Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

A Hybrid Approach for Data Hiding using Twofish Algorithm and Compression Steganography Techniques

Merlin K[1], Pradiksha S[2], Deepa Lakshimi B[3], Ramya G[4]

[1,2,3]Student, Dept of Information Technology, National Engineering College, Kovilpatti, India

[4]Asst. Professor, Dept of Information Technology, National Engineering College, Kovilpatti, India

E-mail: [1]2015036@nec.edu.in, [2]2015045@nec.edu.in, [3]2015021@nec.edu.in, [4]ramya-it@nec.edu.in

Abstract

Data compression and encryption are essential components of information security, facilitating efficient data handling, reduced storage requirements, and secure data transmission. The system presents a novel hybrid data compression algorithm that combines lossy and lossless compression techniques, along with Twofish cryptography. The hybrid approach leverages the strengths of different compression techniques. First, the Huffman coding algorithm is employed to compress textual data efficiently. Subsequently, the cover image undergoes lossy compression using the Discrete Wavelet Transform (DWT) technique. To fortify data security, the Twofish algorithm offers robust and high-level encryption. The encrypted data is then embedded into the compacted cover image using the least significant bit (LSB) technique in a steganographic manner. In the evaluation phase, the system's performance is assessed using key metrics, bits per pixel, mean squared error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM). Comparative analysis with existing methodologies demonstrates the superior performance and methodological efficiency of the system. The results indicate that the hybrid approach strikes a balance between compression efficiency and data security, enabling faster data transmission over limited bandwidth connections and effectively utilizing storage media.

Keywords: Cryptography, DWT, LSB, MSE, PSNR, SSIM, Steganography

DOI: https://doi.org/10.36548/jiip.2023.3.003
Received: 08.08.2023, received in revised form: 28.09.2023, accepted: 12.10.2023, published: 21.10.2023

1. Introduction

Cryptography plays a pivotal role in the cybersecurity domain by providing essential tools and techniques to secure sensitive data and communications. It involves the use of mathematical algorithms to convert plain text into unintelligible ciphertext, ensuring confidentiality, integrity, and authentication. In the cybersecurity context, cryptography is utilized to protect sensitive information during storage, processing, and transmission. Robust cryptographic methods, such as AES (Advanced Encryption Standard) and RSA (Rivest-Shamir Adleman) [1], are employed to safeguard data from unauthorized access and cyber-attacks.

Steganography plays a critical role in the cybersecurity domain as a covert communication technique for hiding sensitive information within files or media. Steganography conceals data within images, audio, videos, or other digital content, making it invisible to unauthorized eyes [4].In cybersecurity, steganography is used to create hidden channels for confidential communication, data exfiltration, or concealing malicious activities [6]. It provides an additional layer of secrecy, enabling threat actors to evade detection and bypass traditional security measures.

1.1 Basic Model [15]

The major purpose of image compression is to reduce the image size by removing unwanted parts, which allows for efficient storage and transmission of data. The algorithm utilizes a basic model for securely transmitting secret information from the sender to the receiver Figure 1. The secret message is first encrypted using Twofish cryptography for enhanced security. Then, both the secret message and the cover-image undergo separate compression processes.

