The Optimal LSB Substitution Matrix Technique for Image Hiding

Sunchai Suttipong[1], and Somboon Anekritmongkol[2]

[1] Rangsit University, Thailand, sunchai.s57@rsu.ac.th
[2] Rangsit University, Thailand, somboon.a@rsu.ac.th

ABSTRACT

Least Significant Bit (LSB) Substitution is a hiding information technique by embedded data in the medium and current information won't be recognized. A safety and the quality of the embedded medium are the main point of hiding data because each bits of the medium will be replaced by bits of confidential data. The mediator of data can be changed and noticed and the quality of those data can be measured from Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR). LSB substitution matrix was presented to increase the quality of the data which was decreased after the implantation. Every bits of the secret data will be counted as one pixel for changing the position within the matrix. In this current research, the researcher had presented the technique of finding optimal LSB substitution matrix by changing the position of each bits within the matrix integrated with changing the position with the pixel. When comparing the changing position of only with pixel, the changing position with pixel and bit give the better result for MSE and PSNR calculating.

Keywords: Image hiding, Least significant bit substitution, LSB substitution matrix, Ant Colony Optimization.

I INTRODUCTION

The communication through the Internet can increase the conveniences of exchanging data but it means that the safety of the data should be focused more. For example, "Operation Shady RAT", their targets are educational institutions throughout the world. A mechanism of an operation is to trick un-careless users opening emails which are created for preparing the back door to users' computers. In the next stages, it will connect to the website and download some files like general HTML or JPEG files. In reality, those files are hidden by the command controlling and can be logged through firewalls without any doubts. These command controlling will order the users' computers receiving an executable code from remote server to allow other users accessing any files within the system (Zielinska et al, 2014).

Besides using hiding data technique (steganography) for illegal purposes, those actions are for a criminal communication, a leakage of internal information systems, and an exchanging of cyber weapons in other aspects, hiding data technique also can be used to protect copyrights of creative works including some communication of significant data of the government and private in business sector, one interesting technique used to protect some works including songs, movies and works of arts is known as watermarking.

Hiding data (steganography) is often misunderstand with encryption (cryptography) if you look at the etymology derived from Greek, steganography means "covered writing" and cryptography means "secret writing" (Zielinska et al, 2014). An encryption is the technique using cipher algorithm to protect un-allowed users accessing. These are similar to some meaningless code groups or letter groups so these will be forwarded to receivers and they can use cryptographer key changing the information to the default format (Wayner, 2002). This means that the third person can visually see the communication even they can or can't decipher data. This is different from hiding data because those data will be hidden and the third person will not be able to know any communication.

Least Significant Bit (LSB) Substitution is the simple hiding information technique and most researchers give a lot of their attention to it. The data of secret message will be separated and embedded in the lowest bit position in each pixel of cover image for hiding data. After embedding secret message, cover image will be called as stego-image (Katzenbeisser, 2000; Kaur and Kochhar, 2012; Das et al, 2008). In some cases, the quality of stego-image may be low, the introduction of substitution matrix is for enhancing the quality of stego-image (Wang et al., 2001). Every *r* bit of secret message will be viewed as one pixel and each pixel will be switched positions referenced with substitution matrix. The finding of optimal substitution matric needs to spend time for a lot of calculating, so Genetic Algorithms (GA) is used to discover optimal substitution matrix taking less time.

Ant Colony Optimization (ACO) is presented by Dorigo which is one of stochastic optimization techniques. Dorigo et al. (Dorigo et al., 1996; Dorigo et al., 1997) and Merkle et al. (Merkle et al., 2002) research have given the result that ACO has more efficiency than GA for solving Traveling Salesman Problem (TSP). Therefore, the research of Ching-Sheng Hsu and Shu-Fen Tu (Ching-Sheng and Shu-Fen, 2010) had taken ACO search an optimal substitution matrix. Other research also has developed the technique of Wang et al. but the what remains the same is to use pixel for position switching (Chang et al, 2003; Yang and Wang, 2010). Thus, to increase the efficiency of stego-image, this research was conducted the comparison of the result in PSNR calculating between the position switching of pixels with the position switching of pixels and bit that is based on the finding of the optimal substitution matrix with ACO.

