SECURED INFORMATION HIDING THROUGH SHANNON-FANO-ELIAS CODING TECHNIQUE

Sruthi K[1], Karthikeyan B[1],Palukuru Venkata Ragha Sharanya[1],Priyadarsini C J[1], Vaithiyanathan V[1]

[1]School of Computing, SASTRA University, Thanjavur-613401, India.

Abstract-

The idea of secretly sending information has seen its importance as old as the communication evolved itself. Steganography is nothing but to embed the secret information in an image, audio or in a video. The sensitive image that contains the secret information is called the stego-image. This information hiding makes sure that only the sender and receiver can suspect the existence of secret information in the image. The main advantage of the steganographic technique is that the secret information does not draw attention (remains subtle). In this way the hidden information is highly secured. Shannon-Fano-Elias is one technique of embedding the secret information where the embedding of secret string is done by generating a code word for each character in a string. By this technique high degree of data encryption is made possible. In order to make sure that the stego images are not being easily intruded, a check of the quality of stego images is done by calculating the Peak Signal To Noise Ratio (PSNR) and Mean Square Error (MSE). This paper provides a simple avant-garde review and the various analysis of the Shannon-Fano-Elias algorithm with some standards and general rules unfolded from various analysis.

Keywords: Steganography, Steganalysis, Cipher text, Code word, Arithmetic coding.

I. INTRODUCTION

Steganography is one technique of information hiding where the receiver decrypts the secret information with the help of a appropriate steganographic key. But this key-based technique can restrict only the attackers to know the algorithm but not the key. Steganalysis is the study of detecting the secret information from the stego-image. But the key detection or key intrusion technique has recently drawn attention.

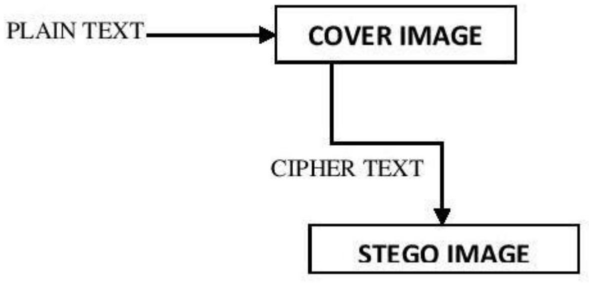

Fig. 1: Basic working of Steganography

The three different aspects that are to be considered on embedding the cipher text are capacity, robustness and security [9]. Capacity refers to how much secret information the image can hold. Robustness refers to the capability to withstand the modifications made and the strength of how far the cipher text is made safer from intruders attack. Security means that the cipher text is statistically undetermined by retaining the originality of the secrecy. Modern steganographic techniques use a secret key. This is analogous to Kerckhoff principle in cryptography [8] which states that the cryptographic system's security entirely depends on the key material.

II. LITERATURE SURVEY

Around a few hundred years ago Jerome Cardan, a famous mathematician procreated an ancient technique of secured message passing where the secret message is written in a paper with holes. This paper which acts as a mask between the sender and receiver appears as an innocent text to the third parties. This technique is referred as Cardan Grille.

During World War II, it was also reported that various steganographic methods such as Microdots and the use of invisible ink [11, 12] came into picture. This is considered to be one of the ancient methods of secret message passing. The use of null ciphers as secret encoders was also prevalent a few years ago. But the recent technological developments in the field of digital world have paved way for various steganographic tools

For the first time in information-theoretic method suffix sorting problem (lexicographical order) without using the suffix tree data structures was coined by Donald Adjeroh. This resulted in efficient space and time complexity [7].According to Fei Nan Lane, permutations of characters of input sequence is being performed that is closely related to suffix array and suffix tree [12]. There is also an indirect method of finding the probability of the source string. The Burrows-Wheller transform is one such method. This paved way for various probabilities modelling for statistical models.

According to Ki-Hyun Jung [13] the highest embedded capacity must be within the acceptable image quality and also the capacity of the embedding is restricted with the lower distortion of the image [11,14].The easiest way to hide data is to embed cipher texts by hiding bitmaps in colour picture[14]. Here the number of bits that is being traversed is perceived. As per the references of Kazem Qazanfari & Reza Safabakhsh, there is yet another technique to secure data which is to protect the encrypted bits of LSB by identifying the sensitive bits among them [16]. This technique also preserves the direct cosine transform (DCT) of the images [6, 7]. There are also various authors who contributed much for developing techniques that would enhance the security of cipher texts.

