# CRYPTOGRAPHY AND STEGANOGRAPHY-A SURVEY

**N. D. Jambhekar and C.A. Dhawale***
Department of Computer Science
S.S.S.K.R. Innani Mahavidyalaya, Karanja Lad Dist Washim (M.S.), India
*Department of Computer Science
P.R. Pote Group of Institutes, Amravati (M.S.), India

## ABSTRACT

*With the advanced technology used today for data exchange in electronic way over the communication network, the information security is becoming more important in data storage and transmission. Data always plays the crucial role for every organization and must be protected from the unauthorized access. In this paper we analyze some data protection techniques based on the Cryptographic and Steganographic algorithms.*

*Keywords — encryption, decryption, cryptography, steganography.*

## Introduction

With the advent technology used for the data communication, importance must be given to the information security. The data communication over the devices which are mobile and uses the wireless communication technology, the data must be encrypted before transmission and at the receiver end, it get decrypted. There are different encryption algorithms available in the Cryptographic and Steganographic approach for the successful data protection from malicious access or theft. To protect the document from the illegal access, media (i.e. information) owner use these technique to embed the desired signal into the original media for secured communication process.

**Cryptography** is the science of writing secret message in a such a way that any one other than the legal user cannot understand the original message. In data and telecommunications, cryptography is necessary when communicating over any untrusted medium, which includes just about any network, particularly the Internet. Cryptography, then, not only protects data from theft or alteration, but can also be used for user authentication.

There are, in general, three types of cryptographic schemes typically used to accomplish these goals: secret key (or symmetric) cryptography, public-key (or asymmetric) cryptography, and hash functions, each of which is described below. In all cases, the initial unencrypted data is referred to as plaintext. It is encrypted into ciphertext, which will in turn (usually) be decrypted into usable plaintext. The cryptographic algorithms is classified as

- Secret Key Cryptography- Uses a single key for both encryption and decryption
- Public Key Cryptography- Uses one key for encryption and another for decryption
- Hash Functions- Uses a mathematical transformation to irreversibly "encrypt" information

**Steganography** is the art and science of invisible communication [C. Cachin, 1998]. This is accomplished through hiding information in other information, thus hiding the existence of the communicated information. The word steganography is derived from the Greek words “stegos” meaning “cover” and “grafia” meaning “writing”, defining it as “covered writing”. In image steganography the information is hidden exclusively in images.

## Overview of Cryptographic algorithms

There are of course a wide range of cryptographic algorithms in use. The following are amongst the most well known.

**Secret Key Cryptography**- Uses a single key for both encryption and decryption. The algorithms included in this are Data Encryption Standard (DES), Advanced Encryption Standard (AES) the DES is the 'Data Encryption Standard'. This is a cipher that operates on 64-bit blocks of data, using a 56-bit key. It is a 'private key' system. The AES i.e. Advanced Encryption Standard cipher is specified as a number of repetitions of transformation rounds that convert the input plaintext into the final output of ciphertext. Each round consists of several processing steps, including one that depends on the encryption key. A set of reverse rounds are applied to transform ciphertext back into the original plaintext using the same encryption key.

KeyExpansion—round keys are derived from the cipher key using Rijndael's key schedule

- Initial Round
  1. AddRoundKey—each byte of the state is combined with the round key using bitwise xor
- Rounds
  1. SubBytes—a non-linear substitution step where each byte is replaced with another according to a lookup table.
  2. ShiftRows—a transposition step where each row of the state is shifted cyclically a certain number of steps.
  3. MixColumns—a mixing operation which operates on the columns of the state, combining the four bytes in each column.
  4. AddRoundKey
- Final Round (no MixColumns)
  1. SubBytes - In this step, each byte in the matrix is updated using an 8-bitsubstitution box, the Rijndael S-box.
  2. ShiftRows - This step operates on the rows of the state; it cyclically shifts the bytes in each row by a certain offset.
  3. AddRoundKey - In this step, the four bytes of each column of the state are combined using an invertible linear transformation.

**Public Key Cryptography**- Uses one key for encryption and another for decryption. Public-key cryptography algorithms that are in use today for key exchange or digital signatures include-

**RSA** - is a public-key system designed by Rivest, Shamir, and Adleman. RSA uses a variable size encryption block and a variable size key. The key-pair is derived from a very large number, n, that is the product of two prime numbers chosen according to special rules; these primes may be 100 or more digits in length each, yielding an n with roughly twice as many digits as the prime factors [Jonsson J. and Kaliski B, 2002].

**Hash Functions**- Uses a mathematical transformation to irreversibly "encrypt" information. A 'hash algorithm' is used for computing a condensed representation of a fixed length message/file. This is sometimes known as a 'message digest', or a 'fingerprint'. MD5- is a 128 bit message digest function. It was developed by Ron Rivest. SHA-1 -is a hashing algorithm similar in structure to MD5, but producing a digest of 160 bits (20 bytes). Because of the large digest size, it is less likely that two different messages will have the same SHA-1 message digest. For this reason SHA-1 is recommended in preference to MD5. HMAC - is a hashing method that uses a key in conjunction with an algorithm such as MD5 or SHA-1.

