Comparative Study of LSB and DWT watermarking Techniques Integrated with DNA Steganography for Digital Document Authentication

Leena Amol Deshmukh[1], Dr. Maithili Arjunwadkar[2]

[1]*Assistant professor, Progressive Education Society's Modern Institute of Business Studies (Autonomous), Nigdi, Pune, Maharashtra India*

[2]*Director, Progressive Education Society's Modern Institute of Business Studies (Autonomous), Nigdi, Pune, Maharashtra India*

***Abstract*—Steganography is used for secret communication and Digital watermark is used for copyright protection and information hiding. This paper proposes DNA steganography using insertion technique and watermark generation framework in which secret message is converted into binary and then DNA bases using DNA dictionary rule and embedded into randomly generated DNA sequence. This new DNA sequence is converted back to binary sequence using DNA dictionary rule and embedded into digital image using Least Significant Bit (LSB) and Discrete Wavelet Transform (DWT) watermarking techniques. The performance of both methods is evaluated using Mean Square Error (MSE), Peak Signal-to-Noise Ratio (PSNR), and payload capacity.**

***Index Terms*—DNA sequence, DNA Steganography, DWT Watermarking, LSB Watermarking, MSE, PSNR**

I. INTRODUCTION

Digital watermarking, Cryptography and Steganography is branch of information hiding. Digital multimedia is made up of text, images, video, animation in the digital format. Due to rapid growth in digital technologies, these digital contents can be easily modified and this leads to major concern to integrity and authentication of multimedia content. Hence copyright protection and content authentication is the biggest challenge. Digital Watermark and Steganography is the solution for this problem. Digital watermark is used to provide copyright protection and Steganography is used to conceal copyright information.

II. DIGITAL WATERMARKING

Digital watermark is the process of embedding identification information, logo, text or image into digital media like text, image, audio or video. This embedded data is known as watermarked data.

Digital watermarking is used in many applications such as Copyright protection Ownership verification, copy control, broadcast monitoring, Authentication of digital content and Tamper detection [1].

The process of watermarking is divided in two parts: - watermark embedding and extraction of watermark to prove authenticity and data integrity. Watermarked data is embedded inside cover medium using certain key. Cover medium can be text, image, audio or video. During extraction, watermarked cover medium is taken as input and algorithm key is used to extract watermarked data and original cover medium [2].

Digital watermark is classified based on visibility, Robustness, cover medium (also called as host) and domain. Digital watermark can be visible or invisible. After certain attacks on watermark, if watermarked data remains unchanged then it is called robust. Watermark can be robust, semi robust or fragile. Based on cover medium watermarked is classified as Text Watermarking, Image Watermarking, Audio Watermarking or Video Watermarking [3].

Based on domain, digital watermark can be classified as Spatial domain and Frequency domain [4].

In Spatial Domain watermarked data is embedded by directly changing pixel value. while embedding watermarked data in spatial domain, pixel values are changed. Actual image is not transformed to another domain. Least Significant Bit, Pixel Value Differencing and Edge based Watermarking are different techniques of spatial domain. Advantages of spatial domain techniques are they are very simple and easy to implement. Hence computational cost is also less. Spatial domain techniques provide comparatively high capacity to embed watermarked data. But these techniques are less resistant to image manipulation attack such as cropping, filtering and adding noise. In Frequency Domain, first image is transformed to frequency component then watermarked data is embedded in this frequency coefficient. In this watermarked data is not directly embedded into image. After inserting watermarked data, inverse transform is performing to reconstruct the image. Discrete Wavelet Transform, Discrete Cosine Transform, Singular value Decomposition are different techniques com under frequency domain image watermarking. Mathematical equation is required to transform image into frequency coefficient; hence implementation of these techniques ae very complex and computational cost is very high. But they are resistant to image manipulation attacks and are more robust [5].

III. LSB (LEAST SIGNIFICANT BIT)

It is spatial domain technique where bits of watermarked data are embedded by modifying least significant bit of a cover image. If any changes made to right most bit of pixel value, pixel intensity does not change drastically. This intensity is very small so it cannot be noticed by human eye. Hence LSB is simple technique and implementation is also very easy. Also, LSB has good payload capacity. The drawback of LSB watermarking is watermark can easily alter or removed. It is poor against image manipulation attack such as cropping, compression, noise addition. LSB watermarking is usually used for authentication, copyright protection [6].

IV. DWT (DISCRETE WAVELET TRANSFORM)

It is Frequency domain watermarking technique where watermarked data is embedded by modifying wavelet coefficient of an image. Int this watermarked data is not embedded directly into pixel value.

