iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** VII **Month of publication:** July 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VIII, July 2017- Available at www.ijraset.com

New Steganalysis Approach for JPEG Image Steganography using F5 Algorithm

Ms. Swati Mahamuni[1], Mr. Anand Sutar[2]
[1] *Department of Electronics and Telecommunication Engineering, MGM's college of Enginerring & Technology, Navi Mumbai, Maharashtra,India*
[2]*Department of Electronics and Telecommunication Engineering, Bharati Vidyapeeth's College of Engineering & Technology, Navi Mumbai, Maharashtra, India*

Abstract: Recently, there has been a lot of interest in the fields of Steganography and Steganalysis. Steganography involves hiding information in a cover (carrier) media to obtain the stego media, in such a way that the cover media is perceived not to have any embedded message for its unintended recipients. JPEG is a popular cover image format used in Steganography. Two well-known Steganography algorithms for hiding secret messages in JPEG images are: the F5 algorithm and Outguess algorithm. Steganalysis is the mechanism of detecting the presence of hidden information in the stego media and it can lead to the prevention of disastrous security incidents. In this paper, we present a steganalytic method that can reliably detect presence of messages hidden in JPEG images using the steganographic algorithm F5. The key element of the method is estimation of the cover-image histogram from the stego-image. By cropping JPEG image more times along the oblique direction, this algorithm can deal effectively with abnormal data and get better results.
Keywords : information hiding, DCT, F5, stego image, Steganalysis

I. INTRODUCTION

The Internet has revolutionized the modern world and the numerous Internet based applications that get introduced these days add to the high levels of comfort and connectivity in every aspects of human life. As the modern world is gradually becoming 'paperless' with huge amount of information stored and exchanged over the Internet, it is imperative to have robust security measurements to safeguard the privacy and security of the underlying data.

Steganography is a Greek word meaning covered or hidden writing. It is the art and science of secret communication, aiming to conceal the existence of the communication. In steganography, the message to be sent is concealed in such a way that an intruder would not know whether any secret communication is going on or not. Hiding information inside digital carriers is becoming popular. A rapid growth in demand and consumption of multimedia has resulted in data hiding techniques for files like audio (.wav), images (.bmp, .pnm, .jpg). Digital images are most common sources for hiding message. The process of hiding information is called an embedding. Least Significant Bit (LSB) embedding is the most widely used steganographic technique. In LSB embedding, the LSBs of uncompressed images are replaced with the message bits. JPEG is a popular cover image format used in steganography. The well-known Steganography algorithm for hiding secret messages in JPEG images is F5 algorithm. F5 is an information hiding algorithm based on frequency domain. It has been widely used because of its high capacity and robustness.

Steganalysis is the art of seeing the unseen. Steganalysis will analyze whether a given content, contains any secret message camouflaged into it. Steganalysis has gained prominence in national security and forensic sciences since detection of hidden (ciphertext or plaintext) messages can lead to the prevention of disastrous security incidents. The objective of steganalysis is to detect steganographic channels. Technically, steganography is considered broken when the mere presence of the secret message can be established [1]. The method is secure if the stego-images do not contain any detectable artifacts due to message embedding. In other words, the set of stegoimages should have the same statistical properties as the set of cover-images. If there exists an algorithm that can guess whether or not a given image contains a secret message with a success rate better than random guessing, the steganographic system is considered broken.

Fridrich present a steganalytic method [8,9,10] that can reliably detect messages (and estimate their size) hidden in JPEG images using the steganographic algorithm F5. This method is known as FR method. FR method can analyse the JPEG digital image and estimate the length of hidden information. However, if the cropping nature of the image more similar to image, the results by the FR method are more accurate. If not, the results may be inaccurate. The reason is that FR method uses a single-way cropping. So, the abnormality of data is inevitable. FR method cannot effectively deal with these abnormal data. Moreover, FR method did not give experimental details. Han[11] algorithm based on a combination of differences in coefficients of F5 steganalysis algorithm: the horizontal and vertical adjacent two sub-block adjacent to the frequency position of 0, a coefficient of portfolio differences in 12-

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VIII, July 2017- Available at www.ijraset.com

dimensional features, and the use of support vector machine classification. The method can detect image used by F5 algorithm. But the accuracy of the method is related with the establishment of FR method. So, there is the same problem.

