Protection of Secret Textual Data Using Steganography Based System

Khalid Buryk S Alqarni

Bishah University, Faculty of computing and Information Technology, Department of Cybersecurity, Bishah City, Saudi Arabia

---***---

Abstract – *Recently, capabilities of attackers have been increased dramatically in terms of compromising secret data exchanged among users over the Internet. Using cryptographic approaches contribute to ensure confidentiality of data, but they do not provide secrecy of communication between the connected parties. Steganography can be used to strengthen cryptographic approaches by hiding secret data within innocent audios after passing an encryption stage. This adds second level of protection against attacks as well as hiding the communication itself between sender and receiver. This work proposed an enhanced steganography-based approach to protect secret data by embedding within audio files. Encryption using 3-DES is employed to create the first defense, while hiding in Least Significant Bit (LSB) creates the second defense with respect of ensuring high matching between cover file and stego file, high resistance against attacks, and high accuracy of retrieving the hidden data. The proposed approach achieves better performance in terms of PSNR, SSIM, and correlation.*

Key Words: Cover_object, Hidden_Data, Stego_key, Stego_object, LSB, Attacker.

1.INTRODUCTION

Steganography and the logic behind it. Steganography is the technique of hiding secret data within an ordinary, non-secret, file or message in order to avoid detection. The logic behind it is related to the fact that states "everything stored on the computer is converted into zeros and ones", and thus, it is possible to hide zeros and ones (secret data) within other zeros and ones (other files) [1].

Terms in steganography domain. Secret data, which is the data that requires protection against malicious actions, such as illegal modification and unauthorized access. Cover-Object, which is the file that hosts the secret data. Stego-Object, which is the file that is generated as an output of performing hiding process. Stego-Key, which is the key that is used for encrypting secret data [2].

Statement of problem. The problem is stated based on the general scenario of encryption, as shown in Figure-1.

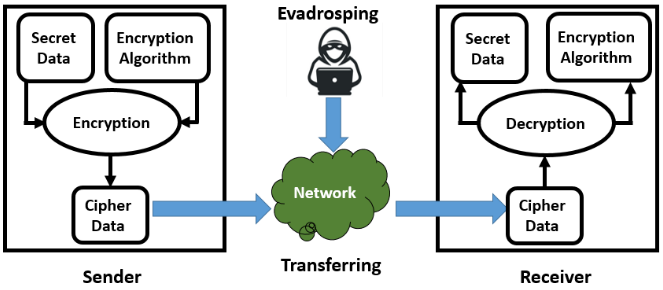


Fig -1: General scenario of encryption-based system.

Attackers can be located between the sender and receiver, where the malicious goal is obtaining cipher data through evadrosping the communication network. In this case, illegal modification can be applied easily, which in turn means losing of data integrity. Moreover, the communication between the sender and receiver is revealed.

Research questions. The research questions can be listed as follows:

1. How to ensure integrity of secret data?
2. How to hide the communication between the sender and receiver?

Contribution. This paper provides the following contribution in the field of steganography:

- Proposing two-level of protection using 3-DES encryption and hiding within audio files.
- Selection of hiding location based on ensuring lowest level of distortion caused by hiding process.
- Providing extensive experiments to validate the proposed approach along with comparison with similar works.

Organization of paper. The paper is structured so that section 1 provides related work. Section 2 presents the proposed approach. Section 3 shows the experimental results. Finally, the paper is concluded in section 4.

2. RELATED WORK

This section provides the state of the art. Systems of steganography can be classified based on the type of cover, as shown in Figure 2.