## Overview of Steganographic algorithms

Almost all digital file formats like text, images, audio and video can be used for steganography. Hiding information in text requires the merging of some characters depending on certain logic criterion [Artz D,

2001]. Images are the most popular cover objects used for steganography [Johnson N.F. and Jajodia S, 1998]. Any type of text or image can be covered by another key image and successful extraction of original information is possible. Hiding information in audio files requires the masking technique [Bender et. al, 1996]. A faint, but audible, sound becomes inaudible in the presence of another louder audible sound.

The steganographic methods are categories in some categories such as-

**Image Domain** - embed secret information using changing the intensity of LSB.

**Transform Domain Technique**- embed secret information in a transform space of the signal

**Spread Spectrum Technique** – hidden data is spread throughout the cover-image making it harder to detect.

**Statistical Method** - encode information by changing several statistical properties of a cover and use hypothesis testing in the extraction process.

**Cover Generation methods**- encode information in the way a cover for secret communication is created.

**Image domain**

Under the image domain, Least significant bit (LSB) insertion is a common [Moulin P and Koetter R, 2005], simple approach to embedding information in a cover file. But, it is vulnerable to even a slight image manipulation. But doing such type of embedding, the image loss can be occur, where it may major of minor depends on the images used [Chandramouli, 2003].

**24-bit images**. To hide an image in the LSBs of each byte of a 24-bit image, you can store 3 bits in each pixel. A 1,024 × 768 image has the potential to hide a total of 2,359,296 bits (294,912 bytes) of information. If you compress the message to be hidden before you embed it, you can hide a large amount of information. To the human eye, the resulting stego-image will look identical to the cover image.

For example, the letter A can be hidden in three pixels (assuming no compression). The original raster data for 3 pixels (9 bytes) may be

(00100111 11101001 11001000)
(00100111 11001000 11101001)
(11001000 00100111 11101001)

The binary value for A is 10000011. Inserting the binary value for A in the three pixels would result in

(00100111 11101000 11001000)
(00100110 11001000 11101000)
(11001000 00100111 11101001)

The underlined bits are the only three actually changed in the 8 bytes used. On average, LSB requires that only half the bits in an image be changed. You can hide data in the least and second least significant bits and still the human eye would not be able to discern it. This same technique is used for the 8 bit images.

**Transform Domain**

The LSB modifications techniques are easy to embed secrete information, but they are highly vulnerable to even small cover modifications. An attacker can simply apply signal processing techniques I order to destroy the secret information entirely. It is clear from the recent practice that embedding information in the frequency domain of a signal can be much more robust than embedding rules operating in the time domain [K B Shiva Kumar et al, 2011]. It is highly robust when using the transform domain technique, while for restriction on the potential attach such as compression, cropping, and image processing techniques than the LSB approach.

One popular method of encoding secret information in the frequency domain is modulation the relative size of two or more DCT coefficients within one image block. During encoding process, the sender splits the cover-image in 8 X 8 pixel blocks; each block encodes exactly one secret message bit. The embedding process starts with selecting a pseudorandom block bi which will be used to code the ith message bit. Both sender and receiver must agree on the location of two DCT coefficients, used in the embedding process.

**Patchwork**

Patchwork is a statistical technique that uses redundant pattern encoding to embed a message in an image [Johnson N.F. and Jajodia S, 1998]. The algorithm adds redundancy to the hidden information and then scatters it throughout the image [Bender W, 1996]. A pseudorandom generator is used to select two areas of the image (or patches), patch A and patch B. All the pixels in patch A is lightened while the pixels in patch B is darkened. The advantage of using this technique is that the secret message is distributed over the entire image, so should one patch be destroyed, the others may still survive [Bender W, 1996]. This however, depends on the message size, since the message can only be repeated throughout the image if it is small enough. If the message is too big, it can only be embedded once.

**Spread Spectrum**

In spread spectrum techniques, hidden data is spread throughout the cover-image making it harder to detect [Marvel et. al, 1999]. A system proposed by Marvel et al. combines spread spectrum communication, error control coding and image processing to hide information in images.

There are two approaches are used in this technique: the direct sequence spread spectrum (DSSS) and frequency hopping spread spectrum (FHSS). Direct-sequence spread spectrum (DSSS) is a modulation technique used in telecommunication. As with other spread spectrum technologies, the transmitted signal takes up more bandwidth than the information signal that is being modulated. Direct-sequence spread-spectrum transmissions multiply the data being transmitted by a "noise" signal. This noise signal is a pseudorandom sequence of 1 and −1 values, at a frequency much higher than that of the original signal, thereby spreading the energy of the original signal into a much wider band. The resulting signal resembles white noise. In contrast, frequency-hopping spread spectrum pseudo-randomly retunes the carrier, instead of adding pseudo-random noise to the data, which results in a uniform frequency distribution whose width is determined by the output range of the pseudo-random number generator.