II. PROPOSED ALGORITHM

A F5 steganography algorithm

F5 steganographic algorithm was introduced by Westfeld[4]. The F5 algorithm embeds message bits as the LSBs of coefficients along a key-dependent random walk through all DCT coefficients of the cover image while skipping the DC coefficients and all coefficients that are zeros. If the coefficient's LSB does not match the message bit, the absolute value of the coefficient is always *decremented*. If the subtraction leads to a zero coefficient (we say that so-called *shrinkage* occurred), the same message bit must be embedded at the next coefficient because at the receiving end the message is extracted only from nonzero coefficients. Fig 1. Shows F5 embedding process. This includes

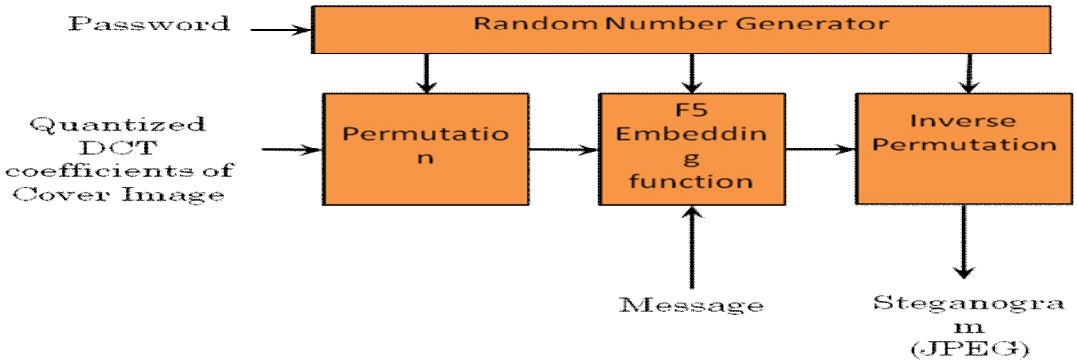


Fig 1 F5 Embedding Process

1) *Permutative Straddling:* The straddling mechanism used with F5 shuffles all coefficients using a permutation first. Then, F5 embeds into the permuted sequence. The shrinkage does not change the number of coefficients (only their values). The permutation depends on a key derived from a password. F5 delivers the steganographically changed coefficients in its original sequence to the Huffman coder. With the correct key, the receiver is able to repeat the permutation.
2) *Matrix Encoding:* Ron Crandall [5] introduced matrix encoding as a new technique to improve the embedding efficiency. F5 possibly is the first implementation of matrix encoding. If most of the capacity is unused in a steganogram, matrix encoding decreases the necessary number of changes. Let us assume that we have a uniformly distributed secret message and uniformly distributed values at the positions to be changed. One half of the message causes changes, the other half does not. Without matrix encoding, we have an embedding efficiency of 2 bits per change. For example, if we embed a very short message comprising only 217 bytes (1736 bits), F4 changes 1157 places in the image. F5 embeds the same message using matrix encoding with only 459 changes, i.e. with an embedding efficiency of 3.8 bits per change. The following example shows what happened in detail. We want to embed two bits $x1$, $x2$ in three modifiable bit places $a1$, $a2$, $a3$ changing one place at most. We may encounter these four cases:

$$x1 = a1 \oplus a3;\ \ x2 = a2 \oplus a3 \ \Rightarrow change\ nothing$$
$$x1 \neq a1 \oplus a3;\ \ x2 = a2 \oplus a3 \ \Rightarrow change\ a1$$
$$x1 = a1 \oplus a3;\ \ x2 \neq a2 \oplus a3 \ \Rightarrow change\ a2$$
$$x1 \neq a1 \oplus a3;\ \ x2 \neq a2 \oplus a3 \ \Rightarrow change\ a3$$

In all four cases we do not change more than one bit. In general, we have a codeword $\boldsymbol{a}$ with n modifiable bit places for k secret message bits $\boldsymbol{x}$.