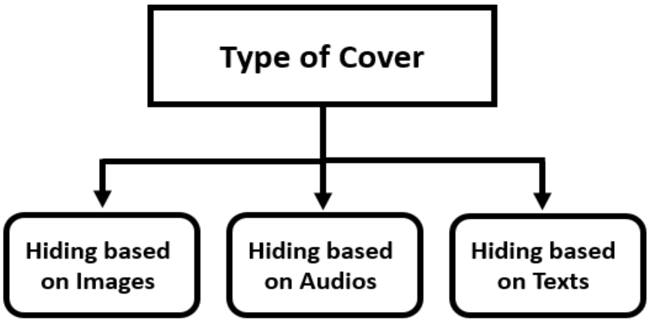


Fig -2: Classification of steganography-based system.

2.1 Hiding based on images

The problem that authors of [3] handled is related to small size of payload (hidden secret data) that can be inserted in the cover file. The proposed system mainly depends on spreading bits of secret data on all pixels of image cover not only on LSB. they used pixel indicator technique for randomizing. Protecting secret data based on one level only is the main issue that the researchers of the work [4] targeted. They strengthen steganography-based level of protection by encryption level using AES algorithm. However, the basic disadvantage is related to high complexity in the process of extraction. In frequency domain, an approach is presented in the work [5], where it uses DCT transformation to hide the secret information in the high frequency coefficients for the purpose of achieving stego-image of high quality.

2.2 Hiding based on audios

In [6] a dual-protection approach is provided, where a level of encryption for hiding data is created based on ASCII coding before inserting the secret message within the signal of audio file. Although [7] achieved steganography depending on LSB, its final objective was providing testing environment for steganography systems that use audio files as covers. The experiments conducted within the proposed environment proved that LSB technique was weak against rotation and compression attacks. In addition, LSB is simple in terms of implementation and may include some code-based vulnerabilities. Relying on modified Vigeneve cipher algorithm, authors in work [8] took into consideration the importance of adding a level of security to protect secret data. However, they only applied this approach on textual data ignoring other types of data. The system proposed in the work [9] starts to convert the audio file into frequencies in wave domain using DCT transformation. Then, secret data is encrypted using Huffman encryption to be embedded in converted audio file. Finally, embedding is performed. The extraction phase is performed using IDCT transformation.

2.3 Hiding based on texts

Using Hidi text as cover, a steganography-based system is proposed in [10]. The key idea behind this system is to replace the first letter of each word by a letter from secret message. Special feature of this system is related to convert numbers in cover into text to increase the space of hiding. Similarly, [11] provided an approach to embed a secret information within Arabic and Persian text file depending on several intersections between these two languages. In their proposed work [12], researchers presented a scheme based on cryptography using SSCE, where the secret key will be directly exchanged between the sender and receiver for both hiding and extracting a secret message using RSA algorithm.

3. PROPOSED SYSTEM

The proposed system consists of two main phases, which are hiding phase and extraction phase. Figure 3 provides the structure of the proposed system.

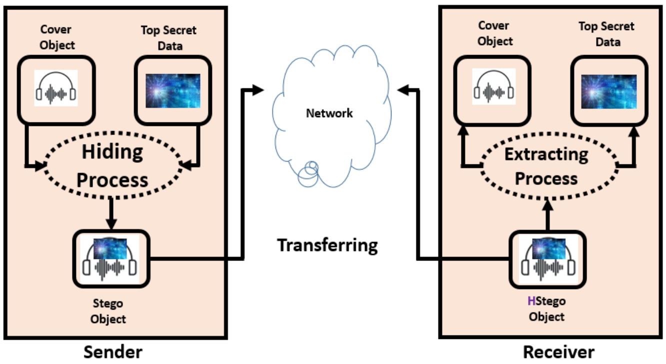


Fig -3: Structure of proposed system.

3.1 Actions at sender side

The main action performed at the sender side is hiding. It is expressed according to the following formula:

$$(\text{Cover\_Object}) + (\text{Hidden\_Data}) = \text{Stego\_Object} \quad (1)$$

To achieve double protection and to answer the first research question, 3-DES encryption algorithm is employed. The secret data is encrypted first before performing the hiding process. As shown in Figure 4.