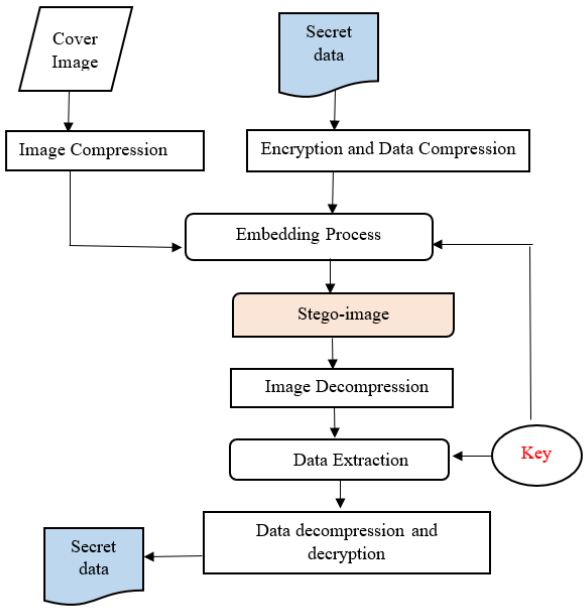

Figure 1. Block Diagram for Image and Data Compression

The algorithm utilizes a basic model for securely transmitting secret information from the sender to the receiver Figure 1. The secret message is first encrypted using Twofish cryptography for enhanced security. Then, both the secret message and the cover-image undergo separate compression processes. The secret message is compressed using Huffman's algorithm, and the cover-image is compressed using the Discrete Wavelet Transform DWT algorithm [1][7]. Next, the compressed cover-image is combined with the compressed secret message using the Least Significant Bit (LSB) technique [2]. The resulting encoded streams are sent over the Internet to the destination as a compressed file. At the receiver's end, a decoder is used to retrieve the final output image by decoding the received streams. The algorithm employs both lossy and lossless image decompression techniques to efficiently reduce the number of bits required to represent the image. This ensures that the transmitted data is compact and optimized for efficient storage and transmission. Overall, the approach provides a secure and efficient method for transmitting secret information in the form of compressed images.

1.2 Twofish Algorithm

Twofish is a robust symmetric key block cipher engineered for secure data encryption and decryption. Its operation centers around fixed-size data blocks, typically 128 bits in size, while offering flexibility in key lengths, supporting 128, 192, or 256-bit keys. The Twofish

algorithm comprises essential components such as key expansion, data whitening, and Feistel network rounds, which together bolster its capacity to provide formidable security measures for safeguarding sensitive information. The process begins with key expansion, where the original key is transformed and expanded into a set of round keys. The process involves applying the key schedule to generate multiple subkeys, which are used in the subsequent Feistel rounds.

The key schedule algorithm utilizes S-boxes, permutation functions, and bitwise operations to create a set of round keys that add complexity and confusion to the encryption process. Next, the data to be encrypted is divided into blocks, and each block undergoes data whitening [12]. This step involves a bitwise addition of the block with a subkey to ensure that the initial data transformation is non-linear and add an extra layer of security. The result is then split into four 32-bit words, forming two pairs of plain text and ciphertext. The core of the Twofish algorithm is the Feistel network, which consists of a fixed number of rounds, depending on the key length Figure 2.

Each round is divided into several steps, including the F-function, mixing, and key mixing. The F-function applies several operations, such as bitwise XOR, S-box substitution, and matrix multiplication, to introduce confusion and diffusion, making it highly resistant to attacks.

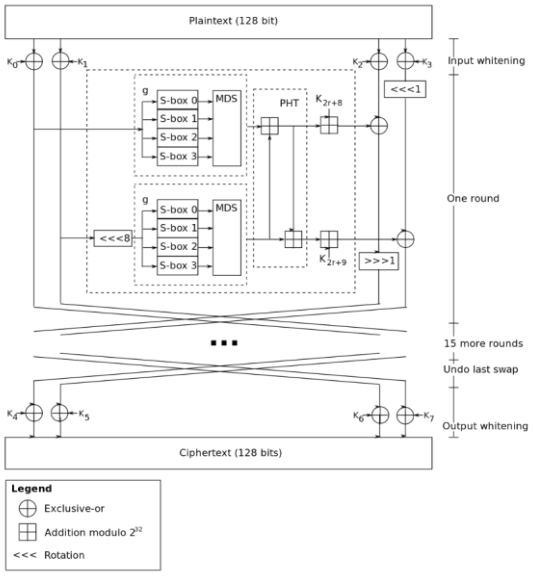

Figure 2. Twofish Algorithm

During each round, the two halves of the data are mixed using bitwise XOR and S-box substitutions, and then the subkeys are applied to each half. The mixing and key mixing operations further add confusion and prevent patterns from emerging in the cipher, enhancing its security. After completing all the rounds, the final ciphertext is obtained by reversing the initial data whitening process [13]. The ciphertext is now ready for secure transmission or storage, and the decryption process simply reverses the steps of the encryption process using the same round keys in reverse order.