It divides image into sub sequent bands known as LL, LH, HL and HH bands.

a. LL (Low-Low) – Basic structure of an image. For example, smooth areas and brightness of an image (Approximation).
b. HH (High-High) – It contains fine details. For example, edges, shapes (Diagonal).
c. LH (Low-High) – Represents Horizontal edge information (Horizontal).
d. HL (High-Low) – Represents Vertical edge information (Vertical).

LL (Approx)	LH (Horiz)
HL (Vert)	HH (Diag)

Once image is divided into sub-bands, coefficient is modified to embed watermarked data. Inverse DWT is applied to reconstruct the image [7].

DWT provides high robustness against image processing attacks. Drawback of DWT watermarking is complexity. It is more complex and hence computational cost is very high.

DWT provides good imperceptivity and robustness. So, it is used for copyright protection, authentication [8].

V. STEGANOGRAPHY

It is Brach of information hiding. Two Greek word “stegnos” and “graphia” form word Steganography. Steganography is used to hide any confidential information inside other Text, Image, Audio or Video form. These types are known as digital media. Confidential information is concealed in such away intruder cannot recognize confidential information inside the cover media. Text Steganography, Image Steganography, audio and video Steganography are traditional types of Steganography based on the medium inside which confidential information is hidden [9].

The main aim of Steganography is not only concealing confidential information into digital media but also hide the fact that confidential information is hidden somewhere in the digital media [10].

Steganography is used in military for passing confidential information. Also, Steganography is act as invisible watermark as confidential information such as copyright information is concealed inside digital media and in case of any disputes this information is extracted and verified against claim. In this way Steganography is used in Authentication and copyright protection [11].

VI. DNA STEGANOGRAPHY

DNA steganography is modern technique used in data hiding where secret data is hidden inside DNA sequence. Instead of using traditional techniques like text or image which is act as cover object to hide secret message, DNA steganography uses DNA for concealing secret message. DNA stands for Deoxyribonucleic Acid. It consists of four nucleotide bases: A (Adenine), T (Thymine), C (Cytosine), G (Guanine) [12].
As length of DNA sequence is large, it can hide large amount of data. and receiver who knows exact decoding knowledge is able to detect secret message otherwise it is very difficult to extract it. DNA steganography provides high level security to secret message [13].

Techniques Used in DNA Steganography are [14]
a. Substitution Method –
It is used to hide secret message inside DNA sequence by substituting some DNA bases with bits of secret message. here secret data first converted to binary format and then using DNA dictionary rule it is converted into DNA sequence. Based on position selected secret message is replaced in original DNA Sequence and rest of DNA sequence is as it. At the time of extraction, receiver must be aware of position and DNA dictionary rule. Based on this secret message is retrieved. This is simple technique and in this length of DNA sequence is unchanged. Sender and receiver must agree on the position and DNA dictionary rule, otherwise secret message is wrongly retrieved. Also using this technique, small amount of secret data is hidden.

b. Insertion Method –
Instead of replacing some bases of original DNA, this technique inserts additional DNA bases to conceal secret message. In this secret message first converted to binary sequence and then using DNA dictionary rule, it is converted to DNA bases. These DNA bases are inserted at specific positions into original DNA sequence. Sender and receiver must know position where secret message DNA bases are inserted. At the receiver side, it receives increased length DNA sequence and then receiver who know exact position and DNA dictionary rule, extract secret message by removing it from exact position in DNA sequence and then convert back to text. This technique can conceal more data than substitute method. Drawback is length of DNA sequence increases.

c. Complementary Pairing –
It uses biological pairing rules means A is mapped to T and C is mapped to G. In this technique secret message is converted into binary and DNA sequence using DNA dictionary rule. Instead of storing this DNA sequence, this technique stores its complement (A-T, C-G). At the time of extraction, receiver again complement DNA sequence and convert back to binary and then text. In this technique DNA structure is not modified but encoding and decoding is complex as compared to other methods.

