E-governance Data Security using Steganography, Concepts, Algorithms and Analysis

Tanmoy Halder¹*, Sunil Karforma² and Rupali Mandal³

¹Dept. of Computer Application, Dr. B.C.Roy Engineering College, West Bengal, India.
²Dept. of Computer Science, The University of Burdwan, Burdwan, West Bengal, India.
³Dept. of Computer Application, Bengal College of Engineering and Technology Durgapur, West Bengal, India

*Corresponding author : Tanmoy Halder; tanmoyhalder@gmail.com

ABSTRACT

Steganography is the art of communicating in such a way that the existence of a secret information within another remains invisible. Since the Greek-era steganography has become a popular mode of secret communication. Modern day digital steganography has traveled through a long path of modification and development from basic LSB matching process to complex artificial intelligence based method. Transferring E-governance related information through non secure channel like internet is just not vulnerable for the data itself but could damage governmental security and privacy. Therefore this types of important documents could be protected form unauthorized viewers by applying steganography. In this paper we have reviewed various steganography techniques and analyzed their positive and negative sides in biref which could be applied for secure transaction of E-governance data.

Keywords: E-governance, Steganography, Steganalysys, Adaptive LSB, Component based LSB, Edge detection filter based Techniques, LSB Matching, LSB Substitution, Pixel Indicator Techniques, Pixel Value Differencing.

E-governance focuses more on interaction among citizens, community actors and stakeholders and their locally elected politicians (Pankowska, 2008). E-governance is defined as the manner in which power is exercised in the management of a country’s economic and social resources for development (Bhatnagar S. 2008). Implicit in the reference to ‘power’ is the concept of accountability. To enrich the power of accountability between government and all other parties related with it communication process must be secure, fast and liable to visibility. Steganography is the best option to fulfill such a safer communication rather than cryptography. Steganography differs from cryptography, hiding secret messages using steganography into a cover-media such that the unauthorized observer will not be able
to know regarding the existence of the hidden messages. It can be used for the benefit of the mankind to serve all type of E-governance digital secret message and data from all kind terrorists and criminal hacking for malicious purposes.

Text, Image, Audio and Video files are used as a medium to hide images are called cover-images. Cover-images with the secret messages embedded in them are called stego-images. For data hiding methods, the image quality refers to the quality of the stego-images. Whale hiding secret data this must be keep in mind that the quality of the cover image. Digital image and video contain high degree of redundancy in representation, thus appealing for data hiding. Steganography finds applications in copyright control of materials, enhancing robustness of image search engines and smart IDs, where individuals’ details are embedded in their photographs, video-audio synchronization, companies’ safe circulation of secret data, TV broadcasting, TCP/IP packets and check sum embedding [Johnson and Jajodia, 1998; Bender, Butera, Gruhl, Hwang, Paiz, Pogreb, 2000; Fridrich, Golan and Du, 2001]. It also finds application in medical imaging systems where a separation is considered between patients’ image data or DNA sequences and their captions, e.g., physician, patient’s name, address and other particulars. Cyber-crime is believed to benefit from steganography (Johnson and Jajodia, 1998). Examples are found for hiding data in music files (Hosmer, 2000), and even in a simpler form such as in HyperText Markup Language (HTML), executable files and Extensible Markup Language (XML) (Hernandez-Castro, Blasco-Lopez, Estevez-Tapiaidor, 2006). In the literature we have discussed about only image steganography. Types of different image steganography are LSB substitution, LSB matching, Adaptive LSB and Pixel-Value Differencing (PVD), Indexed-LSB, Edge detection filter based, Pixel Indicator techniques Component based LSB and Texture based techniques. The paper is organized as follows. In Section 2 classification and analysis of different methods in steganography, section 3 conclusions.

Different Steganography techniques

There are three basic types of steganography: spatial domain steganography, transform domain steganography and adaptive steganography.

Spatial Domain Steganography

LSB Substitution

Sharp.T in 2001 proposed simple Least Significant Bit (LSB) steganography, long-known to steganographers, in which the hidden message is converted to a stream of bits which replace the LSBs of pixel values in the cover image. (Sharp, 2001)

NEIL F. Johnson and Sushil Jajodia, 2000. Discuss three popular methods for message concealment in digital images. These methods are LSB insertion, masking and filtering and algorithmic transformations (Johnson, Duric and Jajodia, 2000).