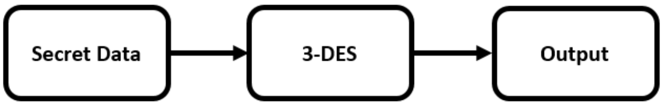


Fig -4: First level of protection.

3-DES algorithm works as follows: Triple DES operates in three steps: Encrypt-Decrypt-Encrypt (EDE). It works by taking three 56-bit keys (K1, K2 and K3) known as a key bundle and encrypting first with K1, decrypting next with K2 and encrypting a last time with K3. A Triple DES two-key version exists, where the same algorithm runs three times but K1 is used for the first and last steps [13]. Therefore, K1 will be the master key.

To answer the second research question, Least Significant Bit (LSB) technique is utilized. The key idea behind the LBS technique is to hide the bits of secret data within the first bit of each sample audio. The logic behind this based on the fact that states "the start point of an audio includes some silence spots during recording". This means that it is a suitable location for hiding since it leads to lowest level of distortion. LBS represents the second level of protection. Figure 5 shows where LSB bits are located within a given audio file.

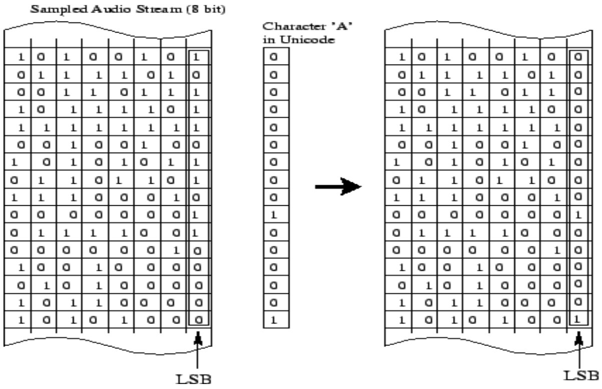


Fig -5: Location of LSB bits within an audio file.

As shown in Figure 5, LSB located in an array that weights the least when compared to other locations, which in turn means that if LBS bits are replaced or modified the damage will be minimum. In contrast, any alternation in Most Significant Bit (MSB) leads to maximum damage or distortion.

After determining the location of hiding which ensures minimum distortion, hiding process starts as described below. Figure 6 illustrates the first two steps.

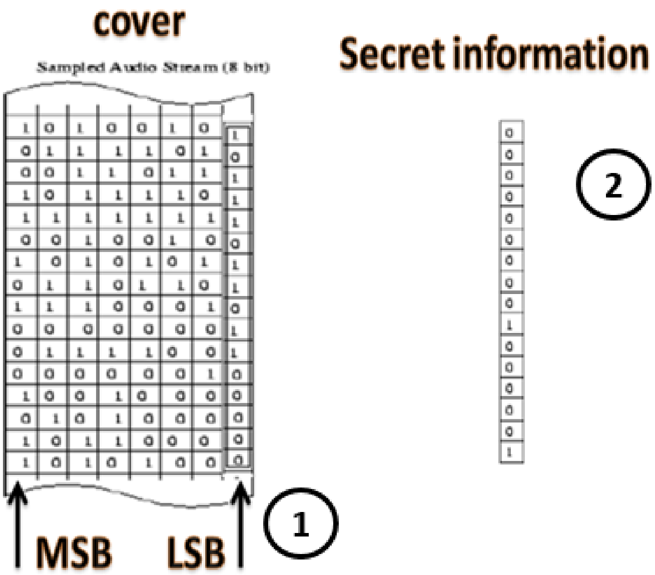


Fig -6: First two steps in hiding process.

The first step is to define the LBS column within the audio cover file. This is done by analyzing the audio file after converting it to binary mod. The second step is to obtain the bits of the secret data. This can be done by some skills in programming or using some tools available on the Internet.

