International Journal of Advanced Science and Technology
Vol. 3, February, 2009

Text Steganography: A Novel Approach

Debnath Bhattacharyya[1], Poulami Das[1],
Samir Kumar Bandyopadhyay[2], and Tai-hoon Kim[3]

[1]*Computer Science and Engineering Department,*
Heritage Institute of Technology, Kolkata-700107, India
{debnathb,dasp88}@gmail.com

[2]*Department of Computer Science and Engineering,*
University of Calcutta, Kolkata-700009, India
skb1@vsnl.com

[3]*Hannam University, Daejeon-306791, Korea*
taihoonn@empal.com

Abstract

In this paper, a security model is proposed which imposes the concept of secrecy over privacy for text messages. In the recent years, we have plenty of security tools which are developed to protect the transmission of multimedia objects. But approaches for the security of text messages are comparatively less. The model proposed by us combines cryptography, steganography (taken as security layers) and along with that an extra layer of security has been imposed in between them. This newly introduced extra layer of security changes the format of normal encrypted message and the security layer followed by it embeds the encrypted message behind a multimedia cover object.

Keywords: *Stenography, Text stenography, Security, Secrecy, Privacy.*

1. Introduction

In the field of Data Communication, security-issues have got the top priority. So, of late the degree of security provided by a security tool has become the main evolutionary criteria of it. Classical cryptography is one of the ways to secure plain text messages. Along with that at the time of data transmission, security is also implemented by introducing the concept of steganography, watermarking, etc. In this types of combined approach, there exits some drawbacks.

In remote networking, at the time of transmission of hidden encrypted text message, if the eavesdroppers get the track of the hidden text, then they could easily get the encrypted text. Now breaking of encrypted text message can be achieved by applying some brute force technique. So, there remains some probability of snooping of information. So, this type of techniques incurs another level of security which can route the Cryptanalyzer or Steganalyzer in a different direction.

The work proposed here represents a heuristic approach to introduce the concept of Multi Layer Data Security algorithm in the field of combined Cryptography and Steganography. The algorithm, that we have proposed here, will secure the text message in multiple protection layers. Here, we have used the concept of Cryptography and Steganography (as two Layers of Security) and in between them an extra layer of security is introduced. In this work, two new methods namely code_matrix mapping and matrix_pix mapping are used to

employ the above mentioned extra layer of security and in this paper we have considered the image file format as our covering multimedia object. But, the same technique can be applied for other multimedia file formats.

This work is specifically focused on protection of any information which is in the form of text. The design of this technique is based on extensive analytical as well as experimental modeling of the data-hiding process.

2. Related works

The most of today's steganographic systems use images as cover object because people often transmit digital images over email and other communication media. Several methods exist to utilize the concept of Steganography as well as plenty algorithms have been proposed in this regard. To gather knowledge in this particular research field, we have concentrated on some techniques and methods which are described below.

Least significant bit (LSB) insertion is a common and simple approach to embed information in a cover object. For images as a covering media, the LSB of a pixel is replaced with an M's bit. If we choose a 24-bit image as cover, we can store 3 bits in each pixel by modifying the LSBs of R, G and B array. To the human eye, the resulting stego image will look identical to the cover image [1, 2].

Hiding data in the features of images is also an important technique which uses the LSB modification concept. In this method, to hide data in an image the least significant bits (LSB) of each pixel is modified sequentially in the scan lines across the image in raw image format with the binary data. The portion, where the secret message is hidden is degraded while the rest remain untouched. An attacker can easily recover the hidden message by repeating the process [2, 3].

An interesting application of steganography and cryptography has been developed by Sutaone, M.S., Khandare, M.V, where a steganography system is designed for encoding and decoding a secret file embedded into an image file using random LSB insertion method. In that method, the secret data are spread out among the cover image in a seemingly random manner. The key used to generate pseudorandom numbers, which will identify where, and in what order the hidden message is laid out. The advantage of this method is that it incorporates some cryptography in that diffusion is applied to the secret message [4].

The next interesting application of steganography is developed by Miroslav Dobsicek, where the content is encrypted with one key and can be decrypted with several other keys. In this process, the relative entropy between encrypt and one specific decrypt key corresponds to the amount of information [5].