Twofish is a robust and efficient symmetric key block cipher that employs key expansion, data whitening, and the Feistel network to provide strong encryption and decryption [14]. Its combination of confusion and diffusion operations makes it highly secure against various cryptographic attacks, making it suitable for protecting the data in a wide range of applications.

1.3 Discrete Wavelet Transform

The Discrete Wavelet Transform stands as a mathematical method employed in the realm of signal processing and data analysis. Its primary function lies in interpreting a provided signal or dataset into numerous constituents, each carrying distinct representations of various frequency ranges or dimensions [10].

Notably, this transformation is of a discrete nature, signifying its operation on a finite collection of data points. Within the framework of the DWT, an assortment of wavelet functions plays a pivotal role. These functions are characterized by their capacity to apprehend information at assorted resolutions or dimensions. Their primary application involves the partitioning the original signal, that results in the emergence of two fundamental categories of coefficients: detail coefficients and approximation coefficients [6][7]. In this division, the detail coefficients become carriers of high-frequency components and finer intricacies within the signal, while the approximation coefficients take on the responsibility of encapsulating low-frequency elements and broader features.

1.4 Huffman Coding

The importance of Huffman coding stems from its capability to achieve efficient data compression by exploiting the statistical properties of the input data. The coding technique has

widespread applications where reducing data size is crucial, such as in image and video compression, file archiving, and data transmission over networks. Huffman coding operates on the principle of constructing a binary tree-based codebook, wherein more frequently occurring symbols in the input data are represented by shorter binary codes, while less frequent symbols are assigned longer codes[7]. The adaptiveness allows Huffman coding to tailor its encoding to the specific statistical characteristics of the data, resulting in optimal compression ratios.

The primary advantage of Huffman coding is its ability to reduce data redundancy and utilize the information's entropy efficiently, achieving significant data compression without any loss of data integrity. As a result, Huffman coding is considered an optimal lossless compression technique. In image and video compression, where storage and bandwidth constraints are crucial, Huffman coding plays a pivotal role in reducing data size while preserving the data's originality. It is often combined with other compression techniques, such as DCT in JPEG or DWT in JPEG2000, to achieve even higher compression efficiencies.

1.5 Least Significant Bit (LSB)

The Least Significant Bit concept assumes a central role in the domain of binary representation, exerting a substantial influence in a wide array of fields, including digital signal processing, data secrecy, and encryption. Positioned as the rightmost bit within binary sequences, the LSB holds minimal sway in the overall computation of a binary number's value [11]. Exploiting the LSB emerges as a pivotal technique in the realm of data concealment, enabling the seamless incorporation of information into digital media, such as images or audio files, without compromising the underlying data's integrity [8]. This distinctive capability endows LSB-centric methods with indispensable utility in the domains of multimedia security and the preservation of intellectual property rights. Furthermore, in the field of cryptography, subtle manipulations of the LSB find utility, facilitating applications in steganography for covert communications and the augmentation of encryption protocols' resilience [9]. Thus, the versatility and significance of the LSB persist as crucial elements within the context of modern information security.

2. Related Work

Azizai.Hussein et al., [1] The scheme, A novel data hiding technique that combines RSA cryptography with lossless image compression steganography. The approach ensures secure data transmission through RSA encryption and utilizes the redundancy in the compressed image for efficient data embedding. DWT compression techniques compress the image for efficient process.

Y. Boarrios et al.,[2] The methodology involves developing a lossy compression algorithm specifically tailored for hyper spectral images. Hyper spectral images have high spectral resolution and contain a large amount of data, making compression crucial for efficient storage and transmission. The algorithm aims to achieve a good trade-off between compression ratio and image quality.

L.K. Tiong et al.,[3] The algorithm is adapted for implementation on resource-constrained wireless sensor nodes. The design takes into account the limited processing power, memory, and energy resources available in these nodes. The methodology includes a comparison of the new algorithm with existing compression techniques commonly used in wireless sensor networks. The comparative analysis helps establish the advantages and limitations of the approach over other methods.

R. Agrawal et al.,[4] A novel steganography scheme that utilizes JPEG compressed cover images to achieve high embedding capacity. The methodology likely includes specific techniques for embedding data in the frequency domain or spatial domain of the compressed image.