II THEORIES AND RELATED RESEARCH

A. The Hiding Data with Simple Least Significant Bit Substitution

Least Significant Bit substitution is the simplest technique of hiding data within the digital images. LBS represents to the position of bit on each byte with the minimum value of the pixel. Thus, when there is a change of bit in each position of pixels, the change of colors within pixels will be soft. For example, there are a pixel $(00100111)_2$ and secret message $(100)_2$, due to the length of the data is equal to three bits so the secret message senders can embed the data in the three rightest bits position of pixels and the result will be shown as $(00100100)_2$. For the receivers, they can extract the secret message by taking the three rightest bits position of pixels out. In case huge secret message, the amounts of bits need to be changed and this would make the differences of stego-image comparing with cover image before the embedding. The simplicity of the technique makes the senders and receivers can embed and pull the data out easily, and the third party can also be easily done as well.

B. The Optimal LSB Substitution

The optimal least significant bit substitution is presented by Wang et al. in 2001 and it aims to increase the qualities and safeties of stego-image. In the first step, the position each *r* bits of secret message *E* will be changed by Toral Automorphisms (Voyatzis, 1996) as follow:

$$f(x) = (k_0 + k_1 \times x) \bmod s \qquad (1)$$

With that k_0 and k_1 are key *x* which is the position of *r* bits and *s* is the length of Secret message. The highest common factor of k_0 and *s* is 1. By this method, we can receive secret message which is switched the position *E'*. Although the stego-image was stolen, the secret message still cannot be accessed without both k_0 and k_1 keys. Calculating substitution matrix $A = \{a_{ij}\}$ can increase the output to calculate PSNR of stego-image. From this procedure, the position of secret message *E'* will be switched to *E** referenced by matrix *A*. After two steps above, *E** will be embedded into the cover image *R* and the stego-image *Z* will show up.

Matrix A shown up in second stage will be used for value changing of the secret message. The differences between cover image and stego-image will be made as low as possible. The substitution matrix A is presented as below:

$$A = \{[a_{ij}] \mid 0 \le i, j \le 2^r - 1\} \qquad (2)$$

where a_{ij} is a member of set {0,1} *i* and *j* replaced by horizontal row (row) and vertical row (column) of matrix *A* and *r* is the amount of bits per pixel which are used to embed the secret message referenced by matrix *A*. *i* is the value of E' replaced by *j*. If a_{ij} is equal to 1 and there is only one part of vertical and horizontal rows. This part will be equal as 1 and the other parts are equal to 0. The example of the secret image *E'* and substitution matrix *A* can be shown as below.

$$E' = \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}_{10} \; and \; A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The value s of a_{00}, a_{12}, a_{21}, a_{33} of matrix *A* have shown that pixel 0, 1, 2, and 3 of *E'* will be replaced by position 0, 2, 1, and 3. After that, there will be secret image *E** as follow:

$$E^* = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}_{10}$$

The basic of finding the best of substitution matrix needs to take time for calculating. If the value *r* is high, all possibility is 2^r! so if *r* is equal to 4, the number of possibilities which must be calculated is 20,922,789,888,000.

Moreover, the research also takes the advantage from hiding data technique within noisy pixel and smoothing pixel areas. 8 bits of each pixel are divided into 2 parts: 4 bits on the left are called as high bits and other 4 bits are called as low byte. Noisy pixel area of the cover image represents to the neighbor of pixel of 8 pixels which have the same high byte value that less than a half or less than 4 pixels. Other pixels apart from those will be called as smoothing pixel. LSB technique can embed the amount of data much more than smoothing pixel without any observable traces of embedding data on the stego-image.

C. Ant Colony Optimization

Ant colony optimization (ACO) was presented by Dorigo et al. in year 1991 by observing the behaviors of ants (Dorigo et al., 2004). The results had been found that ants could walk through the path of trodden ants. This is because ants can release some chemicals called Pheromone by walking along Pheromone. Ants can find the best route or the shortest of foraging and nesting. From the finding, Dorigo had presented "Ant System" which consists of two parts: the first part had shown the steps of tour construction and the second part had presented Pheromone updating. In the system, ants will stimulate their own path after sampling to find initial capacity. Ants *k* are determined to stand on point *i* and point *j* is another point by applying Random Proportional Rule as below:

$$p_{ij}^{k} = \frac{[\tau_{il}]^{\alpha}[\eta_{il}]^{\beta}}{\sum_{l \in N_i^k}[\tau_{il}]^{\alpha}[\eta_{il}]^{\beta}},\ if\ j\ \in N_i^k, \qquad (3)$$

where $\eta_{il} = 1/d_{ij}$ and η_{il} is a heuristic value from point *i* to point *j* and d_{ij} is the length between point *i* and point j. τ_{il} is the concentration of pheromone on the path from i to j. p^k_{ij} is the possibility of ant *k* walking from point *i* to point *j*. N^k_i is the set of the neighbor point not including ant *k* walking to, and α and *β* are adjustable parameters. Every ant can remember all passed path and calculate a distance of the route. Finally, all ants can walk back to their nest and release Pheromone all the way.