In 2007, Nameer N. EL-Emam proposed an algorithmic approach to obtain data security using LSB insertion steganographic method. In this approach, high security layers have been proposed to make it difficult to break through the encryption of the input data and confuse steganalysis too [6].

S. K. Bandyopadhyay, Debnath Bhattacharyya, Swarnendu Mukherjee, Debashis Ganguly, Poulami Das in 2008 has also proposed a heuristic approach to hide huge amount of data using LSB steganography technique. In their method, they have first encoded the data and afterwards the encoded data is hidden behind a cover image by modifying the least significant bits of each pixel of the cover image. The resultant stego-image was distortion less. Also, they have given much emphasis on space complexity of the data hiding technique [7].

There is also a good method proposed by G. Sahoo and R. K. Tiwari in 2008. Their proposed method works on more than one image using the concept of file hybridization. This particular method implements the cryptographic technique to embed two information files using steganography. And due to this reason they have used a stego key for the embedding process [8].

The method proposed by Adnan Abdul-Aziz Gutub, and Manal Mohammad Fattani in 2007 is also a good one. Their approach hides secret information bits within the letters benefiting from their inherited points. To note the specific letters holding secret bits, the scheme considers the two features, the existence of the points in the letters and the redundant Arabic extension character. They have used the pointed letters with extension to hold the secret bit 'one' and the un-pointed letters with extension to hold 'zero'. This steganography technique is found attractive to other languages having similar texts to Arabic such as Persian and Urdu [9].

Unfortunately, modifying the cover image changes its properties as well as its features, so eavesdroppers can detect the distortions in the resulting stego-image's statistical properties. In fact, the embedding of high-entropy data (often due to encryption) changes the histogram of colour frequencies in a predictable way. So, in order to obtain more security in our prescribed method, we have embedded entire text information behind the cover image by modifying the least significant bits. In true colour system, this type of modification does not affect the visual perception of human eye and thus the proposed model becomes more stringent.

3. Our work

Our work deals with the security of text messages at the time of sending it over the network. In our algorithm, we have used asymmetric key cryptography which means different keys are needed to encrypt and decrypt the data. Here we have divide the domain of the key selection into different sub domains (a random prime number, a randomly generated number, decimal value of the pixel (only R) from the cover picture). In this approach we have given strength on division of the domain together with the key length.

According to our concept, we encrypt the original text message letter by letter applying a function which involves certain mathematical operations using corresponding letters and also numbers from the original image. Then, we use two public keys and one private key for encryption and decryption. These keys are generated randomly following some constraints and equations. For encryption and decryption, we have used a mathematical operation called Multiplicative Modulo in between the text and the generated keys. The used mathematical relation is given below.

$$C = \text{remainder of } (a*b)/p$$

(Say, a, b are any numbers and p be a prime number such that $0 < a, b < p$.). This technique constitutes the first layer of Security in our model.

Now, in the next attempt we have used one new method code_matrix mapping. In this method, the encrypted code is first is broken digit by digit. Next digits are converted into binary matrices having size DP (Depth of the cover Picture) X x where x gives the resultant code plus 1 where the code is obtained from the encryption procedure of the text. Here, the content of the matrices are not important and it can have any binary value. This approach incurs second layer of Security in our model.

After that we have used another new method matrix_pix mapping. In this method the matrices obtained previously are mapped into zone of pixels having area DP X x (in bytes)

where again x represents the same (previously mentioned) and DP represents the depth of the picture, by using Steganography (Least Significant Bit of the pixel bytes are modified). Here, also after mapping of each matrix we have left one pixel unchanged after mapping a certain set of matrices (constituting a word) we have left 2 pixels unchanged. This type of operation implements third layer of Security in our work.

The change in Least Significant Bit in the value of Red-Green-Blue (pixel) is likely to be undetectable by human eye. Even if the hackers could predict that a message is hidden inside the image, then they could at most acquire the matrices. These matrices should be effectively converted to obtain the encrypted data. After this, the encrypted data needs to be decrypted with the use of the private key to obtain a cipher code which has to be again decrypted to finally retrieve the original text. So, for the hackers it is very difficult to salvage the data crossing these Multiple Layers of Security.