While the F5 algorithm does modify the histogram of DCT coefficients, the some crucial characteristics of the histogram are preserved, such as its monotonicity and monotonicity of increments. The F5 algorithm cannot be detected using the $\chi 2$ attack because the embedding is not based on bit-replacement or exchanging any fixed Pairs of Values.

In the next section, we describe an attack on F5. It is based on the idea that one can accurately estimate the histogram of the cover-image from the stego-image. Because F5 modifies the histogram in a well-defined manner, we can calculate the number of modified coefficients by comparing the estimated histogram with the histogram of the stego-image.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VIII, July 2017- Available at www.ijraset.com

B Analysis of Histogram Modifications

Let $h(d)$, $d = 0, 1, \ldots$ be the total number of AC coefficients in the cover-image with absolute value equal to d after the image has been compressed inside the F5 algorithm. In a similar manner, we denote $h_{kl}(d)$ the total number of AC DCT coefficients corresponding to the frequency (k, l), $1 \le k, l \le 8$, whose absolute value is equal to d. The corresponding histogram values for the stego-image will be denoted using the capital letters H and H_{kl}.

Let us suppose that the F5 embedding process changes n AC coefficients. The probability that a non-zero AC coefficient will be modified is $p = n/P$, where P is the total number of non-zero AC coefficients ($P = h(1) + h(2) + \ldots$). Because the selection of the coefficients is random in F5, the expected values of the histograms H_{kl} of the stego-image are

$$H_{kl}(d) = (1-p)h_{kl}(d) + ph_{kl}(d+1), for\ d > 0$$

$$H_{kl}(0) = h_{kl}(0) + ph_{kl}(1), for\ d = 0 \qquad \text{--- (1)}$$

Let us further assumLet us further assume that we have an estimate $h\hat{}kl\,(d)$ of the cover-image histogram (the baseline). We can use this estimate to calculate the expected values $H_{kl}\,(d)$ using Eq. (1) and estimate p as the value that gives us the best agreement with the cover image histogram. Because the first two values in the histogram (d=0 and d=1) experience the largest change during embedding (see Fig. 2), we calculate p as the value that minimizes the square error between the stego-image histogram H_{kl}, and the expected values $H\hat{}_{kl}(d)$ calculated from the estimated histogram $h\hat{}_{kl}$ using Eq. (1):

$$p = \frac{h^{\wedge}_{kl}(1)\left[H_{kl}(0) - h^{\wedge}_{kl}(0)\right] + \left[H_{kl}(1) - h^{\wedge}_{kl}(1)\right]\left[h^{\wedge}_{kl}(2) - h^{\wedge}_{kl}(1)\right]}{h^{\wedge 2}_{kl}(1) + \left[h^{\wedge}_{kl}(2) - h^{\wedge}_{kl}(1)\right]^2} \qquad -(2)$$

The final value of the parameter p is calculated as an average over selected low frequency DCT coefficients $(k, l) \in \{(1,2),(2,1),(2,2)\}$.

$$p = \frac{p_{12} + p_{21} + p_{22}}{3} \qquad -\ (3)$$

The main reason why we decided to use histograms of individual low-frequency DCT coefficients rather than the global image histogram is as follows. Even with the low-pass pre-filtering, the spatial shift introduces some non-zero coefficients in high frequencies due to the discontinuities at block boundaries. And the values that are most influenced are 0, 1, and –1, which are the most influential in our calculations. Individual histograms of low frequency coefficients are susceptible much less to this onset of spurious non-zero DCTs.