After creating the route, all ants will update the quantity of pheromone during the way back to the nest called "Global Updating". Pheromones are volatile based on this follow formula:

$$\tau_{ij} \leftarrow (1-\rho)\tau_{ij}, \quad \forall(i,j) \in L \qquad (4)$$

where *ρ* is the rate of the evaporation and $0 < \rho \leq 1$, *L* is the set of point on the route. When *ρ* has high value, ants will forget the trodden path. But when *ρ* has low value, ants can stuck in the same path. The amount of pheromone is increased according to the formula below:

$$\tau_{ij} \leftarrow \tau_{ij} + \sum_{k=1}^{m}\Delta\tau_{ij}^{k}, \quad \forall(i,j) \in L \qquad (5)$$

where *m* is the number of all ants and τ^k_{ij} is the number of pheromone released from ants' bodies *k* on the path from *i* to *j*. The variable of $\Delta\tau^k_{ij}$ can be shown as follow:

$$\Delta\tau_{ij}^{k} = \begin{cases} 1/C^{k}, & if\ arc\ (i,j) belong\ to\ T^{k}, \\ 0, & otherwise. \end{cases} \qquad (6)$$

where T^k is the route created by ant k and C^k is the length of T^k. The increasing of pheromone is the reciprocal of the length of the route. Therefore, we can adjust the amount of pheromone to control the proportion finding the best route.

Next, Dorigo et al. had developed Ant System replacing by Ant Colony System. The difference is the finding of the route or the solution of Ant Colony System based on ants' experiences. Moreover, updating and the evaporation of pheromones are used only with the best route and the update can be appeared both global and local.

D. Finding Optimal LSB Substitution Using ACO

In 2010, Ching-Sheng Hsu and Shu-Fen Tu had presented the technique using ACO for finding matrix *A* to increase the efficiency of hiding data technique. Firstly, the secret message *S* will be accessed for changing the position of each pixel by Toral Automorphisms. With this process, we will have the secret message *S'*. In the system of data embedding on the cover image *H*, every bit of the secret message *S'* will be switched the position reference by matrix *A* from Ant Colony Algorithm (ACO). The first stage is to create an objective function for the substitution matrix. Measuring the quality of digital images can use PSNR (Peak Signal to Noise Ratio) or MSE (Mean Square Error). The objective function represents to the increasing of PSNR results or the decreasing of MSE measured results.

$$f(I') = \sum_{i=0}^{a-1}\sum_{j=0}^{b-1}(I_{ij} - I'_{ij})^2 \qquad (7)$$

The next step is to create problem graph. For example, in Figure 1: the problem graph of substitution matrix with the size of 4 x 4 . Each ant will walk through each node within the graph to create its own route like the ant leave the nest to forage. They have 4 different ways or nodes v_{00}, v_{10}, v_{20} and v_{30} . When ants choose the path assumed as v_{00} , ants will arrive to v_{00} then they have to decide whether going to any nodes. There will be only three nodes left : v_{11}, v_{21}, v_{31} because each row and column can be set only in one node. After repeating the process untill all ants walk to the food, every node which ants walk through will be set as 1 and the other nodes will be set as 0. And we will have the substitution matrix.

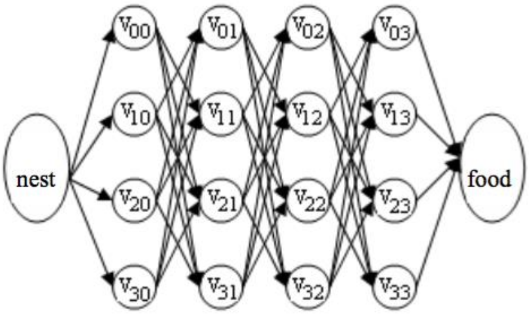


Figure 1. The problem graph.

Each parameter used for creating necessary route has to be determined by these follow details:

m: the number of all ants;
τ_0: the initial value of pheromones;
τ: the concentration of pheromones;
η: the experiences value;
α and β = the weight of the concentration of Pheromones and experience value, respectively;
T: the number of return trips;
ρ: the parameter of the evaporation of Pheromones;
q_0: the random number parameter.