(Chang et al., 2002) (Thien et al., 2003) had insinuate, LSB substitution is basically a method to directly switching the LSBs of pixel in the cover image with secret bits to get the stego-image. This method replaces the LSB of each pixel with the encrypted message bit stream to hide message in a host image. Authorized receivers can extract the message by decoding the host image with the help of a pre-shared
key. Capacity of algorithm is 1 bit per pixel. (Chang, Lin and Hu, 2002; Chang, Hsiao and Chan, 2003; Thien and Lin, 2003)

(Wang et al., 2000 and 2001) examined on to improve the quality of stego-image and employed a genetic algorithm to generate a substitution table. The substitution table consists the value of the secret data to be embedded into each host pixel is transformed to another value in advance which is closer to the original value of the host pixel. But substitution table may not be the optimal solution. (Wang, Lin and Lin, 2001)

To find out the optimal solution (Chang et al., 2003 and 2006) proposed dynamic programming (Chang, Hsiao and Chan, 2003; Chang, Chan and Fan, 2006) strategy. But the optimal substitution process may require huge computational cost because of using genetic algorithm and dynamic programming strategy. Rather to use genetic algorithm, an Optimal Pixel Adjustment Process (OPAP) is used to enhance the visual quality of the stego-image by LSB substitution.

**LSB matching**

(Ker et al., 2004) was established the LSB matching scheme. It is a different technique than LSB Substitution; it modifies the LSBs of the cover image. If one secret bit does not match the LSB of the cover image, then another one will be randomly added or subtracted from the cover pixel value. (Ker, 2004)

(Mielikainen, 2006) proposed a modified version of LSB matching, which progress it by lesser the expected number of modifications per pixel (ENMPP), from 0.5 to 0.375, so the histogram is less significant. (Mielikainen, 2006)

(LI, 2009) presented the simplification of LBS matching, in which sum and difference covering set of finite cyclic group were used to more decrease ENMPP and providing improved protection. (Li, Yang, Cheng and Zeng, 2009)

These techniques had some problems, mainly the artificial noises in the smooth regions of the image, which damage the visual quality of the stego-image. Not all pixels in the image can bear same length of changes without clear alteration, and as a result the stego-image has low quality. After research done, scheme establish that, if an image is processed with simple LSB substitution the histogram of the image will be showed in a “pair-wise” manner which known as Pairs of Values (PoV) which can be identified by Chi-square Test given by (Stanley, 2005). All LSB matching techniques were successfully attacked by best-known detector for LSB matching which is based on the center of mass (COM) of the histogram characteristic function (HCF) discussed by (Ker, 2005).

Halder and Karforma 2013 modified the basic LSB-replacement technique with indexing; this method doesnot directly hide the message in LSB but apply indexing and match the bits .chi-square test can detect the presence of hidden bits but could not retrive the bits as they are not directly embeded.

**Adaptive LSB and Pixel-Value Differencing (PVD)**

To recover the problem of LSB matching technique, a new LAB based method came like Human Visual System (HVS), which cover the characteristics to embed the secret data into the variable sizes of LSBs of each pixel, known as Adaptive LSB.
Halder et al., (2000) invented a piecewise mapping function according to the HVS difference sensitivity to decide the adaptive numbers of LSBs for data hiding.

After research (Lee et al., 2000) broken into the contrast luminance property of HVS and get a variable sized LSB insertion.

(Wu and Tsai, 2003) had introduced the method PVD Pixel value differencing). The method is basically used to progress the quality of stego-image. PVD make use of the HVS sensitivity to get power of different smoothness to high distinction by the choice of the width of the range which the difference value of two neighbor pixels.

(Liu et al., 2004) projected that, pixel of the original image is grouped according to its power, and each group is count up the occurrence of the original pixel, then a bit plane wise data hiding method is used to embed the secret message into the original image by the method of the pixel with high frequency precedence.

(Kekre et al., 2008) determined the fixed capacity of each pixel by considering the radiance from the highest bits of the remaining image.

These enhanced techniques of LSBs not completely use the HVS masking characteristics; particularly the Edge masking effect and they cannot obtain good invisibility.