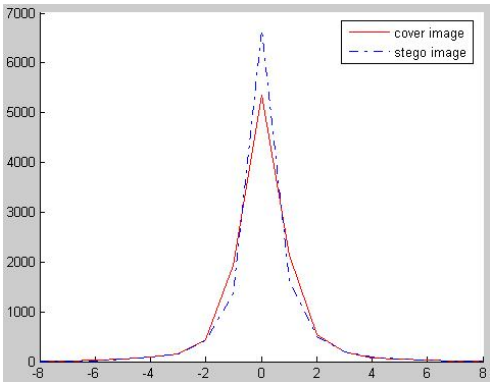


Fig 2 The effect of F5 embedding on the histogram of the DCT coefficient (2, 1)

C. Proposed Algorithm Principle

Accurate estimation of the cover-image histogram h is absolutely crucial for our detection method to work. We first decompress the stego-image to the spatial domain, then crop the image by 2rows and 4 columns, and recompress the cropped image using the same quantization matrix as that of the stego-image. The Image cropping method is a method by removing the image along the oblique direction. The image is not change of the more obviously visual effect. The values of image will be occurred to be more different than before. However, cropping of a single for image will produce abnormal data and have an effect on the steganalysis of the JPEG image. If cropping the image many times and calculating the averages of the DCT coefficients, the process will be able to reduce the effect of abnormal values. This is the Principles we used in our proposed method called as ESF algorithm.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VIII, July 2017- Available at www.ijraset.com

ESF algorithm cropped multiple different blocks more times. The averages of DCT coefficients will be to reduce the abnormal data on the impact of experimental results. In the cutting process, the DCT properties of experimental JPEG image were damaged. The DCT properties of new JPEG image will be to very closely to new DCT properties. The new DCT properties may be approximated to accurate values or not. The average of the new DCT properties by cropping image many times will be approximate to accurate values of the DCT properties of original carrier of JPEG image.

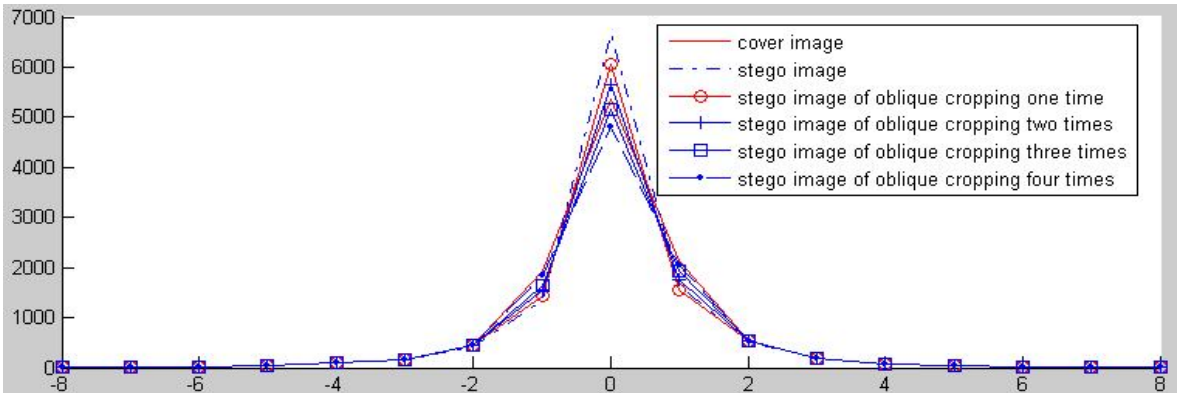


Fig. 3 The estimation of cover image histogram using ESF algorithm

Because the accuracy of the estimates is the major factor influencing the detection accuracy, we include a simple pre-processing step to remove possible JPEG blocking artifacts from the cropped image before recompressing. We use a simple uniform blurring operation with a 3×3 kernel B, B 22=1−4e, B 21 = B23 = B12 = B 32 = e, and B_{ij} = 0 otherwise. This low-pass filter helps remove some spurious non-zero DCT coefficients produced by "discontinuities" at the block boundaries, which are in the middle of the 8×8 blocks of the cropped image.