In the part of creating the route, each ant will choose its own path according to the rule below:

$$\tau_{ij}(t+1) = (1-\rho)\times\tau_{ij}(t) + \rho\times\tau_0 \qquad (8)$$

The pattern of pheromones updating can be shown as follows:

$$\tau^{best}(t+1) \rightarrow (1-\rho)\times\tau^{best}(t) + \rho(\frac{1}{\min MSE}) \qquad (9)$$

When ants create their routes until one set of substitution matrix, the gained result will be used as the determination of value changing in each *r* bit of *S'*. Therefore, *S'* will be changed to *S** and we can embed *S** into *H* by taking the number of *r* bits (the righest *R*) from the cover image H, reaplcing *R* with *S** and then intergrating with *H-R*. After that we will have the stego-image *Z*, the receivers will take data *S* out off the stego-image in each pixel within the amount of *r* bits and the substitution matrix will switch the position back. *S'* from switching the position in each *r* bits by the substitution matrix, there will be decoding process again from the variable k_0 and k_1 by Toral Automorphisms. Finally, the recipient will receive the secret image.

Within the research, the result has been shown that the application of ACO in creating the substitution matrix for increasing the efficiency of hiding data by LSB technique use 21 rounds of walking and the substitution matrix can be created. The good results had shown up even though there will be a lot of possibilities of the paths.

III THE PROPOSED METHOD

This research is based on the basic of hiding image data that have been presented within the report of optimal least significant bit substitution technique (Wang et al., 2001) and the creation of the substitution matrix with ACO (Ching-Sheng and Shu-Fen, 2010) by adding the steps of position switching in each bit. This is to maximize the results of PSNR calculating. The details and processes can be shown as these referencing in Figure 2:

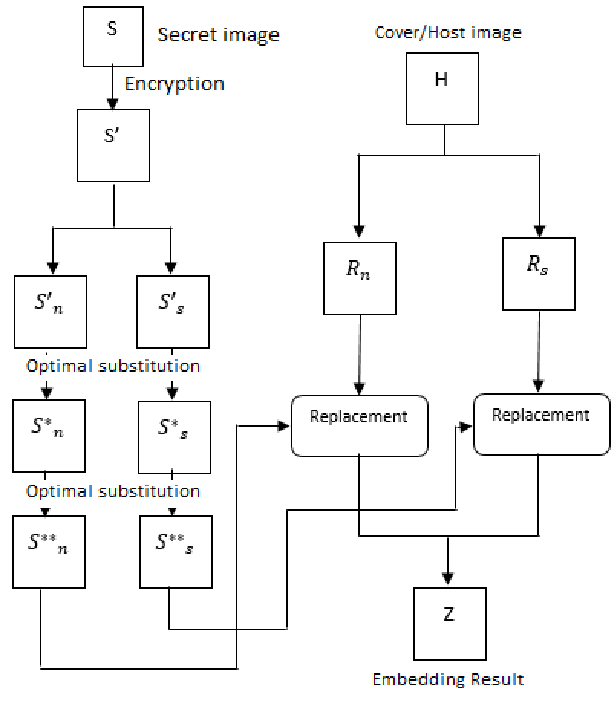


Figure 2. Flowchart of the proposed method.

A. The Embedding Process

Encryption. Switching the position of the secret image *S* in each *r* bit by Toral Automorphisms to increase the safety of data, *S'* will be shown up in this step.

Initial Data. Extract the data from *H* in the number of *r* bits in each pixel for comparing the result of MSE calculating to create the substitution matrix. *R* is divided into two parts: bit from noisy pixel replacing by R_n, smoothing pixel replacing by R_s. Noisy pixel will be embedded in the number of r = 4 bits and smoothing pixel will be embedded in the number of r = 2 bits.

Construct Substitution Matrix for Pixel. Create problem graph with ACO of both R_n and R_s by using the rule from (Ching-Sheng and Shu-Fen, 2010). When there is one set of round, substitution matrix will be shown up. The switching position of *r* bits of S'_n for replacing on R_n and S'_s replacing on $R_{s.}$

Construct Substitution Matrix for Bit. When switching *r* bits, there will be $S*_n$ and $S*_s$ for creating the problem graph with ACO again as in Eq. (3) by this time, each node of ants' paths will be the switching position of each bit of the secret image. After that $S*_n$ and $S*_s$ will be changed the positions by the substitution matrix gaining after $S**_n$ and $S**_{s.}$

Embedding. Replace $S**_n$ with R_n and $S**_s$ on R_s and integrate with *H-R* resulting to stego-image *Z*.