The entire heuristic is given below.

3.1. NATS_KEYGEN()

This is the main function in our algorithm. This function will be used in the sender side and will call other modules of our algorithm like KEYGEN and CRYPTOHIDE.

1. Select the file bearing the information or the message to be hidden as INFO_FILE.
2. Call NATS_KEYGEN() and obtain the two public keys. Distribute it to the intended receiver.
3. Next, call NATS_CRYPTOHIDE(TEXT_MSG, COVER_IMG) and obtain the desired FINAL_IMG
4. Send the FINAL_IMG and COVER_IMG separately to the receiver.

3.2. NATS_MAIN()

This function will generate two public keys. These keys will be used to encrypt the text message.

Arguments: This function will not take any argument.

1. Generate a prime number n, Where n > the highest number assigned to the alpha numeric character (if ASCII is used then n > 255), and n <= 9999.
2. Randomly generate a number b, Such that b < n. Here b and n are the two public keys which are generated randomly.

3.3. NATS_CRYPTOHIDE(TEXT_MSG, COVER_IMG)

This function will be used to encrypt the text message and hiding of the same after encryption.

Arguments: This function will take the required text message (TEXT_MSG) and the cover image (COVER_IMG) as argument and finally it will output the FINAL_IMG which will be a stego-image.

1. Select a bitmap image from the set of bitmap pictures available in the collection of Sender End.
2. Convert the text to the number system which has been followed in the previous method to generate the keys.
3. A mathematical function ϕ is used which gives the number of prime numbers below a given number, say p.
4. Here, the value of p depends on the decimal value of the R array of each pixel of the original image. The first letter corresponds to the first pixel, the next to the second and so on.

5. The ϕ function is then added or subtracted from the number by checking the parity of the R array of the used pixel.
6. Now we have,

$$c = (a*b) \bmod n,$$

Where, "a" is the number assigned to a text and "c" is the encrypted text obtained. So, from here we get the encrypted text corresponding to a normal or plain text.

7. Apply code_matrix mapping in which the number is broken into its digits to form the matrices where each digit plus one corresponds to the number of columns in the matrix and has a fixed row number of DP (Depth of the Picture).
8. Apply matrix_pix mapping in which the zone of pixels are selected whose area is DP (Depth of the Picture) * number of columns of the matrix. Now on the Bytes within that zone, Steganography using LSB bit modification is applied. After each digit a pixel is kept unchanged and after each code 2 pixels are left unchanged for identification.
9. Obtain the resultant stego-image as FINAL_IMG.

3.4. NATS_DECRYPT(FINAL_IMG,COVER_IMG,PUBLIC_KEY1,PUBLIC_KEY2)

This function will first unhide the encrypted message from the cover image. Next, it will decrypt the encrypted message and will generate the desired sent message.
Arguments: This function will take FINAL_IMG (which is sent to the receiver from the sender's side), COVER_IMG and two public keys which were generated previously as argument and finally it will output the normal message (sent by the sender).

1. Declare a binary matrix having dimension H_C X W_C. Starting from the first byte of the INFO_FILE till to the end byte, store the bits in the declared matrix.
2. The FINAL_IMG is compared with the COVER_IMG to obtain the matrices.
3. The number of columns of the matrices gives the digits. The digits are first serially obtained from those matrices which are placed 1 pixel apart. The digits for the next code are obtained from the matrices which are placed after 2 unchanged pixels.
4. The digits are combined to form the encrypted number.
5. Now d is a private key to decrypt the code. "d" is generated using the formula :

$$d*b = (1 + k*n),$$

Where, k is any integer and $d < n$.

6. Decrypt the number using the private key, d with the help of the formula:

$$a = (c*d) \bmod n,$$

Where, "a" is the encrypted number for the text.

7. The previous mathematical function ϕ(p) is again used on the encrypted code but with opposite mathematical operation to obtain the original number for the text.
8. Convert the numbers to obtain the original text.