According to our experiments, the estimated histogram is quite close to the histogram of the original image In Fig. 3; we show a typical example of how good the histogram estimate is when compared to the histogram of the original image. The graph shows the original histogram values $h_{21}(d)$ (red colour), histogram values after applying the F5 algorithm with maximal possible message (dotted) and the estimate of the histogram for each cropped image.

III. EXPERIMENT AND RESULT

For F5 Steganography algorithm a database of BMP images has been created. Selected images are shown in Fig.4. The images were obtained using internet resized to a smaller, randomly chosen size, and saved as BMPs. A test tool using Matlab version 7 was developed to embed a text message in BMP images using F5 steganographic algorithm with quality factor 75 and to automate the testing and comparing processes of the FR [2] and ESF algorithms.

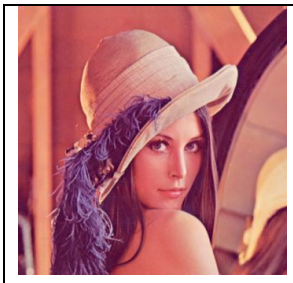


Image 2(256x256)

Image 3(512x512)

Fig. 4 Selected clean images from image database

The experimental results discussed hereafter are obtained by applying the test tool on the sets of test data outlined above. First, all clean BMP images are processed using F5 algorithm with 75% quality factor with embedding any messages of 1Kb and 2Kb and then applied our detection scheme to estimate the number of modifications p.

Stego images are generated by F5 steganography algorithm. F5 preserves the symmetric and monotonic shape of the histogram. The F5 algorithm does modify a macroscopic quantity of the JPEG file – the histogram of DCT coefficients – in a predictable manner. The number of zeros in the histogram increases due to shrinkage, while the histogram values for other coefficients decrease with embedding as shown in Fig.5.

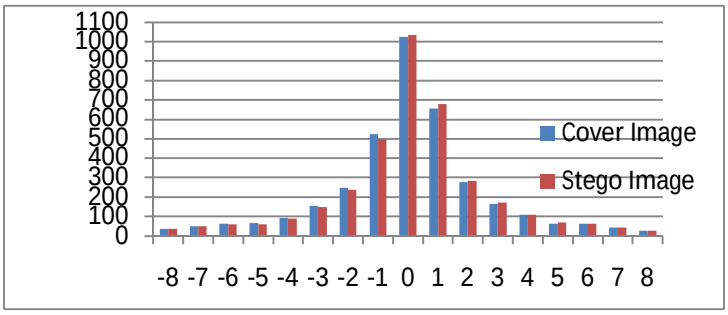


Fig 5. Comparison of DCT Coefficient Histogram for cover image & stego. Each bar represents number of occurrences for each value of DCT coefficients (2, 1).

The experimental results discussed hereafter are obtained by applying the test tool of the FR and the proposed ESF algorithms on the sets of test data outlined earlier**.** Fig. 6 shows a comparison of *p* values obtained from the data set by the FR and proposed ESF algorithms for a 1 Kb stego image of size 512x512 pixels. From Fig 6, it could be observed that the proposed algorithm yields better values for p than the original algorithm.