(Chang and Tseng, 2004) proposed Side Match method and (Park et al., 2005) gave Neighborhood pixel information (NPI) method, that selected more than two neighborhood pixels to establish the consignment of each pixel. The embedding capacity of these methods is comparatively lesser than the PVD methods.

Combination of the LSB insertion and PVD methods, (Wu et al., 2005) proposed a data hiding scheme. The scheme is used by PVD method to get an enhanced image quality. Followed their scheme, two successive pixels are fixed by the LSB replacement method if their dissimilar value falls into a lower level; and likewise, the PVD method is used if their dissimilar value falls into a higher level. In other words, the secret data is hidden into the flat areas by LSB substitution and PVD methods in the border areas.

Again (Yang et al., 2006) gave an idea of Multi pixel differencing (MPD) and (Jung et al., 2008) combined MPD with LSB to guess efficiency of each pixel. Overview of PVD method was provided by (Liu et al., 2008). (Wang et al., 2008) demoralized modulus function with PVD to resolve its falling off border line problem and opposed to RS-analysis attack.

An adaptive LSB steganographic method introduced by (Yang et al., 2008) using PVD and LSB replacement. Using these methods, the dissimilar value of two successive pixels is used to estimate the hiding capacity into two pixels. The pixels are located in the border areas are fixed by a k-bit LSB substitution method with a greater value of k than that of the pixels located in flat areas. In the host image, method embeds maximum secret data into the border areas than the smooth areas. Even various problems are there like Falling off the boundary problem for border pixels, inserting data in the whole image even at low embedding rate, poor in resistance to statistical attacks. These were easily attacked by (Zhang et al., 2004), by change in histogram. All of the valuation principle like quality, capacity, security and complexity of data attaching were not met. More media rather than audio, video and media should also
be undertaken. These methods follow the rule that the border areas can bear more changes than flat areas. However, this opinion conformed by some existing data hiding schemes does not differentiate texture features from border ones; the border areas used by these schemes contain both borders and texture.

(Luo et al., 2010) worked on the problem of consistent inserting at all parts of an image irrespective of size of secret message and proposed LSB matching revisited. This border adaptive method can choose the insertion regions according to the size of secret message and the difference between two successive pixels in the cover image. For lower insertion rates, only sharper border regions are used while keeping the other smoother regions areas it is. When the embedding rate raise, more border regions can be released adaptively for data hiding by adjusting just a few limitations. The tentative results valuated on 6000 natural images with three specific and four universal steganalysis algorithms show that the new scheme can improve the security considerably.

(Maleki et al., 2011) worked on falling of boundary problem and security issues and provided an adaptive data hiding method based on four-pixel differencing collective with modulus function. The standard distinction value of a four-pixel block via an entrance secret key determines whether current block is situated in border or flat areas. Pixels in the edge areas are fixed by Q-bit of secret data with a larger value of Q than that of pixels located in flat areas and provides five secret keys to defend fixed secret data and problem of overflow or underflow does not take place.

(Joo et al., 2011) proposed Adaptive Steganographic Method Using the Floor Function and Modulus function with Practical Message to present better battle to attacks.

A four-pixel differencing and modified LSB substitution method had introduced by (Liao et al., 2011) to improve the quality of image. Secret data are hidden into each pixel by the k-bit modified LSB substitution method, where k is determined by the average disparity value of a four-pixel block. Readjustment has been implemented to remove the secret data exactly and to minimize the perceptual deformation but it negotiation struggle to attacks for realize the quality.

(Mandal et al., 2011) worked on DHPVD in which more number of secret bits is introduced to the border areas than flat areas to progress capacity. 2x2 non overlapping mask is chosen from the source image in row major order. The difference among two successive may fall into any one of four levels such as lower, middle1, middle2 and higher. Then depending upon the difference level, variable numbers of secret bits are fixed using a hash function in the successive two pixels in the non overlapping 2x2 mask. In addition to inserting the contents of the hidden image, dimension of the hidden image has also been fixed. A bit handling is used to minimize the difference between the source and fixed pixels but not challenging to attacks.

(Kumar et al., 2012) incorporates Tri way Pixel Value Differencing and LSB matching return for accomplish file as secret data as a means to show that the steganography can also work with other Medias than that of image, audio and video.