To achieve double level of protection, bits of secret data are encrypted before hiding. However, attackers can obtain the LSB column and modify it, which means loosing integrity. To solve this problem, a randomization step is added, as shown in Figure 7.

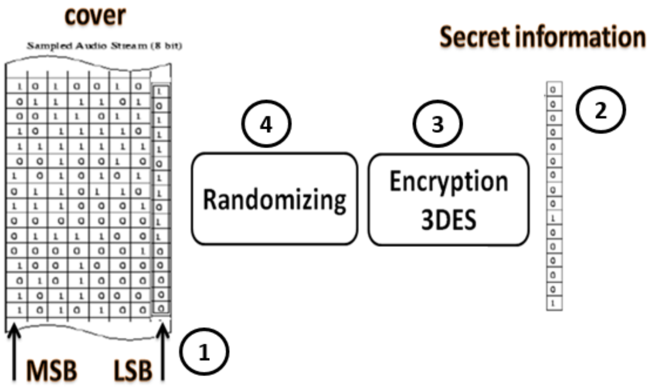


Fig -7: Third and fourth steps in hiding process.

Here, even the attacker tries to obtain LSB, he will face a problem of reconstructing (or reforming) LSB in the correct order. In other words, he needs to know the algorithm of randomization.

3-DES algorithm has a key and also randomization algorithm. There is a need for safely exchange the keys between sender and receiver. In this work, the same key used for encryption is also used as a seed of the randomization algorithm. RSA algorithm is employed for the purpose of safely exchange the key (called Stego-key). RSA algorithm is performed based on four steps (supported by Java statements) [14]. As described below.

RSA algorithm

Step 1: Construction the RSA model

The process starts with selection of two prime numbers (R and O), and then calculating their product W, as:

$$W=R*O \qquad (2)$$

Step 2: Derived Number (S)

Consider number (S) as a derived number that satisfies the condition > 1 and less than (R-1) and (O-1). The primary condition will be that there should be no common factor of (R-1) and (O-1) except 1

Step 3: Public key

The specified pair of numbers (W) and (S) forms the RSA public key and it is made public.

Step 4: Private Key

Private Key V is calculated from the numbers R, O and S. The mathematical formula that adjusts the numbers is as follows:

$$SV = 1 \bmod (R-1)\,(O-1) \qquad (3)$$

Encryption and decryption using RSA

Encryption Formula:

Consider a sender who sends the plain\clear\readable text MSG to someone whose public key is (W, S). To encrypt the plain text MSG, we rely on the following expression:

$$CPH = RS \bmod W \qquad (4)$$

Decryption Formula:

The decryption process is very straightforward and includes analytics for calculation in a systematic approach. Receiver (K) has the private key , the result modulus will be calculated as:

$$Plaintext = K \bmod W \qquad (5)$$

Figure 8 shows the whole process of hiding that achieves two level of protection.

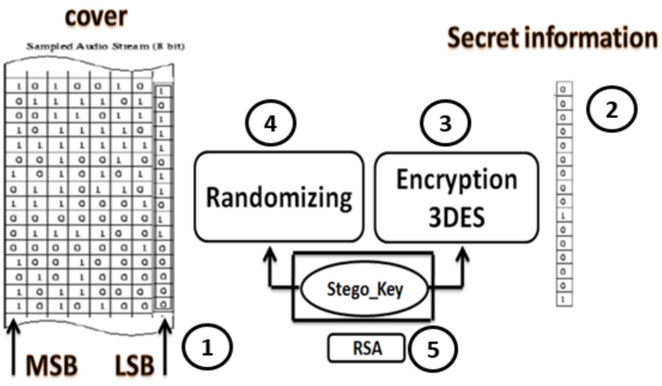


Fig -8: Comprehensive hiding process.