B. The Extraction Process

Extraction. Get $S**_n$ and $S**_s$ out from stego-image *Z*.

Recover Bit Position. Change the position of each bit of $S**_n$ and $S**_s$ with the substitution matrix to receive $S*_n$ and $S*_s$.

Recover Pixel Position. Change the position of each r bit of $S*_n$ and $S*_s$ with the substitution matrix to receive *S'*.

Decryption. Reverse the position of *S'* with the variable of k_0 and k_1 with Toral Automorphisms to receive the secret image.

IV EPERIMENTAL RESULTS

In this experimental, the researcher had compared hiding data with optimal least significant bit substitution technique combining with the creation of the substitution matrix with ACO, the technique presented in Section III. To add the steps of changing the position in each bit and simple LSB substitution with r = 4 bits by the secret image, four images are greyscale size of 256 x 256 pixels shown up in Figure 3 and the cover image of 3 images size 512 x 512 pixels shown up in Figure 4. After that the researcher had calculated the result from three techniques and the results had been shown that those presented techniques can maximize the result of PSNR calculating (Eq. (10)) comparing to the techniques presented in the research (Wang et al., 2001) and (Ching-Sheng and Shu-Fen, 2010) as shown in the Tables 1, 2 and 3.

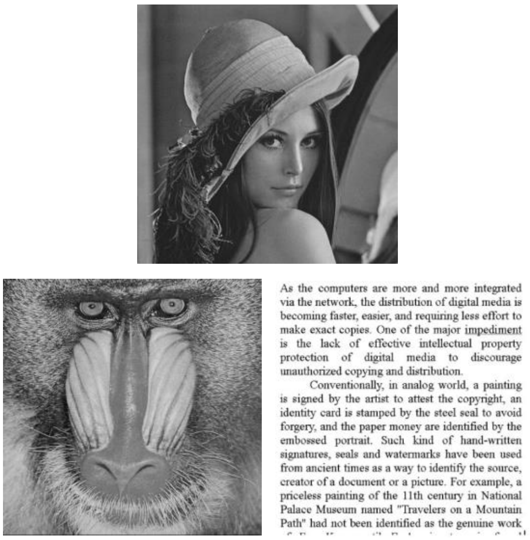
In Figure 5 had shown the comparison the results of the stego-image. After hiding data with presented techniques and simple LSB substitution, the result had found that presented techniques will not show up any noticed trace

The important concept from this history lesson is that communication does not have to occur over standard open channels using well-known methods. The Internet, in its massive, protocol-laden glory, is a playground for the modern steganographer. For example, think of an IP packet as the wax tablet previously mentioned. The packet's data field is equivalent to the writing in the wax. The headers serve as the wood in this analogy – who ever looks at an IP packet's headers, much less the data alignment padding? Most every protocol, language, and data format on the Internet has room for rent.

Uses of Steganography

Steganography is a means of storing information in a way that hides that information's existence. Paired

Figure 3. Secret images: greyscale image of 256 x 256 pixels.



As the computers are more and more integrated via the network, the distribution of digital media is becoming faster, easier, and requiring less effort to make exact copies. One of the major impediment is the lack of effective intellectual property protection of digital media to discourage unauthorized copying and distribution.

Conventionally, in analog world, a painting is signed by the artist to attest the copyright, an identity card is stamped by the steel seal to avoid forgery, and the paper money are identified by the embossed portrait. Such kind of hand-written signatures, seals and watermarks have been used from ancient times as a way to identify the source, creator of a document or a picture. For example, a priceless painting of the 11th century in National Palace Museum named "Travelers on a Mountain Path" had not been identified as the genuine work

Figure 4. Cover images: greyscale image of 512 x 512 pixels.

Table 1. The results of PSNR calculating from hiding image data from Figure 3 on the cover image 'Lena' from Figure 4.

Techniques	Deer	Zooey	Light-house	Text
Simple LSB Substitution	34.5914	34.6367	35.4246	32.9301
Optimal LSB Substitution	43.9888	43.6377	43.8728	42.9878
The Proposed Method	44.0192	44.0214	44.0714	43.0058

Table 2. The results of PSNR calculating from hiding image data from Figure 3 on the cover image 'Baboon' from Figure 4.