4. Result

During our implementation phase, we have tested our algorithm for different sets of images as well as text messages. For each and every normal bitmap images the proposed technique is working fine. We have also calculated that using a standard 1024 X 768 bitmap image, we can hide approximately 23130 numbers of characters. So, to illustrate our model, we are showing only one satisfactory experimental result due to the limitation of space.

To test the algorithm, let we want to send the text message:

IEEE

The keys are first generated by finding a random prime number, n=521. Another random number is generated, the public key, b=297. Let the text is converted to ASCII. Hence, the text becomes:

75(I) 71(E) 71(E) 71(E)

Now, using ϕ-function and parity bit checking the numbers are converted as:

$$71 - \phi(56) = 55$$
$$71 + \phi(75) = 92$$
$$71 - \phi(124) = 41$$
$$75 + \phi(96) = 99,$$

Where, 56, 75, 124, and 96 are the decimal values of the R array of the pixels of the original cover image.

Now using the public keys, the text is encrypted.

$$(55 * 297) \bmod 521 = 184$$
$$(92 * 297) \bmod 521 = 232$$
$$(41 * 297) \bmod 521 = 194$$
$$(99 * 297) \bmod 521 = 227$$

These encrypted numbers are converted into binary matrices using code_matrix mapping method. For example, for the encrypted number 184 we get $A_{DP\ X\ 2}$, $A_{DP\ X\ 9}$, $A_{DP\ X\ 5}$ and similarly for the other numbers we get [$B_{DP\ X\ 3}$, $B_{DP\ X\ 4}$, $B_{DP\ X\ 3}$], [$C_{DP\ X\ 2}$, $C_{DP\ X\ 10}$, $C_{DP\ X\ 8}$], [$D_{DP\ X\ 3}$, $D_{DP\ X\ 3}$, $D_{DP\ X\ 8}$] respectively. We will consider only the pixel zone which will be described by the DP X n (e.g.; DP X 3) not the content of the matrices.

The used Cover Image is given below. It is a simple bitmap image having Bit Depth 24 bits per pixel. So, in our case the DP will be simply 3.

Figure 1. Cover Image [34748 Bytes]

The matrices which are obtained after the encryption process are hidden behind the Cover Image using our matrix_pix mapping method. As stated before, in this method we will consider the least significant bit of each and every pixel for the information hiding and thus the resultant stego-image which will be obtained afterwards, will be similar to the Cover Image. So, the Final Image will be distortion less.

Figure 2. Final Image [34748 Bytes]

This Final Image and the Cover Image will be sent to the receiver separately. In the receiver's side, first the private key has to be generated using the step 5 of section III D. In this example the private will be 207 (denoted as "d"). Now, if the decryption algorithm described in section III D is executed successfully then the receiver will get the original message "IEEE" back.

5. Conclusion

In this paper, the major importance is given on the secrecy as well as the privacy of information. So, to obtain privacy we have used the concept of cryptography and on the other hand to implement secrecy, we have used steganography. But, again we have felt that the introduction of another level of security layer can make the existing technique a stringent one. Thus with the addition of one layer of security, this model has been designed to obtain the Multi layer data security.

This algorithm is supposed to be more efficient as here from the resultant image it is difficult to guess the actual data that is hidden behind it. Application of strong encryption technique is also introducing further security over the hidden data. Again we have followed the zone wise pixel's LSB replacement scheme in the steganography part unlike normal method of data hiding using LSB bit modification. So to unhide the data, a definite heuristic approach is required unlike normal implementation. So, this type of introduction of steganography with cryptography will obviously discern our attempt from the existing one. Again, during the implementation phase we have seen that there is a possibility to hide huge number of characters (approximately 23130) in minimal amount of time which is not present in any kind of existing techniques.

The proposed approach has many applications in hiding and coding messages within standard media, such as audios or videos. Also, the model does not depend on the type of the text that is to be hidden. We can use any type of text (even text of different language) here and to work with it, the corresponding number system has to be chosen (here, we have used ASCII). As future work, we intend to study some more steganalytic techniques for text messages and to extend our model to mobile communication.