The above stated techniques worked on PVD which lacks to separate quality features from edges, time consuming, complex and can be attacked.
Edge detection filter based

Edge detection filtering based approach was introduced by (Alwan et al., 2005) to overcome the problem of PVD. The method use Sobel mask filter for inserting data in images using LSB, gray level connectivity using a fuzzy approach and the ASCII code.

(Negi, Santosh Arjun, N. in, 2006) projected adaptive Steganography based on filtering approach using both global and local image features.

(Hempstalk in 2006) established two new techniques FilterFirst and BattleSteg that is not in favor of more traditional image steganography procedures BlindHide and HideSeek. As a result that FilterFirst hit all the steganalysis procedures until insertion rates became greater than 7% and achieve better than all other steganography algorithms examined. Also, features of the cover, such as border, are better way of hiding information.

(Singh et al., 2007) encrypted messages in noncontiguous and random pixel locations in borders of images on gray images to avoid sequential attacks of all above mentioned techniques.

(Chen et al., 2010) suggested an original steganography method which is based on the LSB steganography mechanism. It utilized a hybrid edge detector which combines the fuzzy edge sensor with the careful edge detector. The hybrid edge detector support the new scheme in generating a better quality stego-image with high embedding rate of 2.86bpp with some confrontation to statistical attacks.

(Hussain, 2011) worked on data hiding method in the order of the edge boundary of an object by varying entrance value of filter. The tentative result shows very high rate of PSNR but projected scheme is aimed for low rate of hidden data capacity.

(Bassil et al., 2011) proposed a replication tool GhostBit based on Parametrized Canny edge detection algorithm to supply high fighting to Steganalysis attack. The limitation used are size of the Gaussian filter, a low and a high threshold value. These parameters can yield to different outputs for the same input image and secret data. Result; determine the inner-workings of the algorithm would be considerably ambiguous, misguiding steganalysts from the correct location of the secret data. The above techniques spotlight on gray images but the advancement in image technology to RGB directs to steganography application for color images.

Pixel Indicator Techniques introduced by (Gutub et al., 2008, 2009). They combined random pixel manipulation method and the stegokey to propose a technique, which uses the least two significant bits of one of the channels to point out the reality of data in the other two channels. His further aim was high capacity in RGB image based steganography by introducing the concept of storing variable number of bits in each channel (R, G or B) of pixel based on the authentic color values of that pixel i.e. lower color component stores higher number of bits. He also proposed Triple-A concealment technique method to hide digital data inside image-based medium. The algorithm attached more randomization by using two different kernels generated from a user-chosen key in order to select the components used to hide the secret bits as well as the number of the bits used inside the RGB image component. The randomization inserts more protection particularly if an active encryption technique is used such as AES. The capacity
ratio is increased above SCC and pixel indicator scheme. Triple-A has a capacity ratio of 14% and can be increased if more number of bit is used inside the components.

(Gandharba et al., 2012) come within reach to RGB channel based steganography technique which uses RSA algorithm for encryption and decryption. In an RGB image, each pixel (24 bits) is having R channel of 8 bits, G channel of 8 bits and B channel of 8 bits. The image is separated into 8 blocks and the cipher text is divided into 8 blocks. One cipher block is allocated to be embedded in only one image block by a user defined sub key. Out of the three channels in each pixel of the image one is used as the indicator channel. The indicator channel for the different blocks is not the same. The other two channels (called data channels) are used for hiding cipher text bits in 4 least significant bit (LSB) locations. In a data channel 4 bits of cipher text can be inserted if after embedding the change in pixel value is less than or equal to 7. The two LSBs of pointer will tell whether the cipher text is inserted in only one data channel or in both data channels, so that recover can be done accordingly at the beneficiary. But pixel indicator practice had a disadvantage that they treated all Red, Green, Blue components equally but in actual the role of all Red, Green, Blue components is not same for visual perception. So component based approaches were introduced.

**Component based LSB**

(Imran et al., 2007) introduced Component based Steganography for color images. They incorporated NPI, customized Least Significant Bits (LSBs) method for data implant and uses the green factor of the image as it is less sensitive to human eye and thus it is totally unfeasible for human eye to calculate whether the image is encrypted or not.

(Chang et al., 2008) also introduced a large payload data inserting method for color images. The advised method modifies the blue value of the color pixel in order to involve the secret data because the blue value is an insensible color to human eyes. Additionally, the number of secret bits that can be inserted into a cover pixel is dynamic and can be useful to both RGB and YUV color systems.