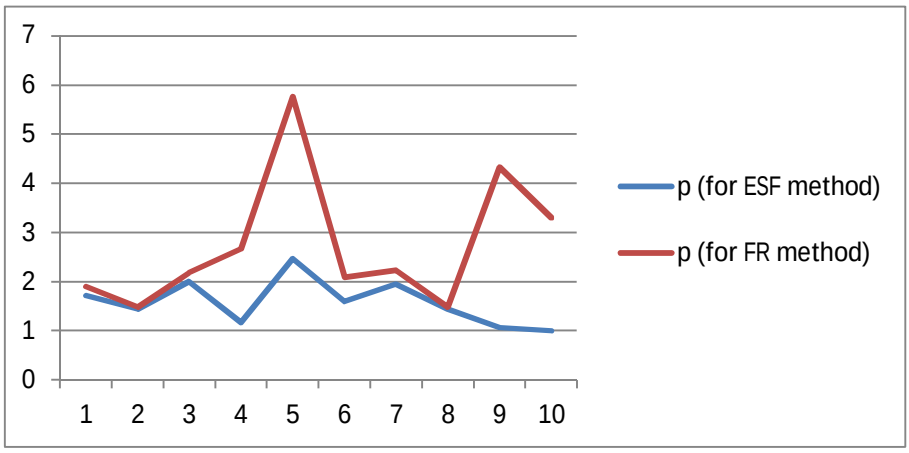


Fig 6 Comparison of p value for 512x 512 stego JPEG images for ESF and FR method for 1Kb hidden Message

Values for relative modifications calculated by proposed ESF method are more stable than that calculated by FR method for both stego and clean JPEG images. As per the given result we can easily detect the stego JPEG image by using proposed ESF method.

IV. CONCLUSIONS

For JPEG images, it is actually possible to construct from the stego image a new JPEG image that will have many macroscopic properties very close to the cover JPEG image. This is because the JPEG file is formed by quantized DCT coefficients, which are "robust" to small distortion, such as the one due to message embedding and previous JPEG compression. By cropping the (decompressed) stego image in oblique direction many times and recompressing it using the quantization table of the stego image, we obtain a JPEG file with macroscopic properties that well approximate the properties of the cover image. Because of the cropping, the newly calculated DCT coefficients will not exhibit clusters due to quantization. Also, because the cropped stego image is visually similar to the cover image, many macroscopic characteristics will be approximately preserved.

REFERENCES

[1] Ghgh T. Pevny ,J fridrich, and A.D. Ker. From blind To Quantitative Steganalysis Proc. SPIE, Electronic Imaging, Media Forensics and Security XI, San Jose, CA, January 18-22, 2009.
[2] Hang Fangju, Huang Jiwu. General-purpose image-based calibration JPEG steganalysis [J]. Science in China(Series F:Information Sciences). 2009 39(4): 383~390)
[3] Mao Jia-Fa; Lin Jia-Jun, Dai Meng. An Attacked Image Based Hidden Messages Blind Detect Technique [J]. Chinese Journal of Computers. Vol. 32 No. 2: 318~327.
[4] Westfeld, "High Capacity Despite Better Steganalysis (F5A Steganographic Algo-rithm)", in LNCS Vol.2137, Springer-Verlag, New York Heidelberg Berlin, pp. 289302,2001
[5] Ron Crandall: Some Notes on Steganography. Posted on Steganography Mailing List, 1998. http://os.inf.tu-dresden.de/~westfeld/crandall.pdf.

[6] Niels Provos Defending against Statistical Steganalysis [C] //Proceeding of the 10th USENIX Security Symosium. USENIX Press, 2001:323-335