III. EVOLUTION OF SHANNON FANO ELIAS ALGORITHM

Professor Robert Fano and Shannon designed an algorithm where the code words are arranged according to decreasing probabilities that is put in binary form. But this suffered a drawback because the binary expansion might be long and infinite .So Shannon and Robert Fano worked together to formulate the algorithm efficiently. The modified Shannon Fano coding was based on the fact that the output sequences are arranged lexicographically (alphabetical order). Then the codeword for S is then terminated binary form of cumulative probability [7]. The codeword is generated for the entire source sequence. Shannon Fano coding satisfies the Coding Theorem for a Single Random Message.

Shannon's idea was to apply the compression of a random sequence of output sequences S. This paved way for the origin of Arithmetic coding. Since the output sequences are ordered lexicographically and not based on probability, it was essential to increase the code word by 1. But Professor Peter Elias (late) correlated this idea with the fact that it would be possible to compute the cumulative probability of a particular source code iteratively without the need of knowing the probability of all other output sequences. Hence the name Shannon Fano Elias coding.

IV. STEPS INVOLVED

Step 1: Assume X= {1, 2 ...m} and p(x)>0.

Step2: We now use a cumulative distributive function F(x) to allot code words for given symbols.

$$\sum_{a\leq x} p(a)$$

The modified cumulative function f'(x) is defined as

$$\sum_{a\leq x} p(a) + 1/2p(x)$$

In general this value which could have infinite number of bits in its binary form which is usually a real number, so it is not possible to use the same value as a codeword.

Step3: The basic idea of Shannon-Fano-Elias coding is to assign unique code words. The probability for each symbol is determined which can be represented by the Shannon code length equation of

$$l(x)= \log_2(\frac{1}{p(x)}) .$$

Step4: But in order to make sure that we get a prefix-free code, we modify the above equation to

$$\log_2(\frac{1}{p(x)}) +1.$$

Here, the extra bits are considered insufficient for large blocks.

Step 5: To check whether the code word for each symbol is prefix-free, we consider each codeword C(x) = c1, c2, c3...cl.or in other words, we consider the binary representation of F'(x) represented as 0.c1...c1.

In the graph given below Fig. 1, the cumulative distributive function of consists of many steps where each is of size p(x).

From the graph given below, it has been concluded that he value of the F(x) is found to be the midpoint of the step corresponding to x.

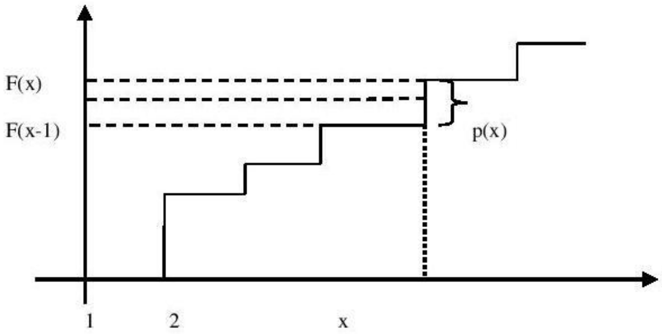

Fig. 2: Graph to show that F(x) is prefix-free to generate a unique code word.

Step 6: The value of F'(x) can be considered as the code for x. But generally, F(x) can be expressed only by an infinite number of bits (since they are real numbers). So it is inefficient to consider the values of F(x) alone as code words. So we truncate the values of F(x) to round off to the values based on l(x).Thus l(x) words suffice to describe x.