VII. CHARACTERISTICS AND PERFORMANCE PARAMETER FOR WATERMARK EVALUATION

After watermarked data is extracted, performance parameters are evaluated. Performance parameter is [15]

1. Robustness: Resistant to attack. Means if there are multiple attacks on watermark, there is no change to watermarked data.
2. Capacity: The maximum quantity of watermarked data can be embedded inside image
3. Imperceptivity: watermarked data is embedded in such a way it cannot disturb the quality of cover image. Watermarked data is embedded in such a way it should be unnoticeable to human eye. Imperceptivity is evaluated by calculating [16]

a. PSNR – PSNR means Peek Signal to Noise
It helps to measure how similar watermarked image is looked to original image. High PSNR means there is less distortion in original image and watermarked image. PSNR value above 30dB is consider as good quality. It is calculated using formula

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right)$$

b. MSE – MSE means Mean Squared Error

It measures average squared difference between original image and watermarked image. It calculates amount of distortion is introduced at the time of embedding watermarked data. Lower MSE shows watermarked image looks similar to original image and visual quality of an image is also good. MSE is calculated as –

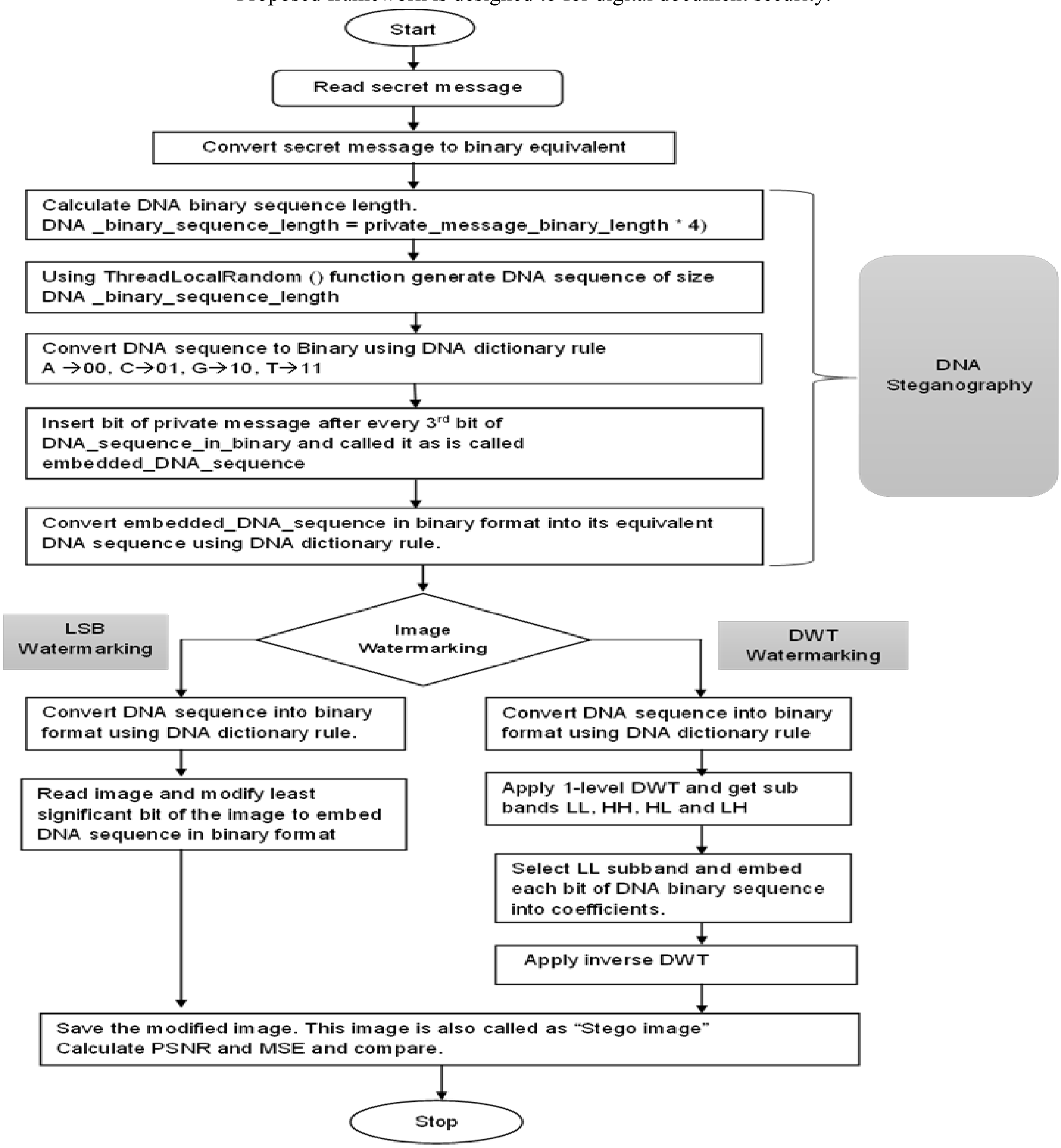
$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j) - I_w(i,j)]^2$$

where:

- M = number of rows in the image
- N = number of columns in the image
- $I(i,j)$ = pixel value of the original image
- $I_w(i,j)$ = pixel value of the watermarked image

VIII. PROPOSED WORK

Proposed framework is designed to for digital document security.