Sooyong Jeong et al.,[5] The solution is to the neural cryptography-based secret key exchange algorithm. These improvements include optimization techniques, parallelization methods, or reducing the number of iterations in the neural network training process. It also contains Quantization and puring technique. Quantization reduces the precision of neural network weights and activations, resulting in smaller memory footprint and faster inference. Additionally, pruning techniques can be applied to remove less important connections, further reducing model size and computational requirements.

E. Setyaningsih et al.,[6] Utilizes DWT. Embeds secret information within the audio signal using Spread Spectrum. Modulation to ensure imperceptibility and robustness which secures hidden data with RSA encryption, allowing authorized access with the appropriate private key. Implements compressive sampling to reduce data redundancy, enabling more efficient storage and transmission.

F. Adhanadi et al.,[7] Using Advanced Encryption Standard for secure data hiding. AES is primarily designed for encryption. The main idea behind Huffman coding is to represent frequently occurring symbols with shorter codes and less frequent symbols with longer codes, resulting in overall data compression.

C.A. Sari et al., [8] the study focuses on hiding data in images using steganography techniques along with compression algorithms. Implementing various steganography methods to embed data discreetly within the image, preserving its visual quality Utilizing compression techniques to reduce image size while retaining essential data Integrating steganography and compression for data hiding, enabling efficient storage and secure transmission of images.

A.I.Hussein et al., [9]Compression and encryption scheme combines compressive sensing and dynamic LSB embedding techniques to achieve high compression ratios while preserving visual quality and ensuring data security. Compressive sensing and dynamic LSB embedding for image compression and encryption.

Gan Zhihua b et al.,[10] The scheme introduces adaptability in DWT-SVD domain watermarking, adjusting the watermark embedding strength based on image characteristics. Watermarking is a technique used to protect digital content, including medical images and information, from unauthorized access and tampering. The improved DWT and Singular Value Decomposition (SVD) domain watermarking method focuses on enhancing the security and robustness of medical information in the digital domain.

Ashima Anand et al., [11] The research focuses on enhancing the security level of message blocks. The technique involves manipulating the position of digits within the message blocks to achieve increased resistance against cryptographic attacks and improve data protection. In this, the original message is first segmented into fixed-size blocks suitable for the cryptographic operation.

Bilal Al-Ahmad et al.,[12] In the methodology for an image enciphering application, the aim is to utilize an improved version of the Twofish algorithm for enhanced security and confidentiality. The process begins with generating a strong encryption key of appropriate length The employ of Cipher Block Chaining (CBC) mode, enhancing the algorithm's resistance against certain attacks by incorporating a randomly generated Initialization Vector (IV) for each block. The secret encryption key must be securely managed and stored, separated from the encrypted image. For decryption, the same encryption key and IV are used to reverse the process and retrieve the initial image.

Huma Jamil et al.,[13] The methodology employs Twofish, a robust 128-bit block cipher for symmetric encryption. A strong encryption key is generated, and data is padded to fit 128-bit blocks. Key expansion generates subkeys, and Twofish cipher operations are performed on each block. Feedback modes like CBC or CFB add complexity.

Fadil Muhammad a et al.,[14] The methodology analyzes the security system performance of MIPv6 during signaling using AES and Twofish encryption algorithms. Data collection, algorithm integration, and performance metrics evaluation are key steps. Cryptographic strength of AES and Twofish is analyzed, and a comparative study reveals their effectiveness in protecting sensitive data during MIPv6 signaling.

3. Proposed Work

The proposed algorithm combines Twofish and Huffman coding, along with DWT, to reduce the bit size of concealed information in steganography. This algorithm encompasses two primary processes which are information embedding and message extraction.

A) Embedding Process

The embedding process[15], in the context of steganography or data hiding, refers to the procedure by which hidden or secret data is concealed within a carrier medium, such as an image, audio file, or any other digital content [11] [12]. The goal of the embedding process is to make the presence of the hidden data as imperceptible as possible to human observers while ensuring that the hidden information can be extracted later when needed Figure 3.