Techniques	Deer	Zooey	Light-house	Text
Simple LSB Substitution	34.6234	34.4546	35.315	33.0287
Optimal LSB Substitution	43.6087	43.2658	43.5488	42.4598
The Proposed Method	43.6461	43.645	43.7244	42.5201

Table 3. The results of PSNR calculating from hiding image data from Figure 3 on the cover image 'Text' from Figure 4.

Techniques	Deer	Zooey	Light-house	Text
Simple LSB Substitution	34.4181	31.960	33.6617	35.990
Optimal LSB Substitution	42.7773	41.931	42.5257	42.535
The Proposed Method	42.9093	42.433	42.6993	43.040

$$PSNR = 10 \times \log \frac{255^2}{MSE}, \qquad (10)$$

where

$$MSE = \frac{1}{M_1 \times M_2} \sum_{j=1}^{M_1} \sum_{j=1}^{M_2} (c_{ij} - c'_{ij})^2. \qquad (11)$$

Figure 5. Stego-image from simple LSB substitution (Left) and presented techniques (Right) with secret image 'Zooey'.

V CONCLUSION

The significant aim of hiding image data is that hidden data will not be noticeable. The proposed method within this research were based on the available data presented in the findings of optimal least significant bit substitution technique and the creation of the substitution matrix with ACO by adding the steps of changing position of each bit from pixel switching process. The result from the experiment has shown that proposed method can increase the quality of the optimal LSB substitution technique. It means that the safety of hiding data has been increase.

The proposed method is simple and not complicated. Thus, this technique can be integrated with other techniques which have been presented in the current and the quality of the stego-image will not much change. There will be better results and the safety will be maximized because of the various applied techniques. Encoding image data from others will be difficult because they cannot know what technique applied together in the created stego-image.

REFERENCES

Chang, C., Hsiao, J. and Chan, C. (2003). *Finding optimal least-significant-bit substitution in image hiding by dynamic programming strategy.* Pattern Recognition, 36(7), pp.1583-1595.

Ching-Sheng, H., & Shu-Fen, T. (2010). *Finding Optimal LSB Substitution Using Ant Colony Optimization Algorithm.* Paper presented at the Communication Software and Networks.

Das, S., Das, S., Bandopadhyay, B. and Sanyal, S. (2008). *Steganography and Staganalysis: Different Approaches.* Int. Journal of Computers, Information Technology and Engineering (IJCITAE), vol.2, no.1.

Dorigo, M. and Gambardella, L. (1997). *Ant colony system: a cooperative learning approach to the traveling salesman problem.* IEEE Transactions on Evolutionary Computation, 1(1), pp.53-66.

Dorigo, M., Maniezzo, V. and Colorni, A. (1996). *Ant system: optimization by a colony of cooperating agents.* IEEE Trans. Syst., Man, Cybern. B, 26(1), pp.29-41..

Dorigo, M. and Stützle, T. (2004). *Ant colony optimization.* Cambridge, Mass.: MIT Press.

Katzenbeisser, S. and Petitcolas, F. (2000). *Information hiding techniques for steganography and digital watermarking.* Boston: Artech House.

Kaur, G. and Kochhar, A. (2012). *A steganography implementation based on LSB & DCT.* International Journal for Science and Emerging Technologies with Latest Trends, vol. 4, no. 1, pp. 35- 41.

Merkle, D., Middendorf, M. and Schmeck, H. (2002). *Ant colony optimization for resource-constrained project scheduling.* IEEE Transactions on Evolutionary Computation, 6(4), pp.333-346.

Voyatzis, G. and Pitas, I. (1996). *Applications of toral automorphisms in image watermarking.* Proceedings of IEEE International Conference on Image Processing, vol. 2, 1996, pp. 237-240.

Wang, R., Lin, C. and Lin, J. (2001). *Image hiding by optimal LSB substitution and genetic algorithm.* Pattern Recognition, 34(3), pp.671-683.

Wayner, P. (2002). *Disappearing cryptography.* Amsterdam: MK/Morgan Kaufmann Publishers.

Yang C.-H. And Wang S.-J. (2010). *Transforming LSB Substitution for Image-based Steganography in Matching Algorithms.* Journal of Information Science and Engineering, vol. 26, pp. 1199-1212.

Zielinska, E., Mazurczyk, W. and Szczypiorski, K. (2014). *Development Trends in Steganography*, Commun. ACM 57(2).