the advantage of redundant bits in image files and modify them according to our needs. There may be very small changes in color of pixels but the human eye can't identify that change.

The resultant stego file will be uploaded to the intended party and the user on the other side will download it and use that stego file as input for the process of extraction and provide the correct password then the actual message file will be extracted. The idea of hiding the message along with actual communication is fulfilled, thus providing more security to the message.

IV. CONCLUSION AND FUTUREWORK

Steganography is used to hide information in innocuous looking files. This is in contrary to Cryptography which transforms the information into incomprehensible cipher text. In our system we have implemented Steganography in images where the sensitive information is hidden in a harmless image file. Our system also allows us to place an image with in another image. To eliminate the overhead of key exchange, we used encryption techniques which use something derived from passwords i.e., Message Digests to encrypt the information whenever encryption is requested for. And since the message digest is usually quite long, the Key Size is also high. So even if the third person manages to extract the information, he can't crack the key. Thus it will be useless to intercept the file. This project also allows us to use compression techniques which compress the data before we embed them into an image. This is extremely useful when we need to embed large amount of information into a single image file. For the purpose of Steganography we only use lossless compression techniques. This project also allows us to opt for noisy image as a carrier for our sensitive information. Finally, we use Least Significant Bit technology to insert data into images. At the receiving end only the password is necessary to obtain the information embedded in the images.

This system can be extended to support Steganography in audio and text files as well. These additions can be introduced into our current project in the form of plug-ins. Instead of using the plain Least Significant Bit (LSB) approach for inserting bits into an image, random bit locations can be chosen for insertion of data bits. With this Steganography becomes even hard to detect or attack. Research is currently being carried in this direction. Compression modules which actually compress the Stego-file can also be designed. This can greatly reduce the bandwidth consumed for transmission of the Stego-file and greatly reduce the time associated with it.

REFERENCES

[1] Cryptography and Network Security, Atul Kahate, TMH second edition.
[2] Applied Cryprtography, Bruce Schneider,John Wiley and Sons, 2001.
[3] A Review of Data Hiding in Digital Images, a paper by E.T Lin and E.J Delp 2002.
[4] Cryptography and Network Security, William Stallings, Prentice Hall India.
[5] Study of Steganography and The Art of Hiding of information, a paper by Alain C.Brainos II East Carolina /University, November 2003.
[6] www.wikipedia.com, Cryptography, the free encyclopedia.
[7] http://www.garykessler.net/library/steganography.htm, A few basics for Steganography.
[8] http://www.securityfocus.com/infocus/168, Steganography Revealed.
[9] Steganography: How to Send a Secret Message By Bryan Clair8 October 2001
[10] Chin-Chen Chang, Chia-Chen Lin, Yih-Shin Hu "An Svd Oriented Watermark Embedding Scheme With High Qualities For The Restored Images" International Journal of Innovative Computing, Information and Control ICIC , 2007
[11] Allan M. Bruce, "A Review of Digital Watermarking", Department of Engineering, University of Aberdeen, 2001
[12] Saraju P. Mohanty., "Digital Watermarking: A Tutorial Review", Department of Computer Science and Engineering, University of South Florida,1999.
[13] Maurice macs, Ton kalker, Jean-paul M.G. Linnartz, Joop Talstra, Geert F.G. Depoveve, Jaap Haistsma, "Digital watermarking for DVD Video copy protection" IEEE Signal Processing magazine, pp.1053-5888, 2000.
[14] Fabien A P P., "Watermarking Schemes Evaluation", IEEE Signal Processing Magazine, 2000.
[15] R.G. van Schyndel, A.Z.Tirkel, N.R.A.Mee, C.F.Osborne. A Digital Watermark. First IEEE Image Processing Conference, Houston TX, pp.86-90, 1994.
[16] A.Z.Tirkel, G.A.Rankin, R.M.van Schyndel, W.J.Ho, N.R.A.Mee, C.F.Osborne. Electronic Water Mark. DICTA-93 Macquarie University, Sydney, pp.666-672, 1993..
[17] Huang, C.-S.C.P.S., Chang, C.-P. and Tu, T.-M., "Robust spatial watermarking technique for colour images via direct saturation adjustment". IEE Proceedings Vision, Image & Signal Processing, pp.561-574, 2005.
[18] Xinzhong Zhu Jianmin Zhao Huiying Xu , "A Digital Watermarking Algorithm and Implementation Based on Improved SVD". International conference on pattern recognition, pp- 651 – 656, 2006.
[19] Liu Liang, Sun Qi, "A new SVD-DWT composite watermarking", Proceedings of 8th International conference on Signal Processing, pp.16-20, 2006.
[20] D Sanchez , "Robust New Method in Frequency Domain Watermarking" , IEEE Transactions on Image Processing, 2001.