(Roque et al., 2009), presents a novel Steganography algorithm based on the spatial domain: Selected Least Significant Bits (SLSB). It works with the Least Significant Bits (LSBs) of one of the pixel color modules in the image and changes them according to the message’s bits to put out of sight. The rest of bits in the pixel color constituent desired are also changed in order get the nearest color to the creative one in the scale of colors. This new technique has been put side by side with others that work in the spatial domain and the huge difference is the fact that the LSBs bits of every pixel color component are not used to insert the message, just those from pixel color component selected.

The pixel of an color image consists of red, green and blue component and each of the component ranges from 0 to 255, in case of 24-bit illustration, (Mandal et al., 2012) projected pixel value differencing (PVD) method for secret data inserting in each of the component of a pixel in a color image while remove overflow (exceed 0-255) difficulty. Further security is present by using different number of bits in different pixel components.
Halder et al.,

Texture based Steganography

(Hamid et al., 2009) planned a LSB steganography approach based on texture analysis i.e. separating the image into simple and complex quality and hiding more data on complex textures than simple one. They planned two approach from LSB algorithm; the 3-3-2 approach without any boundaries on the type of images being used and can reach up to 33.3% of size of secret data, and the second one is the 4-4-4 boundaries on the type of images, the new approach features will enlarge the data hidden in the image by merge the above approaches. The major disadvantage of these texture based algorithms is that after the selection of pixels, they use LSB to embed the message and if we apply any filter in the stego-image the message is nowhere to be found.

Transform domain Steganography

The time when JPEG images are compressed to smaller file sizes, they are firstly transformed into the Discrete Cosine Transform (DCT) field which presents the data as high and low frequencies. The low frequencies are the low detailed areas and high frequencies are the high detailed areas. As the DCT values (referred to as coefficients) are twisted when compressing, it can similarly twist some of the values such that they grip message data. Inserting values in this fashion is much tough to notice from a steganalytical perspective than inserting in the spatial domain as the steganalyst will have to do a bit more digging to find any object of embedding. There are several converts that could potentially be used to embed the hidden data, including the Discrete Wavelet Transform (DWT), Fast Fourier Transform (FFT), Walsh-Hadamard Transform, and many more.

Discrete Cosine Transform-based Steganography

In Discrete Cosine Transform (DCT) based Steganography JPHide algorithm use the quantized DCT coefficients that are used to hide secret message bit are selected randomly by a key. The random generated key is generated by a pseudo random number generator, and the JPHide can also use the two LSBs of the selected coefficients.

(A. Westfeld., 2001) introduced the F5 steganographic algorithm. Rather to replace the LSBs of quantized DCT coefficients with the message bits, the absolute value of the coefficient is compact by the F5 algorithm by one if it needs alteration. Due to the author’s argument, the use the chi-square attack can never detect this type of value insertion. In addition to inserting message bits into haphazardly chosen DCT coefficients, the F5 algorithm utilize matrix embedding that decreases the number of changes necessary for hiding a message of a confident length. Both, the message length and the number of non-zero coefficients are required in the inserting process to determine the matrix insertion needed to decrease the number of alteration required in the cover image.

An another algorithm OutGuess in DCT based bring in by (Provos, 2003) as a UNIX source code for which there are two broadly known released adaptation. The first one is the OutGuess-0.13b, which is exposed to statistical analysis, and the second is OutGuess-0.2, which includes the ability to protect statistical properties. Henceforth, OutGuess refer to OutGuess-0.2. There are two stages representing
the embedding process of OutGuess. The first of which is that OutGuess inserts secret message bits alongside a random walk into the LSBs of the quantized DCT coefficients while omitting 0s and 1s. Presently after modifications are made to the coefficients already left during embedding to make the overall DCT. Histogram of the stego-image matches that of the cover image OutGuess cannot be an issue to a chi-square attack.

Andreas Westfeld’s (2001) steganographic system, F5.17 Instead of replacing the least-significant bit of a DCT coefficient with message data, F5 decrements its absolute value in a process called matrix encoding. As a result, there is no coupling of any fixed pair of DCT coefficients.