- Initially, the confidential message undergoes encryption using the Twofish algorithm, incorporating rounds and keys for robust security.
- Following encryption, the subsequent step involves applying Huffman Coding to compress the concealed message efficiently.
- Simultaneously, the cover image undergoes decomposition through the Discrete Wavelet Transform (DWT), yielding four sub-bands: LL, LH, HL, and HH.
- Subsequently, the previously compressed encoded message image is incorporated into the chosen sub-bands (LH, HL, and HH), effectively blending it with the cover image.
- Employing Least Significant Bit (LSB) embedding, the cover image seamlessly integrates the encrypted secret message, culminating in the creation of a compressed stego-image.
- In the final stage, an array of metrics including Bits Per Pixel (BPP), Structural Similarity Index (SP), Mean Squared Error (MSE), and Peak Signal-to-Noise Ratio (PSNR) are computed to assess the effectiveness of the process.

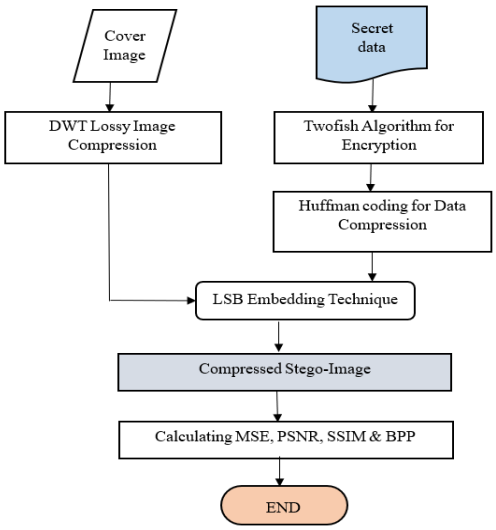

Figure 3. Embedding Process

B) Extracting Process

The extraction process[15], in the context of steganography or data hiding, refers to the procedure by which hidden or secret data is retrieved from a carrier medium, such as an image or audio file, after it has been previously embedded. The goal of the extraction process is to recover the concealed information while minimizing any noticeable impact on the carrier Figure 4.

- To retrieve the secret message, a sequence of steps must be meticulously executed:

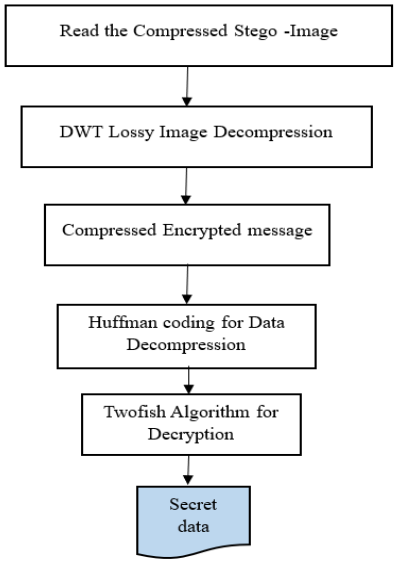

Figure 4. Extracting Process

- Begin by decompressing the previously compressed stego-image, employing Huffman Coding and the Discrete Wavelet Transform (DWT) to generate an uncompressed stego-image.

- Extract the concealed message from the designated sub-bands (LH, HL, and HH), generating a compressed encrypted message image.

- Progress to the next stage, where the Huffman tree file comes into play. Implement Huffman encoding to decompress the compressed encrypted message image, yielding an encrypted message image.

- Employ a Least Significant Bit (LSB) extraction technique from each pixel, facilitating the retrieval of stego-image binaries.
- Finally, utilize the Twofish algorithm instead of RSA for decryption, employing the appropriate decryption key to reveal the concealed message.

3.1. Measuring Compression Performances

1) Bits Per Pixel (BPP)

Bits Per Pixel serves as a measurement for quantifying the data or information necessary to depict each pixel within a digital image. It provides an assessment of the image's color depth and level of detail, commonly employed to characterize the image's overall data density or information richness.