IX. EXPLANATION

In the proposed algorithm, DNA steganography is used to conceal secret message inside DNA sequence and then DNA sequence is embedded into image using either LSB watermarking or DWT watermarking.

Secret message which is to be kept hidden inside image is accepted from user and then it is converted into its binary equivalent. Then length of binary secret message is calculated.

Idea is to generate DNA sequence and convert it into binary equivalent. Then in the DNA binary sequence, use DNA insertion algorithm, where after every 3rd bit of DNA binary sequence, one bit of secret message is embedded. Once this step is done, we get large DNA sequence whose length is approximately equivalent to four times length of binary secret message.

Now this large DNA sequence is embedded into cover image using either LSB watermarking or DWT watermarking Technique.
During retrieval of secret message process, reverse process is applied.

PSNR and MSE is calculated using both techniques. Process is repeated with different length of secret message and both the watermarking techniques are compared based on capacity and robustness

X. EMBEDDING ALGORITHM

Step I: Private message embedding using DNA Steganography

a. Read private message to be concealed inside image
b. Convert private message to its equivalent binary and calculate length of binary message.
c. Calculate DNA binary sequence length.

DNA _binary_sequence_length = private_message_binary_length * 4

d. Using ThreadLocalRandom () function generate DNA sequence of size DNA _binary_sequence_length
e. Convert DNA sequence to Binary using DNA dictionary rule (A- 00, C – 01, T – 10, G – 11)
f. Insert bit of private message after every 3rd bit of _DNA_sequence_in_binary and resultant sequence is called embedded_DNA_sequence.
g. Convert embedded_DNA_sequence in binary format into its equivalent DNA sequence using DNA dictionary rule.

Step II: Embedding DNA sequence on image using watermarking

A. Using LSB watermarking

a. Convert DNA sequence into binary format using DNA dictionary rule.
b. Read image and modify least significant bit of the image to embed DNA sequence in binary format
c. Save the modified image. This image is also called as “Stego image”
d. Calculate PSNR and MSE

B. Using DWT watermarking

a. Convert DNA sequence into binary format using DNA dictionary rule.
b. Read image
c. Apply 1-level DWT and get sub bands LL, HH, HL and LH.
d. Select LL sub-band and embed each bit of DNA binary sequence into coefficients.
e. Apply inverse DWT
f. Save the modified image. This image is also called as “Stego image”
g. Calculate PSNR and MSE

XI. RESULT ANALYSIS

Implementation of above algorithm is done in Java Programming language

Three images of different sizes (256*256), (512*512), (1024*1024) and secret message of different length (100-character,200-character, 1000 character and 5000 characters) are given as input and evaluation is done on below parameters

1. PSNR
2. MSE
3. Payload capacity

Experimental result gives following table:

Based on PSNR and MSE

Image Size	msg Size (char)	LSB PSNR	DWT PSNR	LSB MSE	DWT MSE
256 * 256	100	51.13	39.14	0.50	7.93
256 * 256	500	51.08	39.14	0.50	7.93
256 * 256	1000	51.08	39.14	0.50	7.93
256 * 256	2000	51.03	39.14	0.51	7.93
512 * 512	100	55.28	39.19	0.19	7.81
512 * 512	500	55.11	39.19	0.20	7.82
512 * 512	1000	54.92	39.19	0.20	7.82
512 * 512	2000	54.60	39.16	0.22	7.82
1024*1024	100	61.32	43.59	0.04	2.84
1024*1024	500	61.15	43.43	0.04	2.94
1024*1024	1000	60.96	43.24	0.05	3.07
1024*1024	2000	60.60	42.90	0.05	3.33

Table 1: Comparison based on PSNR and MSE

The table shows LSB has Higher PSNR and lower MSE as compared to DWT watermarking.

Overall PSNR values in LSB watermarking is approximately between 51 to 61 and MSE is in between 0.04 to 0.50 whereas PSNR value of DWT watermarking is approximately between 39 to43 and MSE is between 2.9 to 7.8. This indicates for better imperceptivity LSB is good choice.