From mathematical point of view, hiding process is expressed by the following formula:

$$Cover\_Object+Hidden\_Data+Stego\_Key=Stego\_Object \qquad (6)$$

3.1 Actions at receiver side

At the receiver side, extraction process is performed. This includes extracting LSB firstly. Then, revising the randomization algorithm to reconstruct the correct LSB. Finally, the constructed LSB is decrypted using Stego-key that is safely exchanged using RSA algorithm. Extraction process is expressed by the following formula from mathematical perspective:

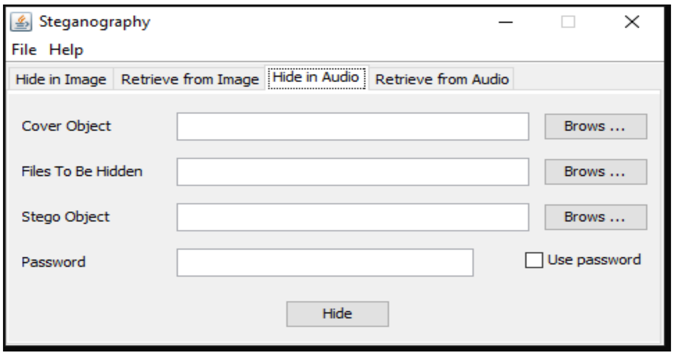
$$Stego\_Object + Stego\_Key = Hidden\_Data \qquad (7)$$

4. Experimental results

This section provides the results based on some metrics and along with comparison with some similar works.

4.1 Setup

The proposed steganography system is implemented using Java programming language. The following screen shows the main interface of the system.

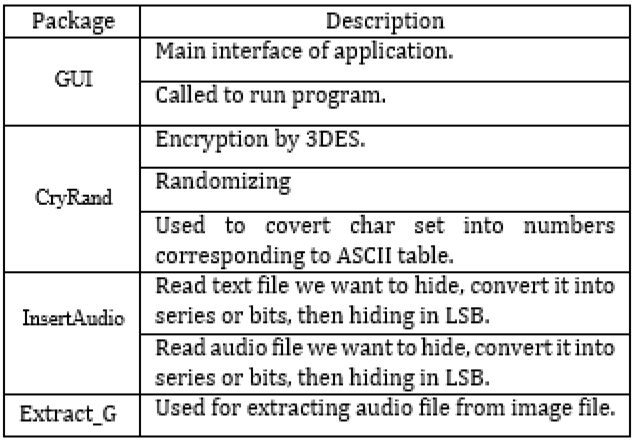


Cover Object
Files To Be Hidden
Stego Object
Password
□ Use password

Fig -9: The main interface of the proposed steganography based system.

From programing point of view, Table 1 summarizes the used classes.

Table -1: Used classes.



Package	Description
GUI	Main interface of application.
	Called to run program.
CryRand	Encryption by 3DES.
	Randomizing
	Used to covert char set into numbers corresponding to ASCII table.
InsertAudio	Read text file we want to hide, convert it into series or bits, then hiding in LSB.
	Read audio file we want to hide, convert it into series or bits, then hiding in LSB.
Extract_G	Used for extracting audio file from image file.

4.2 Security metrics

Three metrics are used for the purpose of evaluating the proposed system, as illustrated in Figure 10.

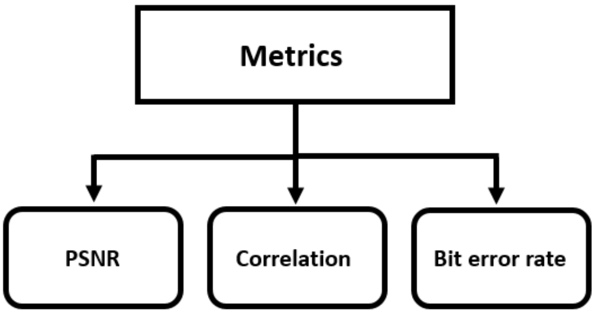


Fig -10: Used metrics.

PSNR [15] measures the Peak Signal to Noise Ratio. It needs to calculate MSE. It is given by the following formula:

$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(f_{ij} - g_{ij})^2$$

$$PSNRindB = 10Log_{10}\frac{L^2}{MSE} \quad (8)$$

To measure the quality of our proposed system, correlation [16] is used, which measures the similarity between the cover and stego_object. It is given by the following formula:

$$corr = \frac{ammoun - of - change}{all - size - of - audio - file} \quad (9)$$

Bit Error Rate [17] which calculates the amount of noise caused by hiding. It is given by the following formula:

$$BER = \frac{number\_of\_errors}{total\_number\_of\_bits\_send} \quad (10)$$

4.3 Results

We used song audio file (downloaded from the Interne) to represent the cover. To ensure covering all types of secret data, a text file is written manually (containing 5 lines taken from this article), another song audio file, and an image taken by mobile phone. Table 2 provides the obtained results.

Table -2: Results.

Stego_object	Original audio		
	PSNR	Correlation	Bit error rate
Stego_T	59.582	97.52	0.443
Setgo_G	58.988	81.68	0.779
Stego_A	58.246	65.98	0.841

As shown in Table 2, the lowest BER is generated when hiding secret data as texts. Actually, most type of secret data is textual form. To evaluate the ability of hiding of the proposed system without increasing the level of BER dramatically, the text file of the secret data is duplicated. Table 3 summarizes the results.

Table -3: Results under duplicated text file of secret data.

Text file order	Size in byte	PSNR	Correlation	Bit error rate
1	150	60.582	98.52	0.453
2	300	58.989	98.11	0.389
3	450	57.658	95.26	0.352
4	600	56.472	94.89	0.324

Table 2 shows that there is a small amount of increasing in the BER. This means that the proposed system is suitable for hiding secret data of textual form.

Visually, the audios (cover and Stego-object) are evaluated based on the wave form, as shown in Figure 11.

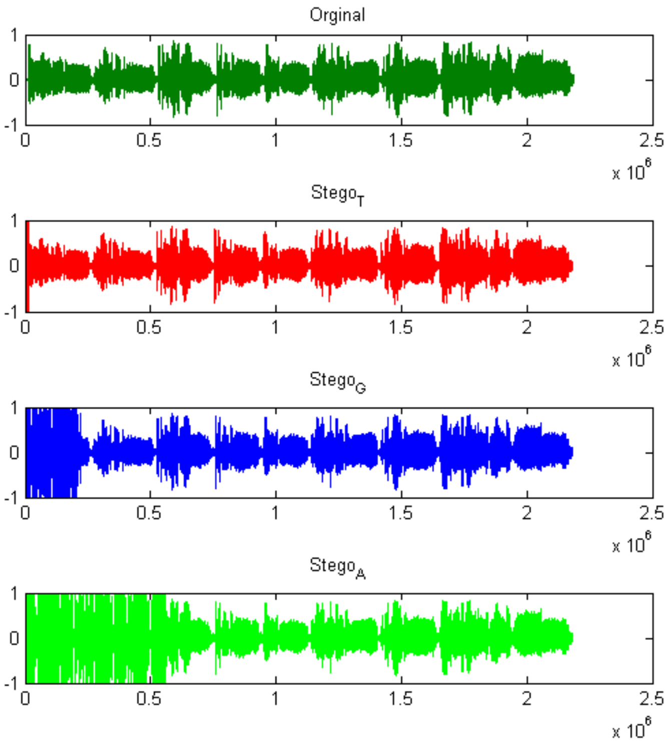


Fig -11: Wave forms.

Figure 11 shows that there is a small change in the wave of original audio (cover) and the Stego-object when hiding textual secret data. This supports the results obtained and summarized in Table 3.

5. CONCLUSION

In this work, a double-level protection technique is proposed using encryption and steganography. The first line of defence is represented by encryption of secret data before hiding. The second line of defence is represented by hiding process. An additional confusion is added to face the ability of attacker when it comes to talking about obtaining the LSB bits from the Stego-object. Results showed that the proposed system provides lower level of BER and it is suitable for hiding textual secret data.

As a future work, the proposed system will be enhanced based on deep learning techniques to discover the best location of hiding.

REFERENCES

[1] Chanu, Yambem Jina, Themrichon Tuithung, and Kh Manglem Singh. "A short survey on image steganography and steganalysis techniques." *2012 3rd National Conference on Emerging Trends and Applications in Computer Science*. IEEE, 2012.

[2] Hashim, Mohammed, MOHD SHAFRY MOHD RAHIM, and ALI ABDULRAHEEM ALWAN. "A REVIEW AND OPEN ISSUES OF MULTIFARIOUS IMAGE STEGANOGRAPHY TECHNIQUES IN SPATIAL DOMAIN." Journal of Theoretical & Applied Information Technology 96.4 (2018).

[3] Pan, Yi-Lun, and Ja-Ling Wu. "Rate-Distortion-Based Stego: A Large-Capacity Secure Steganography Scheme for Hiding Digital Images." Entropy 24.7 (2022): 982.

[4] Sakshi, Sakshi, et al. "Least Significant Bit Steganography for Text and Image hiding." 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM). IEEE, 2022.

[5] Xu, Youmin, et al. "Robust invertible image steganography." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

[6] Abdulkadhim, Hussein Abdulameer, and Jinan Nsaif Shehab. "Audio steganography based on least significant bits algorithm with 4D grid multi-wing hyper-chaotic system." International Journal of Electrical and Computer Engineering (IJECE) 12.1 (2022): 320-330.

[7] Mahmoud, Mahmoud M., and Huwaida T. Elshoush. "Enhancing LSB Using Binary Message Size Encoding for High Capacity, Transparent and Secure Audio Steganography–An Innovative Approach." IEEE Access 10 (2022): 29954-29971.

[8] Geleta, Margarita, et al. "Pixinwav: Residual steganography for hiding pixels in audio." ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.

[9] Abood, Enas Wahab, et al. "Audio steganography with enhanced LSB method for securing encrypted text with bit cycling." Bulletin of Electrical Engineering and Informatics 11.1 (2022): 185-194.

[10] Thabit, Reema, et al. "CSNTSteg: Color spacing normalization text steganography model to improve capacity and invisibility of hidden data." IEEE Access 10 (2022): 65439-65458.

[11] Osman, Omnia Mohammed, et al. "Hybrid multistage framework for data manipulation by combining cryptography and steganography." Bulletin of Electrical Engineering and Informatics 11.1 (2022): 327-335.

[12] Roy, Sangita, and Manini Manasmita. "A novel approach to format based text steganography." proceedings of the 2011 international conference on communication, Computing & Security. 2011.

[13] Ardiansyah, Giovani, Christy Atika Sari, and Eko Hari Rachmawanto. "Hybrid method using 3-DES, DWT and LSB for secure image steganography algorithm." 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE). IEEE, 2017.

[14] Zhou, Xin, and Xiaofei Tang. "Research and implementation of RSA algorithm for encryption and decryption." Proceedings of 2011 6th international forum on strategic technology. Vol. 2. IEEE, 2011.

[15] Sara, Umme, Morium Akter, and Mohammad Shorif Uddin. "Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study." Journal of Computer and Communications 7.3 (2019): 8-18.

[16] Kasetty, Praveen Kumar, and Aniruddha Kanhe. "Covert speech communication through audio steganography using DWT and SVD." 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2020.

[17] AlSabhany, Ahmed A., et al. "Digital audio steganography: Systematic review, classification, and analysis of the current state of the art." Computer Science Review 38 (2020): 100316.

BIOGRAPHY

Khalid: received the B.S. degree in computer science from Bishah University in 2016, and the M.S. degrees in cybersecurity from Bishah University in 2022. His current research interests include information security, IoT, artificial intelligence and deep learning.