BPP = No of bits in compressed image/total *x* number of pixels

2) Mean Squared Error (MSE)

This metric, denoted as the disparity between the compressed image data and the original, is a fundamental measure, as described in Equation below. Its primary purpose is to evaluate the fidelity of the compressed image. Ideally, this disparity should be minimized, and when it reaches 0, it signifies a high degree of resemblance between the compressed and original images, often referred to as a "lossless image compression technique."

$$MSE = \frac{1}{MxN}\sum_{X=1}^{M}\sum_{Y=1}^{N}(f(X,Y) - f'(X,Y))^2$$

3) Peak Signal To Noise Ratio(PSNR)

PSNR represents the relationship between signal strength and the presence of noise within a given signal, with its significance tied to the image's quality. In practical terms, a higher PSNR value corresponds to a superior image quality. This metric is contingent on the Mean Squared Error (MSE) of the chosen image, where a smaller disparity between the two images results in a higher PSNR value, indicating enhanced image quality.

$$PSNR = 10\log_{10}(\frac{MAX^2}{MSE})$$

4) Structural Similarity Index (SSIM)

SSIM quantifies the perceptual dissimilarity between two images that may initially appear identical. Rather than directly determining which image is "better," SSIM's primary function is to assess the degree of dissimilarity between the two images, allowing for inferences based on prior knowledge, such as distinguishing the "original" image from one that has undergone additional processes like data compression.

$$SSIM\ (x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_x\sigma_y + C_2)}{(\mu_x^2+\mu_y^2 + C_1)(\sigma_x^2+\sigma_y^2 + C_2)}$$

4. Results and Discussion

This methodology is executed through PyCharm software, operating on the Windows 11 platform. PyCharm is selected due to its proficiency in handling numerical computations and data analysis, along with its utility as a development tool for building applications. The user-friendly interface of PyCharm streamlines the processes of encryption and decryption, ensuring the accurate and effective management of information. PyCharm streamlines the processes of encryption and decryption, ensuring the accurate and effective management of information the table .1illustrates the comparison of the proposed method the existing methods.

Table 1. Comparison between the Proposed Method with the other Existing Methods.

Avg. Values	MSE	PSNR	SSIM	BBP
Pepper.png	3560.032	13.28708323	1.62452346	0.795
Lenna.png	4.0228019	43.0285758	1.439881	0.565
Cat.png	8266.21920	9.71813375	1.51050258	0.885

Figure 5. Input Images

Figure 6. Output Images

In the research, a subjective analysis of various images and employed both lossy and lossless compression techniques to represent the image outputs were conducted. Specifically, the color images in various formats such as jpeg, png, and bmp were considered. When utilizing Huffman coding, the resulting stego-image exhibits a gradual pixel alteration, preserving the visual appearance of the original cover image to the human eye Figure 5. As part of the study, a 128x128 pixel image was concealed within a 512x512 pixel canvas, resulting in a high-quality stego-image Figure 6.

Within this study, an assessment of several image quality metrics was conducted, including Table 1 Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), Bits Per Pixel (BPP), Saving Percentage and Structural Similarity Index (SSIM). These metrics were computed for a set of three images, and the corresponding results were analyzed and presented.

5. Conclusion

The hybrid techniques represent a comprehensive and powerful approach to secure data communication and storage. The encoding process efficiently conceals encrypted data within images, ensuring robust security against unauthorized access. The decoding functionality allows for the retrieval of hidden information, promoting seamless and secure communication between authorized parties.

As technology continues to evolve, this process is well-positioned to adapt and remain vital component of secure data management. It contributes to the broader field of information security by combining the strengths of cryptography and steganography, offering a reliable solution for real-time data protection and privacy.

References

[1] Wahab, Osama Fouad Abdel, Ashraf AM Khalaf, Aziza I. Hussein, and Hesham FA Hamed. "Hiding data using efficient combination of RSA cryptography, and compression steganography techniques." *IEEE access* 9 (2021): 31805-31815.

[2] Y. Boarrios, A. Rodríguez, A. Sáchez, A. Pérez, S. López, A. Otero, E. De La Torre, And R. Sarmiento, ‘‘Lossy hyper spectral image compression on a reconfigurable and fault-tolerant FPGA-based adaptive computing platform’’ Electronics, vol. 9, no. 10, p. 1576, 2020

[3] C. Chen, L. Zhang, and R. L. K. Tiong, ‘‘A new lossy compression algorithm for wireless sensor networks using Bayesian predictive coding,’’ Wireless Netw., vol. 26, no.8, pp. 5981–5995, Nov. 2020

[4] A. K. Pal, K. Naik and R. Agrawal ‘‘A steganography scheme on JPEG compressed cover image with high embedding capacity,’’ Int. Arab J. Inf. Technol., vol. 16, no. 1, pp. 116–124,2019.