Based on Capacity
In the implemented algorithm for LSB watermarking, a watermarked bit is directly embedded into the least significant bit of an image. so, for LSB watermarking

LSB capacity = image width * image height
In DWT watermarking technique, watermark is embedded in every 2*2 image block

So

$$\text{DWT Capacity} = \frac{\text{Image width * image height}}{2*2} = \frac{\text{Image width * image height}}{4}$$

So, if we take ratio of LSB Capacity: DWT Capacity We get

LSB Capacity = 4 * DWT Capacity
That means proposed LSB watermarking technique embed four times more payload than proposed DWT watermarking Technique.

XII. CONCLUSION

Experimental result shows LSB watermarking archives higher PSNR (51 - 61) and lower MSE as compared to DWT watermarking. So, for high capacity and high imperceptivity applications, hybrid DNA -LSB watermarking is best choice. DNA steganography is used to hide secret message
By this additional security is provided as document.

REFERENCES

[1] M. Begum and M. S. Uddin, “Digital image watermarking techniques: A review,” Information, vol. 11, no. 2, p. 110, Feb. 2020, doi:10.3390/info11020110.

[2] R. Patel and P. Bhatt, “A review paper on digital watermarking and its techniques,” International Journal of Computer Applications, vol. 110, no. 1, pp. 10–13, 2015.

[3] E. Hussein and M. A. Belal, “Digital watermarking techniques, applications and attacks applied to digital media: A survey,” International Journal of Engineering Research & Technology, vol. 1, no. 7, pp. 1–8, 2012.

[4] S. Patel, A. Katharotiya, and M. Goyani, “A detailed study of digital image watermarking techniques,” in Proc. Int. Conf. Information Technology, Electronics and Communications (ICITEC), Hyderabad, India, Nov. 2011, pp. 81–86.

[5] W. Wan, J. Wang, Y. Zhang, J. Li, H. Yu, and J. Sun, “A comprehensive survey on robust image watermarking,” Neurocomputing, vol. 488, pp. 226–247, Jun. 2022, doi: 10.1016/j.neucom.2022.02.083.

[6] A. Mehto and N. Mehra, “Techniques of digital image watermarking: A review,” International Journal of Computer Applications, vol. 128, no. 9, pp. 21–33, 2015.

[7] V. B. Sankpal and R. N. Patil, “A review on different digital watermarking techniques,” International Journal of Innovative Science and Research Technology, vol. 2, no. 10, p. 458, 2017.

[8] D. Samiappan and K. Ammasi, “Robust image watermarking using discrete wavelet transform,” Journal of Computer Science, vol. 7, no. 1, pp. 1–5, 2011, doi: 10.3844/jcssp.2011.1.5.

[9] J. Kour and D. Verma, “Steganography techniques – a review paper,” International Journal of Emerging Research in Management & Technology, vol. 3, no. 5, pp. 132–135, 2014.

[10] M. A. Saleh, “Image steganography techniques – a review paper,” International Journal of Advanced Research in Computer and Communication Engineering, vol. 7, no. 9, pp. 52–58, 2018.

[11] P. Yadav and M. Dutta, “An overview of various steganographic domains and its applications,” International Journal of Engineering Trends and Technology, vol. 52, no. 3, pp. 137–141, 2017.

[12] G. Hamed, M. Marey, S. A. El-Sayed, and M. F. Tolba, “Comparative study for various DNA based steganography techniques with the essential conclusions about the future research,” in Proc. 2016 11th Int. Conf. Computer Engineering & Systems (ICCES), Cairo, Egypt, Dec. 2016, pp. 302–307, doi: 10.1109/ICCES.2016.7822003.

[13] M. A. Farahat, A. Abdo, and S. K. Kassim, “A systematic literature review of DNA-based steganography techniques: Research trends, data sets, methods, and frameworks,” in Lecture Notes in Networks and Systems, vol. 224, A. E. Hassanien et al., Eds., Singapore: Springer, 2022, pp. 495–505, doi: 10.1007/978-981-16-2275-5_31.

[14] O. H. Alhabeeb, F. Fauzi, and R. Sulaiman, “A review of modern DNA-based steganography approaches,” International Journal of Advanced Computer Science and Applications, vol. 12, no. 10, pp. 184–196, 2021, doi: 10.14569/IJACSA.2021.0121021.

[15] B. Singh and G. Kasana, “A review of digital watermarking techniques: Current trends, challenges and opportunities,” Web Intelligence, vol. 22, no. 4, pp. 523–553, 2024, doi: 10.3233/WEB-230280.

[16] N. Tiwari and S. Sharmila, “Digital watermarking applications, parameter measures and techniques,” International Journal of Computer Science and Network Security, vol. 17, no. 3, pp