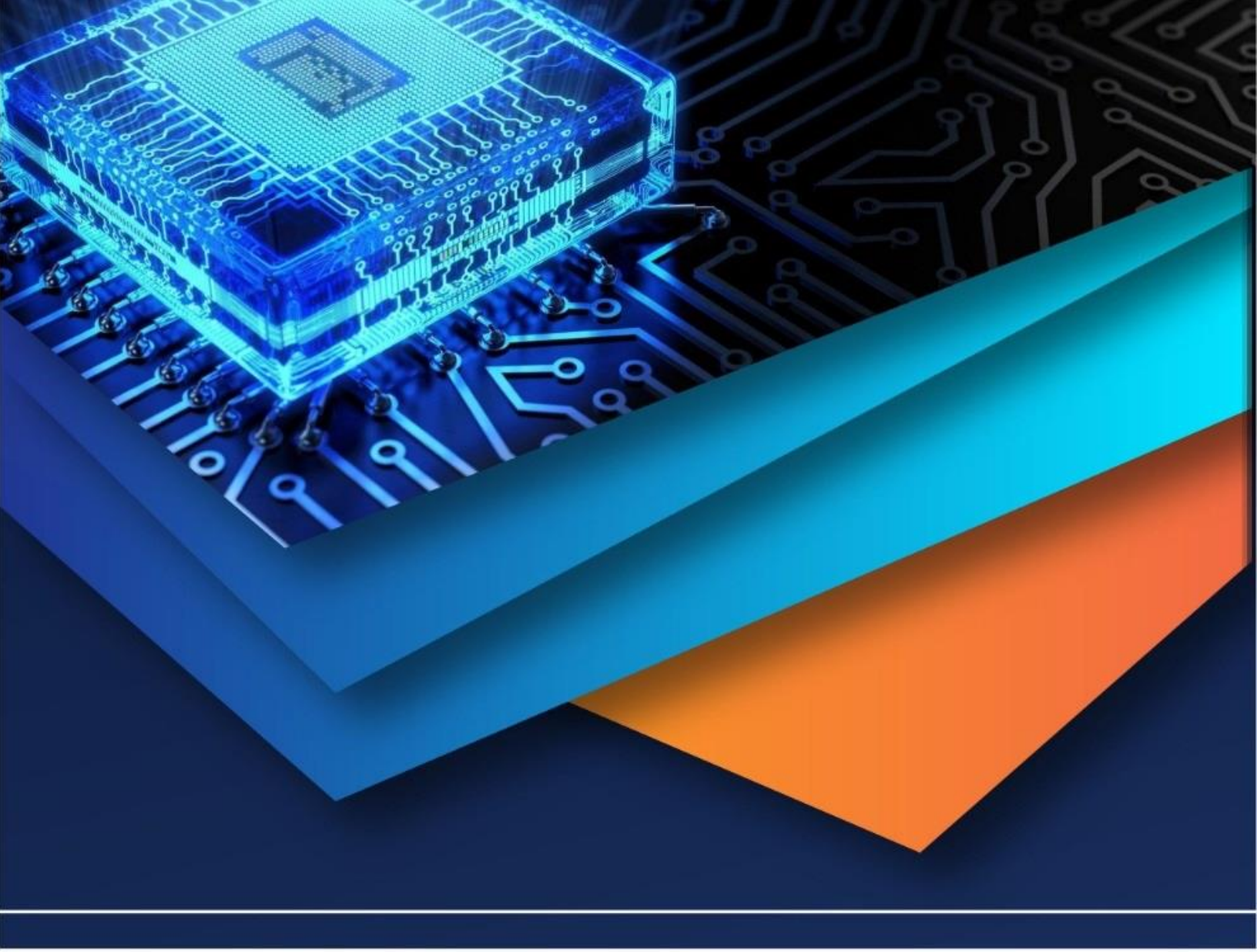
[7] P Sallee, Model-Based Steganography [C]//International Workshop on Digital Watermarking . Springer-Verlag 2004, 2939: 154 -167.

[8] Fridrich J,Goljan M, Hogea D.New methodology for breking steganographic techniques for jPEGs[C].CA:Proc EI SPIE Santa Clara,2003.143-155.

[9] Fridrich J, Soukal D.Quantitative Steganalysis of digital images:Estimating the secret message length[J]. ACM Multimedia Systems Journal, 2003,9(3):288-302.

[10] Jessica Fridrich, Miroslav Goljan, Dorin Hogea Steganalysis of JPEG Images: Breaking the F5 Algorithm[C]. Proceedings of the 5th Information Hiding Workshop,Lecture Notes in Computer Science,2002,2578:310-323.

[11] Han Xiao-dong; Ping Xi-jian; Zhzng Tao. Steganalysis Based on the Differences of Coefficient Combinations of 0,1 for Detecting F5 Steganography [J]. Journal of Information Engineering University.2009Vol. 10 No.2:184~187.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY