

Detection of Stego Images by principle of Suspicion Value for Distributing Stego Algorithms

Kaustubh Choudhary

ABSTRACT: Image based Steganography has been in use for securely broadcasting, sending malicious secret messages over the internet by means of hiding them in the images. Over a period of time Terrorist Organizations have mastered this technique of manipulating images to send data away from the preying eyes of law enforcement agencies. The cyber space is full of such malafide images containing hidden secret informations. In a single day over millions of images are uploaded and downloaded from internet and the most difficult aspect of tracking such malicious images is the absence of efficient, fast, reliable algorithms for identifying and isolating them from the bulk of innocent images. This paper endeavours to develop Distributing Spatial Domain Steganalysis Algorithm for detection by assigning it a Suspicion Value. The Suspicion Value is a number associated with the image and is bigger for those images which contain information and smaller for innocent images. The Suspicion Value thus sets a benchmark for images that need to be scrutinized or not, the thorough interrogation of an image requires large amount of computation time but it accurately pinpoints the hidden information using various techniques) or not. The stego-identifier algorithm developed has the capability to carry out tedious task of quick filtering of the suspicious images flowing through the web servers, routers, layer three switches and any other electronic media. Using this algorithm the suspicion values of various cover and stego-images generated by three different steganographic algorithms are computed and evaluated. All the graphs and tables are generated using MATLAB @Image Processing Tool Box.

Index Terms— Bit Plane Slicing, Cyber Crime, Distributing Steganographic Algorithms. Global Terrorism, Image Steganalysis, Multicolor LSB Transform, Pixel Aberration, SDT based Image Steganography, Suspicion Value.

1. INTRODUCTION

Image based steganography is a special technique of hiding secret messages in the image in such a way that no one apart from the sender and intended recipient suspects the existence of the message. This technique hides the contents in the picture and thereby makes detection very difficult, making it the preferred choice of criminals and terrorists.[1][2][3] Various agencies even claim that 9/11 attacks have been masterminded and planned using image based steganography.[4] The advantages like Anonymity, Electronic Dead Dropping, Secure Broadcasting and above all very high Secrecy and Security (explained in detail in Section 5.2.1 of [5]) come into play when this technique is used to modify images. Thus an innocent looking digital image on any Web Portal, Online Auction Site or even a Social Networking Site may be probably hiding a malicious and deadly terrorist plan or any other significant criminal Information. Steganalysis involves the process of identifying such stego- images (original image which is used for hiding data is called the Cover-Image whereas the image obtained after inserting the Secret Information in it is called Stego Image) from the bulk of innocent images. The next

step of steganalysis involves either the extraction of the hidden information or destroying the information by adding visually imperceptible noise in the image or can be even used for embedding counter-information in the Stego-Image thus defeating the purpose for which it was generated. Thus the most important part of Steganalysis is identifying the images given the fact that millions of images are on internet at any point of time.

The steganographic algorithms have a different signature that is algorithms depend on the way data has been encrypted on to the image. As mentioned in [7] the most spatial domain steganographic algorithms can be broadly classified in to two types- Distributing Steganographic Algorithms and Concentrating Steganographic Algorithms. In other words all spatial domain steganographic algorithms either embed the information in the Least Significant Bits of the pixel or change the entire color code of the pixel by inserting information in more than 2 bits of the pixel. Thus the Distributing Type algorithms are for images having data embedded in LSB while in the latter algorithm the entire information can be stored in very few pixels because large numbers of bits are available from every pixel for storing

information and hence called as Concentrating Type.

In this paper the suspicion value related only with the distributing steganographic algorithms is being determined using the properties of information pixels determined in [8]. Based on the suspicion value calculated in this paper (i.e. for distributing stego algorithms here onwards termed as Distributing Suspicion Value and represented by Λ) and detailed analysis of the concentrating stego algorithms a holistic suspicion value (catering for both the concentrating and distributing algorithms) is performed in [9].

2. PRELIMINARIES FOR DETERMINATION OF DISTRIBUTING SUSPICION VALUE (Λ)

The fast mathematical stego-identifier algorithm designed in this paper analyses any given digital image (for the presence of distributing spatial domain steganographic signatures) and quickly generates a Numerical Value (called in this text as Distributing Suspicion Value and denoted by Λ) corresponding to every image it has analyzed. This Suspicion Value is a number which is greater for those images which are more likely to have stego information and lower for innocent images.

2.1 Preliminaries and Definition

In order to generate the Suspicion Value few concepts will have to be defined for understanding and clarity. These preliminary concepts are derived from the concepts mentioned in [6] and [7].

Definition 1 (Image)

Every digital image is collection of discrete picture elements or pixels. Let M be any digital image with N pixels. So any particular pixel of image M is represented as $M(z)$ and z can be any value from 1 to N . This $M(z)$ can be a gray level intensity of the pixel in gray scale image or RGB or YCbCr value of the pixel in a color Image. The the individual RGB components of the pixel $M(z)$ in image M is represented as $M^R(z)$, $M^G(z)$ and $M^B(z)$ respectively. Thus $M(z)$ can be a set $\{ M^R(z), M^G(z), M^B(z) \}$ or equivalent gray scale representation or $(M^R(z) + M^G(z) + M^B(z))/3$. But it is always better to consider each R, G and B components individually because the averaging effect cause loss of vital steganographic information. Further $\langle \{M\}, m \rangle$ is multiset of Image M such that $M(z) \in \{M\}$ for every $z = 1$ to N and m is a vector corresponding to the occurrence or count of every element $M(z)$ in $\{M\}$. Mathematically an image M with N pixels is explained in (1) as:

$$\begin{aligned} &\text{For any Image } M \text{ with } N \text{ pixels} \\ &M(z) \text{ represents its pixels } \forall z : 1 \leq z \leq N \\ &M(z) = \{M^R(z), M^G(z), M^B(z)\} \\ &\quad \text{or} \\ &M(z) = (M^R(z) + M^G(z) + M^B(z))/3 \\ &\{M\} = \{M(z) : z = 1 \text{ to } N\} \\ &m = \text{occurrence of any particular } M(z) \in \{M\} \\ &\langle \{M\}, m \rangle \text{ is multiset representing image } M \end{aligned} \tag{1}$$

Definition 2 (Cardinality or Size of Image)

Any Image M consists of certain number of pixels. So any particular pixel of image M is represented as $M(z)$ and z can be any value from 1 to total number of pixels in the image. The cardinality or the size of the image M is the total number of pixels present in the image and represented as $n(M)$. So any Image M has $n(M)$ pixels. Thus

$$M = \bigcup_{z=1}^{n(M)} M(z) \tag{2}$$

Definition 3 (Component of an Image)

Any sub part of an Image is a component of the image. In other words any Image M can be broken down into pixel groups (or clusters) and each such cluster forms a component of the image and is identified by its unique set of pixels. Thus in a Image M the Pixels $M(z)$ from $z = 1$ to $n(M)$ are the elements of the image M and the subsets of the image M are composed of some of those pixels ($M(z)$ from $z = 1$ to $n(M)$) and thus forms the components of the image. Thus if the image M is broken in to K components then any component M^i of the image M is mathematically explained as:

$$M^i \subset M \text{ and } \forall i : 1 \leq i \leq K$$

$$M = \bigcup_{\substack{M^i \subset M \\ \forall i : 1 \leq i \leq K}} M^i \tag{3}$$

Also for every component M^i of the image M the $M^i(z)$ represents the pixels of the component M^i and $n(M^i)$ represents the number of pixels in M^i .

Definition 4 (Neighborhood or Locality of Pixel)

If $\ell(M(z))$ is said to be set of neighboring pixels of any pixel $M(z)$ in image M . Then any $n_i \in \ell(M(z))$ will be such that $d(n_i, M(z)) \leq \lambda$ where d is a function which calculates distance (can be Euclidean, City-Block, Chess Board or any other type depending upon the steganographic

algorithm) between its inputs (ie n_i and $M(z)$) and λ is measurement of degree of neighbourhood and should be minimum (Generally equal to 1 pixel) but also depends upon the steganographic algorithm used. Mathematically this can be represented as:

$$\ell(M(z)) = \{M(x) : M(x) \in M \text{ and } d(M(x), M(z)) \leq \lambda\} \quad (4)$$

An arbitrary pixel Y is shown with its neighbors P, Q, R, S, T, U, V and W. We represent this pixel Y as $\ell(Y)$ in mathematical notation. Thus $\ell(Y) = \{P, Q, R, S, T, U, V, W\}$ is set of neighboring pixels of pixel Y. Here $\lambda = 1$ and distance function d calculates Chess Board Distance.

P	Q	R
S	Y	T
U	V	W

Pixel Y

Definition 5 (Multicolored LSB Transform of Image)

The Multicolored LSB Transform of any digital image (elaborated in detail in Section 2.2.2 of [7]) is the Multicolored LSB plane corresponding to the image and is obtained by bit plane slicing of the same. In other words it is a variant of the image corresponding to the contribution of the LSB to the complete image. Since the intensity levels for any of the R, G or B value of any image in LSB plane is at max 1 and hence is negligibly small when compared with total of 255 intensity levels (assuming 24 bit BMP Image). As a result the image in its LSB Plane appears black. As a remedy a technique used here is that if any of the R, G or B value of the LSB Plane of the image is 1 then it is replaced by the intensity level of 255 whereas the intensity level of 0 is left unchanged. So the new image obtained has R, G or B value of every pixel as either 255 (if the corresponding true LSB value was 1) or else it remains 0. Therefore the modified LSB Plane of the image will appear multicolored (may have one of the seven possible colors ranging from three primary colors (Red, Green and Blue), four other colors obtained by their combination i.e. Yellow (R+G), Magenta (R+B) and Cyan (B+G), White (R+G+B) and Black indicating absence of any of the three primary color components). The Multicolored LSB Plane of the image has the maximum concentration of the hidden information and its significance is elaborated in detail in Section 2.2.2 of [7].

Multicolored LSB Transform of Image M is represented by M_{LSB} . Thus corresponding to the

pixel $M(z)$ of the image M the pixel in M_{LSB} is represented as $M_{LSB}(z)$. The relation between $M(z)$ and $M_{LSB}(z)$ is explained mathematically in (5).

Let M be the true image then the Multicolored LSB Transform of M is represented as M_{LSB} . R, G and B components of pixel $M(z)$ are $M^R(z), M^G(z), M^B(z)$. Then the corresponding pixel in M_{LSB} is given as

$$\begin{aligned} M_{LSB}^R(z) &= (M^R(z) \text{ modulo } 2) \times 255 \\ M_{LSB}^G(z) &= (M^G(z) \text{ modulo } 2) \times 255 \\ M_{LSB}^B(z) &= (M^B(z) \text{ modulo } 2) \times 255 \end{aligned} \quad (5)$$

Definition 6 (Deviation in the Multicolored LSB Transform)

In [7] it was concluded that any statistically significant Component (at least more than 50×50 pixels) of Multicolored LSB Transform of an innocent image has nearly equal distribution of Red, Green and Blue Components. The Degree of Deviation in Multicolored LSB Transform of the image is a quantifier that measures the amount of deviation from equal distributions of the red, green and blue components in the image.

This degree of deviation in Multicolored LSB Transform of the image M_{LSB} is represented as $\mathcal{D}(M_{LSB})$. Algorithm for determining $\mathcal{D}(M_{LSB})$ is given in (6).

Definition 7 (Gödel Color Code of a Pixel Color in Multicolored LSB Transform)

Gödel Code of any symbol is a unique number corresponding to the symbol. Since any pixel in the Multicolored LSB Transform of any image can have only of the seven possible colors (R, G, B, R+G, R+B, B+G, R+G+B and Black (absence of all the three R, G and B) as mentioned in Section 2.2.2 of [7]. Thus we can uniquely identify any pixel color in Multicolored LSB Transform by a unique numeric code called in this text as Gödel Color Code. We represent this Gödel Color Code for the color of the pixel $M_{LSB}(z)$ by $\varphi(M_{LSB}(z))$. Gödel Color Code $\varphi(M_{LSB}(z))$ is determined by associating every pixel's R,G and B value by a unique set of three different numbers i.e a unique triplet of numbers $\langle W_R, W_G, W_B \rangle$ acting as RGB _Identifiers. These numbers are such that the sum of any two numbers is different from any of the three numbers or the sum of any other two numbers or the sum of all the three numbers (eg: $\langle 1, 2, 5 \rangle$ or $\langle 1, 2, 4 \rangle$ (this triplet is used in chmod command of Unix based systems for directory r-w-x permissions and has similar purpose) but not $\langle 1, 2, 3 \rangle$ because $1+2 = 3$). Since every pixel is made up of the combination of the R, G and B values so by RGB _Identifiers each pixel

color can be uniquely identified by a unique number. In other words the

```

count_red = 0
count_green = 0
count_blue = 0
FOR z = 1 to n(MLSB)

    IF (MLSBR(z) = 255)
        count_red = count_red + 1;
    END IF

    IF (MLSBG(z) = 255)
        count_green = count_green + 1;
    END IF

    IF (MLSBB(z) = 255)
        count_blue = count_blue + 1;
    END IF

END FOR

maximum_component = MAX ( count_red , count_green , count_blue )

ratio_red =  $\frac{\text{count\_red}}{\text{maximum\_component}}$ 

ratio_green =  $\frac{\text{count\_green}}{\text{maximum\_component}}$ 

ratio_blue =  $\frac{\text{count\_blue}}{\text{maximum\_component}}$ 

D(MLSB) = 3 - (ratio_red + ratio_green + ratio_blue )
    
```

(6)

Gödel Color Code for any pixel in Multi Colored LSB Transform of an image can be obtained using RGB _Identifiers by adding only those values of W_R, W_G, W_B in the RGB _Identifier corresponding to which the RGB colors are present in the pixel. Mathematically this can be explained in (7).

let Multicolored LSB Transform of image M is M_{LSB}
 RGB _Identifiers = < W_R, W_G, W_B >
 where W_R, W_G, W_B are such that:
 $W_R \neq W_G \neq W_B \neq W_R + W_G \neq W_B + W_R \neq W_G + W_B \neq W_R + W_G + W_B$

$$\varphi(M_{LSB}(z)) = \frac{M_{LSB}^R(z)}{255} W_R + \frac{M_{LSB}^G(z)}{255} W_G + \frac{M_{LSB}^B(z)}{255} W_B$$

(7)

Definition 8 (Pixel Difference in Multicolored LSB Transform)

Two pixels in the Multicolored LSB Transform of the image M are said to be different if they are differently colored. The pixel difference is a function which determines whether two pixels are having same colors or are differently colored. If the two pixels are differently colored then it produces value 1 and if they are same it produces a 0 value.

The Pixel Difference of two pixels $M_{LSB}(y)$ and $M_{LSB}(z)$ is represented as $\vartheta(M_{LSB}(y), M_{LSB}(z))$ and its value is one if the difference of $\varphi(M_{LSB}(y))$ and $\varphi(M_{LSB}(z))$ is non-zero otherwise the Pixel Difference $\vartheta(M_{LSB}(y), M_{LSB}(z))$ is zero. Mathematically the Pixel Difference function ϑ in Multicolored LSB Transform is explained as:

$$\vartheta(M_{LSB}(y), M_{LSB}(z)) = \vartheta(M_{LSB}(z), M_{LSB}(y)) = \begin{cases} 1 & \text{if } \varphi(M_{LSB}(y)) - \varphi(M_{LSB}(z)) \neq 0 \\ \text{else} \\ 0 & \text{if } \varphi(M_{LSB}(y)) - \varphi(M_{LSB}(z)) = 0 \end{cases}$$

(8)

Definition 9 (Pixel Anomaly in the Multicolored LSB Transform)

In [7] it was shown that any image’s Multicolored LSB Transform can be broadly classified as Fine Grained, Coarse Grained, Boulder Grained and Continuous Grained. This classification was based on the degree of difference of any particular pixel in the Multicolored LSB Transform of the image from the immediately neighboring pixels. The immediate neighbors of any pixel $M_{LSB}(z)$ is the set of neighboring pixels of $M_{LSB}(z)$ obtained using locality function $\ell(M_{LSB}(z))$ with $\lambda = 1$ and distance function d calculating the Chess Board Distance (Definition 4, (4)). This degree of difference of any particular pixel in the Multicolored LSB Transform of the image from the immediately neighboring pixels is called Pixel Anomaly of the given pixel $M_{LSB}(z)$. Though logically the concept of Pixel Anomaly is similar to the concept of Pixel Aberration (Section 2.1 of [7] and Definition 7 in Section 2.2 of [6]) but since the Multicolored LSB Transform of any image has only 7 possible colors where as any normal image can have $256 \times 256 \times 256$ colors so the concept of Pixel Aberration (which is based on standard deviation) does not apply satisfactorily in the given case. A more specific method for determining the Pixel Anomaly in any Multicolored LSB Transform of the image is obtained using Pixel Difference function ϑ (Definition 8). The value of Pixel Anomaly for any particular pixel is said to be 1 if it is different from every other pixel in its immediate neighborhood where as its value is 0 when the given pixel is same (in color) as all its eight immediate neighboring pixels. Any value in between 1 and 0 indicates some of the pixels (the number of pixels depends on the magnitude of the value) in the neighborhood are different from the pixel concerned.

Thus Pixel Anomaly of any pixel in the multicolored LSB Transform of the Image is obtained by determining the mean of the modulus

of the Pixel Differences of the concerned pixel from its immediate neighbors (locality function $\ell(M_{LSB}(z))$ with $\lambda = 1$ and distance function d calculating the Chess Board Distance). Mathematically the Pixel Anomaly of any pixel $M_{LSB}(z)$ in multicolored LSB Transform of the Image M is represented as $\Omega(M_{LSB}(z), \ell(M_{LSB}(z)))$ and given in (9).

$$\begin{aligned} \ell(M(z)) &= \{M(x); M(x) \in M \text{ and } d(M(x), M(z)) \leq \lambda\} \\ \text{Where distance function } d &\text{ calculates} \\ &\text{the Chess Board Distance and } \lambda = 1. \\ \Omega(M_{LSB}(z), \ell(M_{LSB}(z))) &= \frac{1}{8} \sum_{\substack{p \in \ell(M_{LSB}(z)) \\ M_{LSB}(z) \in M_{LSB}}} \vartheta(M_{LSB}(z), p) \end{aligned} \tag{9}$$

Definition 10 (Mean Pixel Anomaly of the Multicolored LSB Transform)

The Mean Pixel Anomaly $\Omega(M_{LSB})$ of a Multicolored LSB Transform M_{LSB} of an Image M is the mean of the pixel anomalies of all the pixels in the given Multi Color LSB Transform of the image. Its value is said to be 1 when every pixel in the image is different from its immediate neighborhood. The value of $\Omega(M_{LSB})$ for the Image M_{LSB} is given in (10).

$$\Omega(M_{LSB}) = \frac{1}{n(M_{LSB})} \sum_{\substack{z=1 \\ M_{LSB}(z) \in M_{LSB}}^{z=n(M_{LSB})}} \Omega(M_{LSB}(z), \ell(M_{LSB}(z))) \tag{10}$$

Definition 11 (Zero Count of Pixel Anomaly of the Multicolored LSB Transform)

The Zero Count of Pixel Anomaly $\mathcal{Z}(M_{LSB})$ of a Multicolored LSB Transform M_{LSB} of an Image M is the measure of the number of pixels in M_{LSB} having Pixel Anomaly value as zero per unit pixel of M_{LSB} . I.e. it is the measure of the percentage of pixels with absolutely no difference in color with all the eight immediately neighboring pixels in M_{LSB} . $\mathcal{Z}(M_{LSB})$ is explained mathematically in (11).

$$\begin{aligned} \text{zero} &= 0 \\ \text{FOR } z &= 1 \text{ to } n(M_{LSB}) \\ \text{IF } \Omega(M_{LSB}(z), \ell(M_{LSB}(z))) &= 0 \\ \text{zero} &= \text{zero} + 1 \\ \text{END IF} \\ \text{END FOR} \\ \mathcal{Z}(M_{LSB}) &= \frac{\text{zero}}{n(M_{LSB})} \end{aligned} \tag{11}$$

Definition 12 (One Count of Pixel Anomaly of the Multicolored LSB Transform)

The One Count of Pixel Anomaly $\mathcal{O}(M_{LSB})$ of a Multicolored LSB Transform M_{LSB} of an Image M is the measure of the number of pixels in M_{LSB} having Pixel Anomaly value as absolutely one per unit pixel of M_{LSB} . I.e. it is the measure of the percentage of pixels which are having all its eight immediate neighbors absolutely different (in color) from the given pixel in M_{LSB} . $\mathcal{O}(M_{LSB})$ is explained mathematically in (12).

2.2 Theoretical Base for Determination of Distributed Suspicion Value of an Image

The distributed suspicion value Λ for any given image is the measure of the presence of information hidden in it using distributing stego algorithm. Thus before the Λ value is calculated for

$$\begin{aligned} \text{one} &= 0 \\ \text{FOR } z &= 1 \text{ to } n(M_{LSB}) \\ \text{IF } \Omega(M_{LSB}(z), \ell(M_{LSB}(z))) &= 1 \\ \text{one} &= \text{one} + 1 \\ \text{END IF} \\ \text{END FOR} \\ \mathcal{O}(M_{LSB}) &= \frac{\text{one}}{n(M_{LSB})} \end{aligned} \tag{12}$$

any image it is necessary to know the properties of stego-image (original image which is used for hiding data is called the Cover-Image whereas the image obtained after inserting the Secret Information in it is called Stego Image) generated using distributing algorithms. A 506 x 340 pixel BMP Image represented as M (Figure 1) is used as test image for analyzing three different stego-algorithms. It is a characteristic typical image which helps in creating all possible difficult scenarios which are possible in any real time situation and was selected after analyzing 200 different images. The three steganographic algorithms used in this paper were analyzed in [5] and are referred in Section 5 of [5] as *Algorithm designed in section 4*, *QuickStego Software* and *Eureka Steganographer* respectively. The stego-images obtained after inserting information (first section of this paper is inserted as secret message) from these three different algorithms are represented as S , T and U respectively. The first two stego-algorithms (stego-algorithms producing stego-images S and T) are distributing stego-algorithms while the third one (stego-algorithms producing stego-images U)

is concentrating stego-algorithm. First Distributing Stego algorithm (Producing Stego Image S) uses one bit LSB Insertion and inserts data column by column while the second Distributing Stego algorithm (Producing Stego Image T) uses two bit LSB Insertion and inserts data row by row. From visual appearance the images M, S, T and U look same due to inherent weaknesses in the Human Visual System (HVS) produced by Psycho-Visual effects of perception. As a result images S, T and U are not shown, because they look exactly same as M (Shown in Fig 1).



Fig 1 Typical Test Image Used for Creating All possible Difficulties in Steganalysis (represented M)

Although Multicolor LSB Transform of all these images is different (except that of Stego-Image U which is produced by Concentrating Algorithm). The Multicolor LSB Transforms of all the four Images (M, S, T and U) represented as M_{LSB} , S_{LSB} , T_{LSB} and U_{LSB} are shown in Figure 2.

2.2.1 Properties of Stego Image

Properties of stego-images (images containing information) depends upon the properties of those pixels in the stego-image which are storing the information. In other words the properties of stego-images become different from the innocent image (image without information) due to deformation produced in certain pixels of stego-image due to embedding of information in those pixels. The basic concepts of steganalysis of Distributing Stego Algorithms are given in Section 3.2 of [7]. In this section those concepts are examined with the help of mathematical quantifiers obtained using definitions given in Section 2.1 of this paper.

From [6] and [7] it can be conclusively said that information pixels have following 4 main properties:

1. Since the information pixels are suffering deformations so they are generally quite different from their immediate neighbors. As a result the pixel aberration of information

pixels is quite high. Since the concentrating algorithms bring bigger changes in the pixel so Pixel Aberration (Section 2.1 of [7] and Definition 7 in Section 2.2 of [6]) based analysis is more responsive to the steganalysis of Concentrating Stego-Algorithms but it is applicable occasionally on distributing stego algorithms too.

2. Information has maximum concentration in the LSB Plane of the image. But the LSB Plane of any image appears black and hence its contrast is increased by obtaining the Multicolored LSB Transform of the image. Thus in the Multicolored LSB Transform we can clearly see the information pixels differently colored from the innocent pixels. But since concentrating algorithms change only few pixels and as pixels are very small in size so few counted modified pixels in the Multicolored LSB Plane are imperceptible to human eye and are also from the statistical point of view are insignificantly less in number. But this method applies perfectly well in steganalysis of Distributing Stego-Algorithms because they modify large number of pixels.
3. In any statistically significant component (found to be 50×50 pixels) of the Multicolored LSB Plane the distribution of Red, Green and Blue components is significantly unequal among information pixels where as they are nearly equal for innocent pixels. Thus the degree of deviation (from R:G:B = 1:1:1) is more in the information pixels then the innocent pixels.
4. The information pixels are always present in the Fine Grained Pixel Clusters and rarely in the Coarse grained pixel clusters. They are always absent in the Continuous and Boulder Grained pixels Clusters. Refer Section 3.2 of [7] for details of the classification of pixel clusters. The Multi Color LSB Transform of the images with fine grained pixel clusters have majority of pixel with large value of Pixel Anomaly. Thus the value of the Mean Pixel Anomaly is largest in the fine grained pixel clusters and is lesser in coarse grained and even lesser in boulder grained and least in continuous grained pixel clusters.

2.2.2 Quantification of the Properties of Stego Image

Using definitions given in Section 2.1 the last three properties (the first property is related to Pixel Aberration which is more closely associated with Concentrating Stego Algorithms and hence dealt in

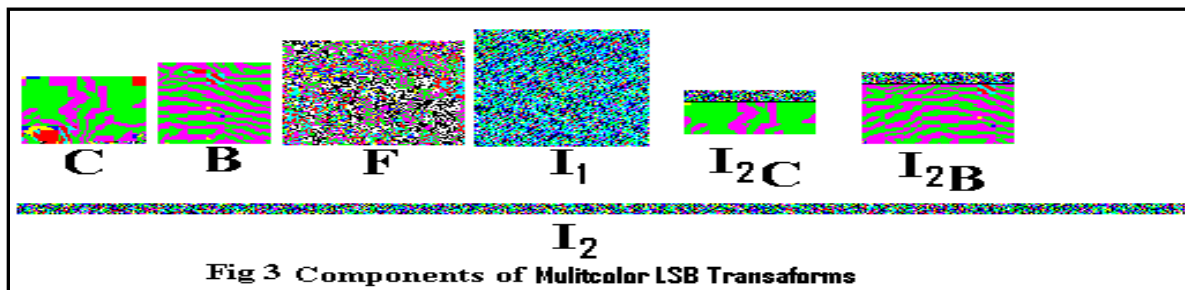
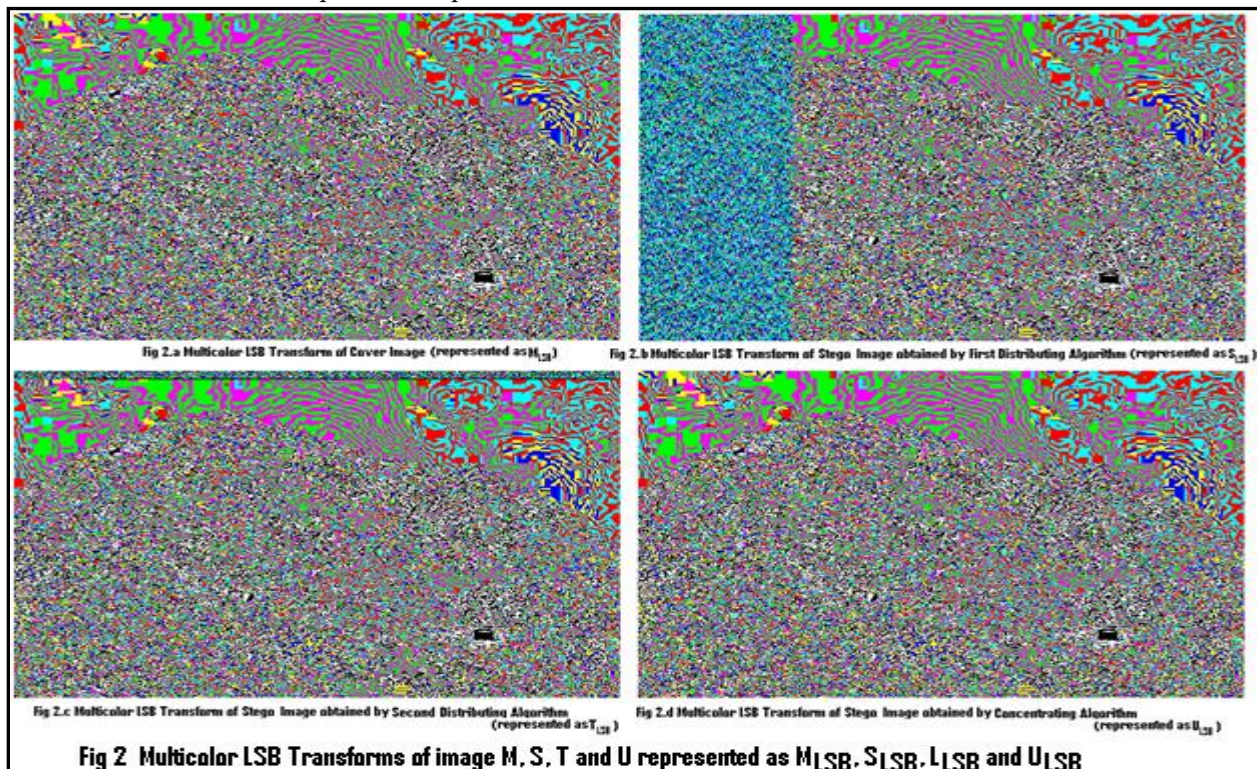
detail in [8]) associated with the stego-image is quantified in to a single equivalent numerical value corresponding to the given stego-image. This numerical value is distributed suspicion value Λ and is computed in such a manner that the images having information have bigger Λ and innocent images will have smaller Λ value.

3 DETERMINATION OF DISTRIBUTING SUSPICION VALUE (Λ)

To begin with we analyze the different components like Continuous Grained, Boulder Grained, Fine Grained, Information Components and even the Complex Components (Components consisting of either Fine grained or Information pixels along with Continuous Grained or Boulder grained pixels) of the Multicolor LSB Transforms of Cover Image and the Stego Images. The Continuous Grained Component is represented as C, The Boulder grained Component is represented as B, The Fine Grained Component is represented

as F and the Information Component obtained from Distributing Algorithm 1 (1 bit LSB Insertion) is represented as I_1 and Information Component obtained from Distributing Algorithm 2 (2 bit LSB Insertion) is represented as I_2 . In addition two complex components are also analyzed. The complex component consisting of Information Pixels (from I_2) and Continuous Grained Pixels is represented as I_2C whereas the complex component consisting of Information Pixels (from I_2) and Boulder Grained Pixels is represented as I_2B . These image components (C, B, F, I_1 , I_2 , I_2C and I_2B) are shown in Fig 3 and mathematically represented in (13).

$$\begin{aligned}
 C &\subset M_{LSB}, B \subset M_{LSB}, F \subset M_{LSB} \\
 I_1 &\subset S_{LSB}, I_2 \subset T_{LSB} \\
 I_2C &\subset T_{LSB}, I_2B \subset T_{LSB}
 \end{aligned}
 \tag{13}$$



The Mean Pixel Anomaly Ω , Zero Count z , One Count O , ratio of Zero Count to One Count z/O and Degree of Deviation D are calculated for these seven image components C, B, F, I₁, I₂, I₂C and I₂B. The results are tabulated in Table 1.

From Table 1 it can be concluded that the continuous grained pixels always have high value of Zero Count and degree of deviation but lower values of pixel aberration and One Count. Thus the value of the ratio of Zero Count to One Count z/O is higher in continuous grained components. The reason for this is that all the pixels in Continuous grained clusters are primarily composed of same colored pixels causing high values of zero count z and lower values of One Count O leading to reduction in pixel aberration Ω and hence are also biased towards any particular color causing higher deviation D . The boulder grained pixels also have lower Ω and O and higher value of z . But the degree of deviation D in boulder grained components depends upon the constituent colors forming the boulder grained cluster. For example the boulder grained image component B is

Table 1

$\Omega, z, \Omega \times D$ and z/O values for image components C, B, F and I

Image	Code	Ω	z	O	D	z/O	$\Omega \times D$	$\omega = \frac{\Omega \times D}{z + 0.01}$
Continuous	C	0.2025	0.4475	0.0049	0.8822	91.0769	0.1786	0.0020
Boulder	B	0.3073	0.1519	3.4590	0.0428	439.000	0.0132	2.9954e-005
Fine	F	0.7136	0.0018	0.1238	0.0417	0.0148	0.0298	1.2002
Information (distributing 1)	I ₁	0.8009	0	0.2427	0.7974	0	0.6387	63.8681
Information (distributing 2)	I ₂	0.8611	0	0.2313	0.4448	0	0.3830	38.2988
I ₂ with Continuous	I ₂ C	0.3180	0.4035	0.0658	1.1037	6.1333	0.3509	0.0571
I ₂ with Boulder	I ₂ B	0.3936	0.1355	0.0434	0.0879	3.1208	0.0346	0.0111

composed of colors Green and Magenta (Red + Blue) causing almost equal distribution of Red, Green and Blue components in the image causing degree of deviation to be close to zero but it is not necessary that D value will be close to zero for all boulder grained pixel clusters.

The information pixels I₁ and I₂ are perfect depicter of fine grained pixels and they have z value very close to zero (absolutely zero in most cases). In fact it was found to be absolutely zero for every other information pixel samples (including the information pixel samples of size 600 x 594). But this fact (information pixels having z values as absolutely zero) is still being examined and it may depend upon the stego-algorithm's encoding mechanism.

In fine grained pixel sample F and Information Pixel I₁ and I₂ the values of Ω and O are quite significantly high causing lower value of z . Thus the value of the ratio of Zero Count to One Count z/O is very low for all the three image components F, I₁ and I₂. The only difference between Fine Grained Innocent Pixel F and Information Pixel I₁ and I₂ (which is best example of Fine Grained Pixels) is due to the value of Degree of Deviation D which is very low for Innocent fine grained pixel and very high for Information pixel. The degree of deviation D is lower for I₂ because distributing algorithm 2 uses 2 bit LSB Insertion (i.e. uses both 1st LSB and 2nd LSB) causing lesser concentration of information in the 1st LSB.

The values of D and z for complex components I₂C and I₂B are very different from the values of D and z in Information Components I₁ and I₂. The values of D and z have been distorted due to the presence of Continuous Grained and Boulder Grained pixels in the information pixels. The same is Mathematically this is summarized in (14).

$$\begin{aligned}
 &D(\text{Information Pixels}) \gg D(\text{Innocent Fine Grained Pixels}) \\
 &D(\text{Continuous Grained Pixels}) \gg D(\text{Innocent Fine Grained Pixels}) \\
 &\Omega(\text{Information or Fined Grained Pixels}) \gg \Omega(\text{Continuous or Boulder Grained Pixels}) \\
 &z(\text{Information or Fined Grained Pixels}) \ll z(\text{Continuous or Boulder Grained Pixels}) \\
 &O(\text{Information or Fined Grained Pixels}) \gg O(\text{Continuous or Boulder Grained Pixels}) \\
 &z/O(\text{Information or Fined Grained Pixels}) \ll z/O(\text{Continuous or Boulder Grained Pixels}) \\
 &z(\text{Information Pixels}) = 0 \text{ (universality of this is yet to be verified)} \\
 &z/O(\text{Complex Components}) \ll z/O(\text{Continuous or Boulder Grained Pixels}) \\
 &z/O(\text{Complex Components}) \gg z/O(\text{Information or Fined Grained Pixels})
 \end{aligned}$$

Complex Components are those image components which have information pixels or fine grained pixels along with Continuous grained or Boulder Grained pixels like I₂C and I₂B in Fig 3.

(14)

Thus the entire outcomes of (14) can be summarized by a single quantifier ω (the last column entry in Table 1) and given in (15) as:

$$\omega(M_{LSB}) = \frac{\Omega(M_{LSB}) \times D(M_{LSB})}{z(M_{LSB}) + 0.01} \tag{15}$$

The value of ω is quite large for information pixels and much smaller for Continuous Grained, Fine Grained, Boulder Grained and even complex components.

The distortions in the values $\Omega, z, \Omega \times \mathcal{D}$ and z/\mathcal{O} for I2C and I2B indicate that in real time conditions these values may not yield suitable results. This is because in true scenario only part of the pixels of the image has information while rests of the pixels are image pixels (consisting of continuous, boulder, coarse and fine grained pixels) which are clearly seen in Fig 2. Thus multicolor LSB transform of entire image can be thought to be as a Single Complex Component.

To counter-verify this, the values of Mean Pixel Anomaly Ω , Zero Count z , One Count \mathcal{O} , ratio of Zero Count to One Count z/\mathcal{O} and Degree of Deviation \mathcal{D} is calculated for the entire image (cover image as well as the stego images i.e. images M_{LSB} , S_{LSB} , T_{LSB} and U_{LSB} .) in Table 2. The distortions in these values are quite evident from table 2 and make identification of stego image quite difficult. These distortions are due to the averaging effect introduced by majority of the innocent pixels.

Table 2

$\Omega, z, \Omega \times \mathcal{D}$ and z/\mathcal{O} values for images M_{LSB} , S_{LSB} , T_{LSB} and U_{LSB}

Image Type	Symbol	Ω	z	\mathcal{O}	\mathcal{D}	z/\mathcal{O}	$\Omega \times \mathcal{D}$	$\omega = \frac{\Omega \times \mathcal{D}}{z/\mathcal{O} + 0.01}$
Cover Image	M_{LSB}	0.6223	0.0550	0.1018	0.0358	0.5398	0.0222783	0.0405
Stego (Distributing 1)	S_{LSB}	0.6578	0.0398	0.1333	0.2583	0.2984	0.1699097	0.5509
Stego (Distributing 2)	T_{LSB}	0.6365	0.0469	0.1085	0.0360	0.4320	0.022914	0.0518
Stego (Concentrating)	U_{LSB}	0.6324	0.0549	0.1144	0.0360	0.4794	0.0227664	0.0465

In order to avoid this averaging effect the Multicolor LSB Transforms of the entire image is broken down into K smaller image components represented as M_{LSB}^i (for $i = 1$ to K) whose ω values are determined separately for each M_{LSB}^i . Since the ω is very high for information pixels and very low for all other pixels so the mean of the ω values for each component of the M_{LSB}^i will be much higher for the images having information pixels than the innocent images.

Also either the entire image may consist of information or the information may be present in the part of the image. If the information is present only in part of the image then for certain image components (consisting of information) the ω will be very high where as for all other image components (the innocent components) the ω will

be quite low. In such cases the variance of ω represented as $\text{Var}(\omega)$ will be quite high.

Thus for such images the value $I_1 = \frac{\bar{\omega} \times \text{Var}(\omega)}{100}$ will be quite high.

But if the entire image consists of information pixels only then although the mean of ω represented as $\bar{\omega}$ will be very high but the variance of ω represented as $\text{Var}(\omega)$ will be quite low (because ω is constantly high for every image component).

Thus for such image the value of $I_2 = \frac{5 \times \bar{\omega}}{\text{Var}(\omega)}$ will be quite high.

Value of ω is very low for all innocent components of image (Ranging from fine grained to continuous grained pixels). But its value is low even for complex components (components with either fine grained or information pixels along with continuous or boulder grained pixels) too so if such complex components are not separately handled then their presence may go unnoticed as smaller values of ω make negligible contribution in determination of mean of ω . Thus such components have to be reported separately as ambiguous components for detailed intensive analysis for presence of any information. The complex components can be easily identified by the value of z/\mathcal{O} which is neither extremely high nor very low but average (around 1.5 to 15).

In order to have minimum possible complex components the size of each image component M_{LSB}^i should be as small as possible. But it was observed that the statistical rule of R:G:B = 1:1:1 holds much better at larger pixel samples. Thus it was observed that at 50×50 pixel component size the results were most optimum.

For better understanding of the same we consider four different stego-images. One of them consists entirely of the information (every pixel contains information) and is represented as Entire. While three others represented as Partial₁, Partial₂ and Partial₃ contain information only in some of the image components. The Cover Image corresponding to Partial₁ is the rare occurring Biased Image (The Images Which do not follow R:G:B = 1:1:1 rule). We have selected this biased image (also used in Section 3.2 Fig 9 of [7]) to increase the difficulty level and to validate the universality of this technique. Partial₂ is the image corresponding to Fig 1. These four images along with their multicolor LSB Transform are shown in Fig 4. Corresponding to these four images the values of mean of ω given as $\bar{\omega}$, Variance of ω represented as $\text{Var}(\omega)$ and associated values

($I_1 = \frac{\bar{\omega} \times \text{Var}(\omega)}{100}$ and $I_2 = \frac{5 \times \bar{\omega}}{\text{Var}(\omega)}$) are determined for the

Cover as well as the Stego Images in Table 3. These values are also shown graphically in Figure 5. In figure 5 those complex image components are also shown which were reported ambiguous by the program and hence were isolated for detailed analysis. Only Eleven image components of size 51 x 51 pixels were reported ambiguous although the four images (seven if we include cover as well as the stego) which were being analyzed were of the size 1024 x 768 (Partial₁), 506 x 340 (Partial₂), 800 x 600 (Partial₃) and 102 x 337 (Entire). Thus the gigantic suspicious images could be reduced to small number (11 in the given case) of very small in size suspicious images and rest of the image components could be analyzed by using numeric value for identification of the authenticity of the image.

Thus from Table 3 we conclude that the mean of the values of $I_1 = \frac{\bar{\omega} \times \text{Var}(\omega)}{100}$ and $I_2 = \frac{5 \times \bar{\omega}}{\text{Var}(\omega)}$ perfectly

identify any image for presence of information (if hidden using Distributing Stego Algorithms). Thus we conclude that Distributing Suspicion Value for any image M represented as $\Lambda(M)$ is the mean of I_1 and I_2 or $\Lambda(M) = \frac{I_1 + I_2}{2}$. Mathematically this can be understood as:

$$\Lambda(M) = \frac{\frac{\bar{\omega}(M_{LSB}) \times \text{Var}(\omega(M_{LSB}))}{100} + \frac{5 \times \bar{\omega}(M_{LSB})}{\text{Var}(\omega(M_{LSB}))}}{2} \tag{16}$$

Thus based on preliminaries from Section 2 in this section (Section 3) we have devised the algorithm

4 CONCLUSION

In Section 3 a unique number $\Lambda(M)$ associated with any given image M is determined. From Fig 5 and the corresponding Table 3 it can be clearly seen that the value of this number $\Lambda(M)$ is larger for those images which have information embedded in them using distributing stego algorithms. But $\Lambda(M)$ does not respond well to concentrating steganographic algorithms because concentrating algorithms do not bring any statistically significant changes in the multicolor LSB transform of the image. Therefore this number $\Lambda(M)$ represents the distributing suspicion value of the image M. Thus although the Distributing Suspicion Value $\Lambda(M)$ is

a very good estimator of presence of information in the image (if hidden using distributing algorithm) but it fails if concentrating algorithms are used. Hence a holistic (which applies on both the Concentrating as well as Distributing Stego Algorithms) Suspicion value must be determined. for determining the distributing suspicion value Λ for any given image. The entire concept of determining the distributing suspicion value Λ for any given image M is given in (17).

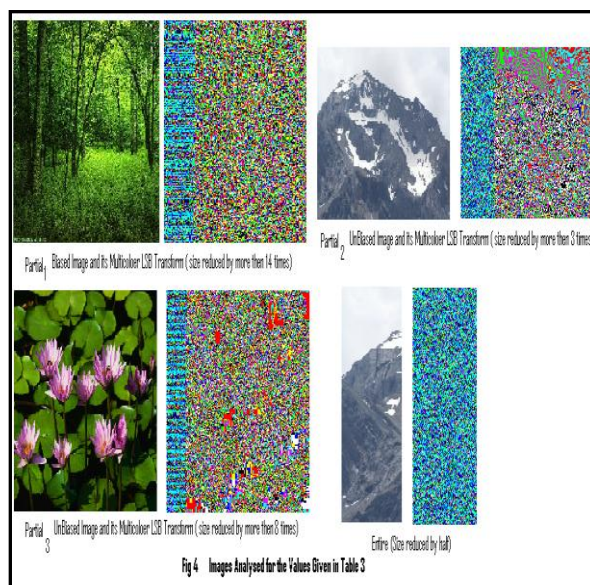
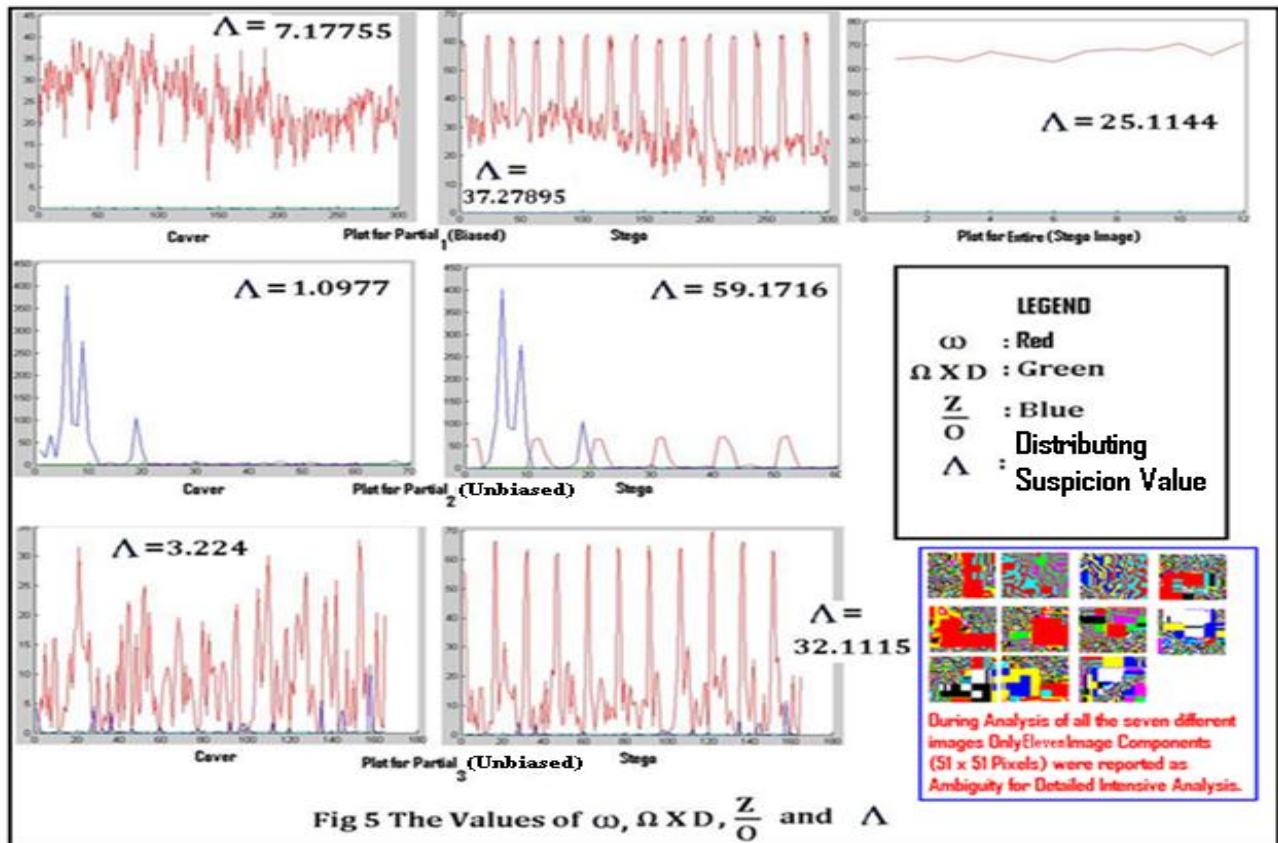


Fig 4 Images Analyzed for the Values Given in Table 3

Table 3

$\bar{\omega}, \text{Var}(\omega), I_1 = \frac{\bar{\omega} \times \text{Var}(\omega)}{100}, I_2 = \frac{5 \times \bar{\omega}}{\text{Var}(\omega)}, \Lambda = \frac{I_1 + I_2}{2}$ values for images given in Fig 4

Image	$\bar{\omega}$	$\text{Var}(\omega)$	$I_1 = \frac{\bar{\omega} \times \text{Var}(\omega)}{100}$	$I_2 = \frac{5 \times \bar{\omega}}{\text{Var}(\omega)}$	$\Lambda = \frac{I_1 + I_2}{2}$
Biased Cover (Partial ₁)	25.1721	46.2075	11.6314	2.7237	7.17755
Biased Stego (Partial ₁)	33.3868	221.0538	73.8029	0.7550	37.27895
Entire	66.7002	7.3585	4.9081	45.3207	25.1144
Unbiased Cover(Partial ₂)	1.2412	2.8736	0.0357	2.1597	1.0977
Unbiased Stego (Partial ₂)	16.7330	706.5376	118.2247	0.1184	59.1716
Unbiased Cover(Partial ₃)	8.9149	64.5869	5.7579	0.6901	3.224
Unbiased Stego (Partial ₃)	16.3910	390.5377	64.0131	0.2099	32.1115



M_{LSB} is Multicolor LSB Transforms of image M which is to be steganalysed
 Let M_{LSB} be broken into K Components given as $M_{LSB}^1, M_{LSB}^2, M_{LSB}^3, M_{LSB}^4, \dots, M_{LSB}^K$
 Let mean of $\omega(M_{LSB}^i)$ for all $M_{LSB}^i \subset M_{LSB}$ be represented as $\bar{\omega}(M_{LSB})$
 and Variance of $\omega(M_{LSB}^i)$ for all $M_{LSB}^i \subset M_{LSB}$ be represented as $Var(\omega(M_{LSB}))$
 Thus

$$M_{LSB} = \bigcup_{\substack{i=1 \\ M_{LSB}^i \subset M_{LSB}}}^K M_{LSB}^i$$

$$\omega(M_{LSB}^i) = \frac{\Omega(M_{LSB}^i) \times \mathcal{D}(M_{LSB}^i)}{\mathcal{Z}(M_{LSB}^i) + 0.01}$$

FOR $i = 1$ to K
 IF $(1.5 < \frac{\mathcal{Z}}{\mathcal{O}}(M_{LSB}^i) < 15)$
 Report M_{LSB}^i as ambiguous image component ;
 END IF
 END FOR

$$\bar{\omega}(M_{LSB}) = \frac{1}{K} \sum_{\substack{i=1 \\ M_{LSB}^i \subset M_{LSB}}}^K \omega(M_{LSB}^i)$$

$$Var(\omega(M_{LSB})) = \frac{1}{K} \sum_{\substack{i=1 \\ M_{LSB}^i \subset M_{LSB}}}^K (\omega(M_{LSB}^i) - \bar{\omega}(M_{LSB}))^2$$


$$\Lambda(M) = \frac{\bar{\omega}(M_{LSB}) \times Var(\omega(M_{LSB})) + \frac{5 \times \bar{\omega}(M_{LSB})}{Var(\omega(M_{LSB}))}}{2} \tag{17}$$

REFERENCES

- [1]. Infosecurity Magazine article dated 02 May 2012 reports that Al-Qaeda uses Steganography to hide documents.
<http://www.infosecurity-magazine.com/view/25524/alqaeda-uses-steganography-documents-hidden-in-porn-videos-found-on-memory-stick>
- [2] Daily Mail Online, UK article dated 01 May 2012 reported that a Treasure trove of Intelligence was embedded in porn video.
<http://www.dailymail.co.uk/news/article-2137848/Porn-video-reveals-Al-Qaeda-plans-hijack-cruise-ships-execute-passengers.html#ixzz1uIgxpire>
- [3].The New York Times article dated Oct 30, 2001 with title "Veiled Messages of Terror May Lurk in Cyberspace" claims 9/11 attacks planned using Steganography.
- [4]Wired article dated 02nd July, 2001 nicknamed Bin Laden as "the Steganography Master"
<http://www.wired.com/politics/law/news/2001/02/41658?currentPage=all>
- [5] Kaustubh Choudhary, Image Steganography and Global Terrorism, IJSER Volue 3, Issue 7, July 2012.

- http://www.ijser.org/ResearchPaperPublishing_July2012_Page2.aspx
- [6] Kaustubh Choudhary, Mathematical Approach to Image Steganography and Steganalysis, IJSER Volume 3, Issue 8, August 2012
<http://www.ijser.org/onlineResearchPaperViewer.aspx?Mathematical-Approach-to-Image-Steganography-and-Steganalysis.pdf>
- [7] Kaustubh Choudhary, Fast, Efficient and Reliable Approach to Spatial Domain Image Steganalysis, IJSER Volume 3, Issue 8, August 2012
<http://www.ijser.org/onlineResearchPaperViewer.aspx?Fast-Efficient-and-Reliable-Approach-to-Spatial-Domain-Image-Steganalysis.pdf>
- [8] Properties of Images in LSB Plane (A Steganalytic Perspective) by Kaustubh Choudhary (Submitted in your esteemed journal, having **Paper ID - I017925**)
- [9] Quick Identification of Spatial Domain Steganographic Signatures using Suspicion Value (Special reference to Concentrating Algos) **Paper ID – I018029**

BIBLIOGRAPHY OF AUTHOR

	<p>Kaustubh Choudhary Scientist, Defence Research and Development Organisation (DRDO) Ministry of Defence, Govt of India Current attachment: Attached with Indian Navy at Naval College of Engineering, Indian Naval Ship Shivaji, Lonavla - 410402, Maharashtra, India Email: choudhary.kaustubh@gmail.com</p>
--	---