(X. Li and J. Wang, 2007) build up another steganographic method that adjusted the Quantization Table (QT) of JPEG compression and their method inserts the hidden bits in middle frequency coefficients.

Model-based steganography (MB) developed by (P. Sallee, 2003) for JPEG images. It accomplishes a high message capacity while remains secure against several first order statistical attacks.

Singular Value Decomposition-based Steganography

(Bergman and Davidson, 2005) enlarge an image steganographic technique based on Singular Value Decomposition (SVD). The cover image is factorized into three matrices, where two are orthogonal matrices, and represents a diagonal matrix, whose diagonal elements are the singular values of arranged in descending order of magnitudes. The top secret message bits are inserted into column elements of the environment by adjusting the controllable attributes such that it is still orthogonal after addition.

(Hadhoud and Shallan, 2009) proposed an image steganographic technique based on SVD that embeds the secret message in the orthogonal matrix U, leaving untouched the diagonal matrix S, for less embedding error and better image fidelity.

(Raja et al, 2009) suggested robust and high capacity image steganography using SVD (RHISSVD), which embeds message bits in singular values of the cover image.

Adaptive Steganography

Adaptive steganography is a special case of the two former methods. It is also known as “Statistics-aware embedding, masking or Model-Based. This method takes statistical global features of the image before attempting to interact with its LSB/DCT coefficients. The statistics will dictate where to make the changes (Kharrazi, et al.; Tzschoppe, Baum, Huber and Kaup, 2003).

Wayner (2002) dedicated a complete chapter in a book to what he called “life in noise”, pointing to the usefulness of data embedding in noise. It is proven to be robust with respect to compression, cropping and image processing.

The model-based method (MB1), described in, generates a stego-image based on a given distribution model, using a generalized Cauchy distribution, that results in the minimum distortion (Sallee, 2003).
Halder et al., Chin-Chen et al. (2004), propose an adaptive technique applied to the LSB substitution method. Their idea is to exploit the correlation between neighboring pixels to estimate the degree of smoothness. They discuss the choices of having 2, 3 and 4 sided matches. The payload (embedding capacity) was high.

Embedding is performed by replacing selected suitable pixel data of noisy blocks in an image with another noisy block obtained by converting data to be embedded.

Kong et al., (2009) proposed a content-based image embedding based on segmenting homogenous gray scale areas using a watershed method coupled with Fuzzy C-Means (FCM). Entropy was then calculated for each region. Entropy values dictated the embedding strength where four LSBs of each of the cover’s RGB primaries were used if it exceeded a specific threshold otherwise only two LSBs for each were used.

Chao et al., presented a 3D steganography scheme. The embedding scheme hides secret messages in the vertices of 3D polygon models.

Bogomjakov et al., hide a message in the indexed representation of a mesh by permuting the order in which faces and vertices are stored. Although, such methods claim higher embedding capacity, however time complexity to generate the mesh and then rendering can be an issue. Moreover 3D graphics are not that portable compared to digital images.

Conclusion
In this paper we have suggested a number of steganographic methods devolved and being used for secure communication in digital world. We have concentrated on image based methods because images are the most popular medium for hiding secret data due to greater capacity and easier embedding process. Our survey paper discussed almost all types of approaches starting from basic LSB replacement technique to recent FUZZY based steganography techniques and also discussed their later advancement. Almost all the techniques fulfill the basic necessity of data hiding, security, robustness and good PSNR. But all of them more or less suffer from any of these constraints. LSB matching and LSB replacement methods are simple but lack security level due to its simple well known process. PVD based methods have greater capacity but by histogram analysis they could be detected. Frequency domain techniques are tough to detect by steganalysis but includes complexity and sometimes results with low PSNR. Irrespective of all the good and bad points regarding all these algorithms we may come to the ultimate conclusion that steganography has secured the digital communication of important data, E-governance has come to successful implementation project in all levels due to the fact of security and authenticity of the documents, otherwise the novel project would have been a far from practical implementation. Steganography is a successful support of E-Governance if it could be used properly for removing threats against vulnerable attacks.

References
Philip Bateman, ‘Image Steganography and Steganalysis’ Supervisor: Dr. Hans Georg Schaathun. Department of Computing. Faculty of Engineering and Physical Sciences University of Surrey, Guildford, Surrey, United Kingdom.