For example consider a string X= {"Shannon"}

The Shannon-Fano-Elias algorithm is applied to this string and their results are tabulated in Table.1

Table.1 Tabular plot to generate code words

X	p(x)	f'(x)	f'(x) in binary	l(x)	C(x)
1	0.1429	0.0714	0.0001	4	0001
2	0.1429	0.0214	0.0011	4	0011
3	0.4286	0.5000	0.0110	3	011
4	0.1429	0.7857	0.1100	4	1100
5	0.1429	0.9286	0.1110	4	1110

V. INTERLEAVING MECHANISM PROPOSED

This embedding technique of Shannon Fano Elias is combined with Least Significant Bit (LSB) approach. The LSB approach is very popular in spatial domain Steganographic method. The basic idea behind the LSB technique is to embed the secret information in the LSB of the cover images without extirpating the original features of it. Here, in this paper we have embedded the code words generated by the Shannon Fano Elias technique into the LSB of the images. Thus the secret bits are said to have a dual security from the scene of the intruders. There is always a bound for the number of LSB bits to be manipulated. If lesser the number of bits embedded, the more it becomes difficult for the interloper to intrude into the embedding technique used. Here, 4 bits of LSB are manipulated. PSNR and MSE for the stego images are calculated, which serves as a metric to review the image compression quality.

This was tested on various images of different dimensions and different string lengths. Based upon those results, the compression quality of each image by calculating their PSNR and MSE is shown in below.

Tab. 2 MSE and PSNR values of the images

S.No	ORIGINAL IMAGE	STEGO IMAGE	DIMENSIONS	PSNR (db)	MSE
1	Image 1	Stego Image 1	1200*1600*3	89.1620	0.0122
2	Image 2	Stego Image 2	1200*1600*3	86.0113	0.0183
3	Image 3	Stego Image 3	1200*1600*3	87.2227	0.0177
4	Image 4	Stego Image 4	1200*1600*3	86.8890	0.0191
5	Image 5	Stego Image 5	1200*1600*3	86.8290	0.0169

Also, by applying the LSB technique, it is visible that the pixel values differ from -4 to +4 which is quite negligible. Hence the embedding technique remains passive to the intruders.

The original images and stego images on applying the interleaving mechanism of Shannon Fano Elias technique and LSB technique is shown below.

Fig. 3(a): Image 1

Fig. 3(b): Stego Image 1

Fig. 3(c):Image 2

Fig. 3(d):Stego Image 2

Fig. 3(e):Image

3Fig. 3(f):Stego image 3

Fig. 3(g): Image 4

Fig. 3(h): Stego image 4

Fig. 3(i):Image 5

Fig. 3(j):Stego Image 5

VI. APPLICATIONS

1.Shannon-Fano-Elias coding finds its application in the IMPLODE compression method, which is a part of ZIP file format.

2.Arithmetic coding which finds its idea from Shannon-Fano-Elias coding, though produces a strong encoding, it is computationally expensive because they involve more patents.

3.More importantly, the intervals corresponding to the code words pattern are said to be disjoint, only if the code words are prefix-free.

4.Further it is also possible of coding several symbols in each codeword. This leads to arithmetic coding.

5.Various algorithms of Steganography is employed in copyright control of materials, smart.

6.Ids where the individual's full details are encoded in the photograph itself.

VII. RESULTS AND DISCUSSION

Here from Tab 2, it is clearly visible that the PSNR values of images range between 80db to 90db and the MSE values lie close to zero. Thus the PSNR values are found to be high and MSE values to be very low, proving that the embedded bits in the Stego images do not deviate much from the original images, hence showing a powerful level of image compression strength. This calculation proves to be very demanding as the image compression level is verified by standard objective methods of Steganography. Various techniques like Signal to Noise Ratio (SNR), Root Mean Square Error (RMSE) can also be used to check the compression level of the images.

These techniques are widely used in Arithmetic encoding, Digital Image Processing (DIP) and also in differing procedures to eavesdropping decoding.

Arithmetic encoding is a type of entropy encoding which is mainly used in lossless data compression. According to this, the most frequently occurring characters in a string are embedded with fewer bits and the least frequently used characters with more bits. This is in par with technique like Huffman coding. Shannon-Fano-Elias is an antecedent to Arithmetic coding. Here the probability of words are taken into account to determine the code words [9].They involve using lesser bits (code words) than the original representation. It is useful because the usage of data storage space or transmission capacity. Since it is a lossless compression technique, it allows for the original data to be reconstructed perfectly from the compressed data. This proves this technique to be efficient.

According to Welsted [3], this Shannon-Fano-Elias Coding is intuitively natural but not optimal. But the unique features of imperceptible, robustness and the capacity to hold hidden data as a highly secure one makes it more powerful than the other related techniques of water marking and cryptography.

Xiaoyu Ruan had discussed an alternation in the method of using Shannon-Fano-Elias coding where he mainly focuses on the problem of decrypting a binary sequence. Here a method is introduced to facilitate the cryptanalyst where if he knows the original dictionary words and the Probability Mass Function (PMF), instead of decrypting the coded sequence by eavesdropping method, he could find it by just within finite time by using the exhaustive search [2]. Thus the code word that is used is being truncated. This similar strategy can also be used to reduce the code length Arithmetic coding.

VIII. CONCLUSION

Various computer statistical study through Steganalysis provided promising ways to detect a distinction that is strenuous for humans to discern. Spatial-domain based technique has seen its growth to a greater extent by developing certain steganographic tools like StegoDos and S-Tools. Thus this massive growth of internet in addition to the evolution of Digital Image Processing (DIP), the information theory and coding theory of image processing has revolutionized the steganographic world. With the advent of digital coding theory of steganography, it has paved way for an environment that has generated various fascinating applications. Hence its evolution has seen tremendous improvement. This technique uses probability of occurrences of various symbols that occur. So it minimises overall computation time. Hence we use this technique of Shannon Fanon Elias Coding interleaved with LSB technique has paved a way to produce a stego image with very high quality. Also, the calculation of PSNR and MSE provides a duo proof of the secret text exposure to the intruders. In this way the security is greatly preserved.

IX. ACKNOWLEDGEMENT

We would like to thank SASTRA University for their support of providing us good infrastructure facilities for our research.

REFERENCES

[1] T. M. Cover and Joy A. Thomas (2006) Elements of information theory(2nd ed.). John Wiley and Sons.pp. 127–128. ISBNHYPERLINK "https://en.wikipedia.org/wiki/Special:BookSources/978-0-471-24195-9" 978-0-471-24195-9.

[2] Xiaoyu Ruan , "Cryptanalysis of Shanon-Fano- Elias codes " ,Proceedings International Symposium on

[3] Information Theory 2005,ISIT 2005.

[4] Suffix-Sorting via Shannon-Fano-Elias Codes Donald Adjeroh ? and Fei Nan Lane West Virginia University, Morgantown, WV 26506-6109, 2010.

[5] Liu J-, Tian Y-, Han T, Yang C-, Liu W-. Steganographic payload location for JPEG-decompressed images. Digital Signal Process Rev J. 38:66-76 , 2015.

[6] Aanandaprova Majumdar,Suvamoy Changder,"A novel approach for text steganography:Generating text summary using Reflection Symmetry",vol10,2013.

[7] Jung K-, Yoo K-. Steganographic method based on interpolation and LSB substitution of digital images. Multimedia Tools Appl. 1-13, 2014.

[8] Rajendra katti, Information Theory, 2005. ISIT 2005. Proceedings. International Symposium, September 2015.

[9] Aziz M, Tayarani-N MH, Afsar M. A cycling chaos-based cryptic-free algorithm for image steganography. Nonlinear Dyn 2015.

[10] Liu, C.-L., Liao, S.-R.: High-performance jpeg steganography using complementary embedding strategy. Pattern Recognition. 41(9) , 2945–2955, 2008.

[11] Kanso, A., Own, H.S.: Steganographic algorithm based on a chaotic map. Commun. Nonlinear Sci. Numer. Simul. 17(8), 3287–3302 ,2012.

[12] Using Improved Shannon-Fano-Elias Codes for Data Encryption, Xiaoyu Ruan ,North Dakota State University, IEEE 2006.

[13] Zhibin Pan , Sen Hu , Xiaoxiao Ma , Lingfei Wang

[14] ,"A new lossless data hiding method based on joint neighbouring coding "

[15] Chao RM ,Wu HC ,Le CC ,Chu YP," A novel image data hiding scheme with diamond encoding" EURASHIP J Inf Secure ,2009:1-9

[16] [Qin C, Chang CC, Hsu TJ, "Reversible data hiding scheme based on exploiting modification direction with two steganographic images "Multimed tools, April 2014.