**Statistical Method**

Statistical steganography techniques utilize the existence of "1-bit" steganographic schemes, which embed one bit of information in a digital carrier. This is done by modifying the cover in such a way that some statistical characteristics change significantly if a "1" is transmitted [Provos N, 2001][Solanki K. et. al, 2005].

Assuming m is the secret message and l(m) is the length of the message in bits. A cover is divided into l(m) disjoint blocks $B_1,\ldots\ldots,B_{l(m)}$ . A secret bit $m_i$ is inserted into the i th block by placing a "1" into $B_i$ if $m_i = 1$ , otherwise the block is left unchanged. The detection of a specific bit is done via a test function that distinguishes between modified and unmodified blocks.

$f(B_i) = 1$ if block $(B_i)$ was modified and = 0 if it was not

The receiver successively applies f to all cover blocks to restore the secret message.

**Cover Generation methods** do not embed messages in randomly chosen cover-objects, but create covers that fit a message that need to be hidden [ Wayner, Peter, 1992]. In contrast to systems where secret information is added to a specific cover by applying an embedding algorithm, cover generation techniques generate a digital object only for the purpose of being a cover for secret communication. Due to the tremendous volume of information that is out there, it is impossible for a human to observe all communications around the world. Mimic functions can be used to hide the identity of a message by changing its statistical profile in such a way that it matches the profile of an innocent looking text. The English language possesses several statistical properties. For instance, distribution of characters is not uniform e occurs a lot more frequently than z. This fact is used in data compression schemes such as Huffman encoding. A mimic function can be constructed out of Huffman compression functions. These functions can only fool machines; to a human observer the mimicked text will look completely meaningless because of

grammatical errors. To overcome these limitations mimicry has been enhanced by the application of context free grammars. Context Free Grammars explain the rules of constructing sentences in languages from different parts of speech. Context Free Grammar can be used to create grammatically correct English text to hide messages. Spam mimic[i] provides a good example of a cover generation method.

## Conclusion

This paper covers the overview of some cryptography and Steganography methods, helps to successfully hide a message into other message or cover medium like image. These all techniques must face the challenges such as Robustness of the technique that hide the secrete message successfully, Security guarantees the safe.

## References

**Cachin, C. (1998),** An Information-Theoretic Model for Steganography, in proceeding 2nd Information Hiding Workshop, vol. 1525, pp. 306-318,.

**Artz, D. , (May-Jun 2001),** Digital Steganography: Hiding Data within Data, IEEE Internet Computing, pp. 75-80.

**Shiva Kumar K. B. et al, (JULY-AUGUST 2011),** Int. J. Comp. Tech. Appl., Vol 2 (4), 1035-1047.

**Johnson N.F. & Jajodia, S., (April 1998),** Steganalysis of Images Created Using Current Steganography Software, in Proceeding for the Second Information Hiding Workshop, Portland Oregon, USA, , pp. 273-289.

**Johnson N.F. and Jajodia, S., (1998)** "Exploring Steganography: Seeing the Unseen", IEEE, pp. 26-34.

**Tanako S. Tanaka K. and Sugimura T., (February, 2000),** Data Hiding via Steganographic Image Transformation", EICE Trans. Fundamentals, vol. E83-A, pp. 311-319.

**Bender W., Gruhl D., Morimoto N.and. Lu A, (1996),** Techniques for Data Hiding, Systems Journal, vol. 35.

**Shannon C. E., (1949).** Communication theory of secrecy systems, Bell Syst. Tech. J., vol. 28, no. 4, pp. 656–715

**Moulin P and Koetter R, (2005),** Data-hiding codes, Proceedings of the IEEE, 93 (12) pp 2083-2126.

**Bender W., Gruhl D., Morimoto N., Lu A, (1996).** Techniques for data hiding, IBM SYSTEMS JOURNAL, VOL 35, NOS 3&4.

**Marvel, L.M., Boncelet Jr., C.G. & Retter, C., (1999),** "Spread Spectrum Steganography", IEEE Transactions on image processing, 8:08

**Chandramouli, R., Kharrazi, M. & Memon, N., (October 2003)** Image steganography and steganalysis: Concepts and Practice, Proceedings of the 2nd International Workshop on Digital Watermarking.

**Provos N. , (2001),** "Defending against statistical steganalysis", in Proc. 10th USENIX Security Symposium, vol. 10, pp. 24-24, Washington DC.

**Solanki K., Sullivan K., Madhow U., and Manjunath B.S., and Chandrasekaran S., (Sep. 2005)** Statistical restoration for robust and secure steganography, in Proc. IEEE Int. Conf. on Image Processing, Genova, Italy, vol. 2, 11-14, pp. 1118-1121.

**Wayner, Peter (1992).** Mimic Functions. Cryptologia XVI:3. 192-213.

**Jonsson J. and Kaliski B. (2002)** On the Security of RSA Encryption in TLS. In Advances in Cryptology – Crypto