References

[1] Johnson, N. F. and Jajodia, S, "Exploring steganography: Seeing the unseen", IEEE Computer Magazine, pp. 26-34, February 1998.

[2] S.K.Bandyopadhyay, Debnath Bhattacharyya, Poulumi Das, S. Mukherjee, D. Ganguly, "A Tutorial Review on Steganography", IC3 Noida, pp. 106-114, August 2008.

[3] Kh. Manglem Singh, S. Birendra Singh and L. Shyam Sundar Singh, "Hiding Encrypted Message in the Features of Images", IJCSNS, VOL. 7, No.4, April 2007.

[4] Sutaone, M.S., Khandare, M.V, "Image based steganography using LSB insertion technique", IEEE WMMN, pp. 146-151, January 2008.

[5] M. Dobsicek, "Extended steganographic system", 8th International Student Conference on Electrical Engineering, FEE CTU 2004, Poster 04.

[6] Nameer N. EL-Emam, "Hiding a large amount of data with high security using steganography algorithm", Journal of Computer Science, Page(s): 223 – 232, April 2007.

[7] S.K.Bandyopadhyay, Debnath Bhattacharyya, Poulumi Das, S. Mukherjee, D. Ganguly, "A Secure Scheme for Image Transformation", IEEE SNPD, pp. 490-493, August 2008.

[8] G. Sahoo, R. K. Tiwari, "Designing an Embedded Algorithm for Data Hiding using Steganographic Technique by File Hybridization", IJCSNS, Vol. 8, No. 1, pp. 228-233, January 2008.

[9] Adnan Abdul-Aziz Gutub, Manal Mohammad Fattani, "A Novel Arabic Text Steganography Method Using Points and Extensions", Proceedings Of WASET, pp. 28-31, May 2007.

Authors

Debnath Bhattacharyya, M.Tech in Computer Science and Engineering from West Bengal University of Technology, Kolkata. He is working as a Lecturer with the Computer Science and Engineering Department at Heritage Institute of Technology, Kolkata. He was an Education Officer in Computer Society of India, Kolkata for 10 years. His research interests include Bio-Informatics, Image Processing and Pattern Recognition. He has 14 Years of experience in the line of Teaching and Projects. He is working towards his research, since, middle of 2006 under the supervisions of Prof. Samir Kumar Bandyopadhyay and Prof. Tai-hoon Kim. He has published 40 Research Papers in International Journals and Conferences and Three Text Books for Computer Science.

Poulami Das, M.Tech in Computer Science and Engineering from the University of Calcutta. She is working as a Lecturer with the Computer Science and Engineering Department at Heritage Institute of Technology, Kolkata. Her research interests include Bio-Informatics, Image Processing and Pattern Recognition. She has more than 4 Years of experience in the line of Teaching. She is working towards his research, since, middle of 2006 under the supervisions of Prof. Samir Kumar Bandyopadhyay and Prof. Tai-hoon Kim. She has published 36 Research Papers in International Journals and Conferences and one Text Book for Computer Science.

Prof. Samir Kumar Bandyopadhyay, B.E., M.Tech., Ph. D (Computer Science & Engineering), C.Engg., D.Engg., FIE, FIETE, currently, Professor of Computer Science & Engineering and visiting Faculty Dept. of Comp. Sc., Southern Illinois University, USA, MIT, California Institute of Technology, etc. His research interests include Bio-medical Engg, Mobile Computing, Pattern Recognition, Graph Theory, Software Engg.,etc. He has 25 Years of experience at the Post graduate and under-graduate Teaching & Research experience in the University of Calcutta. He has already got several Academic Distinctions in Degree level/Recognition/Awards from various prestigious Institutes and Organizations. He has published 300 Research papers in International & Indian Journals and 5 leading text books for Computer Science and Engineering.

Prof. Tai-hoon Kim, M.S., Ph. D (Electricity, Electronics and Computer Engineering), currently, Professor of Hannam University, Korea. His research interests include Multimedia security, security for IT Products, systems, development processes, operational environments, etc. He has 14 Years of experience in Teaching & Research. He has already got distinctive Academic Records in international levels. He has published more than 100 Research papers in International & National Journals and Conferences.