[5] Kim, Juyoung, Sooyong Jeong, Dowon Hong, and Namsu Jho. "Improvement of the Efficiency of Neural Cryptography-based Secret Key Exchange Algorithm." IEEE Access (2023).

[6] E. Setyaningsih, R. Wardoyo, and A.K. Sari, ‘‘Securing color image transmission using compression-encryption model with dynamic key generator and efficient symmetric key distribution,’’ Digit. Commun. Netw., vol. 6, no.4, pp. 486–503, Nov 2020

[7] F. Adhanadi, L. Novamizanti, and G. Budiman ,‘‘DWT-SMM-based audio steganography with RSA encryption and compressive sampling,’’ Telkomnika, vol.18, no. 2, pp.1095–1104, 2020.

[8] C. A. Sari, G. Ardiansyah, and E.H. Rachmawanto ,‘‘An improved security and message capacity using AES and Huffman coding on image steganography,’’ Telkomnika, vol. 17, no. 5, pp.2400–2409, 2019.

[9] O. F. AbdelWahab, A. I. Hussein, H. F. A. Hamed, and H. M. Kelash, ‘‘Hiding data in images using steganography techniques with compression algorithms” Telkomnika, vol. 17, no. 3, pp.1168–1175, 2019

[10] Chai, Xiuli, Haiyang Wu, Zhihua Gan, Yushu Zhang, Yiran Chen, and Kent W. Nixon. "An efficient visually meaningful image compression and encryption scheme based on compressive sensing and dynamic LSB embedding." *Optics and Lasers in Engineering* 124 (2020): 105837.

[11] Anand, Ashima, and Amit Kumar Singh. "An improved DWT-SVD domain watermarking for medical information security." *Computer Communications* 152 (2020): 72-80.

[12] Al-Hyari, Abeer, Khaled Aldebei, Ziad A. Alqadi, and Bilal Al-Ahmad. "Rotation left digits to enhance the security level of message blocks cryptography." *IEEE Access* 10 (2022): 69388-69397.

[13] Haq, Tanveer Ul, Tariq Shah, Ghazanfar Farooq Siddiqui, Muhammad Zafar Iqbal, Ibrahim A. Hameed, and Huma Jamil. "Improved twofish algorithm: a digital image enciphering application." *IEEE Access* 9 (2021): 76518-76530.

[14] Supriyanto Praptodiyono a,1, Fadil Muhammad a, Dhandy Wiriyadinata a,” Analysis security system performance MIPv6 in signaling process using AES and Twofish algorithms”, Teknika: Jurnal Sains Dan Teknologi VOL 17 NO 02 (2021) 158-16

[15] Wahab, Osama Fouad Abdel, Ashraf AM Khalaf, Aziza I. Hussein, and Hesham FA Hamed. "Hiding data using efficient combination of RSA cryptography, and compression steganography techniques." *IEEE access* 9 (2021): 31805-31815.

Author's biography

Merlin K currently pursuing B Tech in Information Technology and a student of National Engineering College, Kovilpatti. Her research interests include cybersecurity, cryptography, Data science and web technologies and the experiences include web development and web applications.

Pradiksha S is a student at National Engineering college. She holds a highest degree B Tech at Information technology. Her research interests revolve around data compression techniques, image processing, and digital steganography fields and web applications.

Deepa Lakshimi B currently pursuing B Tech Information Technology and a student of National Engineering College, Kovilpatti. Her research interests are Cybersecurity, Cryptography, Machine Leaning and then Web Development and the experiences are like Working in FastApi and PowerBi.

Ramya G is Assistant Professor at National Engineering college. she has 1.5 years teaching experience she had completed her bachelors in Information Technology in Sivandhi Aadhithanar college of engineering and masters in Francis Xavier college of engineering. she is interested to work in the area of Networks and Cryptography.