UDC 003
DOI: 10.15587/1729-4061.2022.263695

IDENTIFYING OPTIMAL MESSAGE EMBEDDING LOCATION IN AUDIO STEGANOGRAPHY USING GENERATIVE ADVERSARIAL NETWORKS

Muatamed Hajer
Doctor of Information Technology
Department of Computer Science
University of Sumer
Rifai str., 1, Thi_Qar, Iraq, 57009
Mohammed Anbar
Corresponding author
Doctor of Advanced Internet
Security and Monitoring
Department of National Advanced IPv6 Centre
Universiti Sains Malaysia
Gelugor str., 12, Penang, Malaysia, 11800
E-mail: anbar@usm.my

Audio steganography (AS) uses the auditory redundancy of the human ear to conceal the hidden message inside the audio track. In recent studies, deep learning-based steganalysis has swiftly revealed AS by extracting high-dimensional stego acoustic features for categorization. There is still an opportunity for improvement in the current audio steganography required for managing communication confidentiality, access control and data protection. The main objective of this research is to improving the data protection by identifying the data embedding location in the audio. Generative Adversarial Network-based Audio Steganography Framework (GAN-ASF) is presented in this study, and it can automatically learn to provide better cover audio for message embedding. The suggested framework's training architecture comprises a generator, a discriminator, and a steganalyzer learned using deep learning. The Least Significant Bit Matching (LSBM) message embedding technique encrypts the secret message into the steganographic cover audio, which is then forwarded to a trained steganalyzer for misinterpretation as cover audio. After performing the training, stenographic cover audio has been generated for encoding the secret message. Here, Markov model of co-frequency sub images to generate the best cover frequency sub-image to locate an image's hidden payload. Steganographic cover audio created by GAN-ASF has been tested and found to be of excellent quality for embedding messages. The suggested method's detection accuracy is lower than that of the most current state-of-the-art deep learning-based steganalysis. This payload placement approach has considerably increased stego locations' accuracy in low frequencies. The test results GAN-ASF achieves a performance ratio of 94.5 %, accuracy ratio of 96.2 %, an error rate of 15.7 %, SNR 24.3 %, and an efficiency ratio of 94.8 % compared to other methods

Keywords: generative adversarial network, least significant bit matching, markov model, audio steganography, jpeg image

Received date 01.07.2022
Accepted date 24.08.2022
Published date 31.08.2022

How to Cite: *Hajer, M., Anbar, M. (2022). Identifying optimal message embedding location in audio steganography using generative adversarial networks. Eastern-European Journal of Enterprise Technologies, 4 (9 (118)), 59–68. doi: https://doi.org/10.15587/1729-4061.2022.263695*

1. Introduction

Steganography hides the payload, which is the substance of a secret chat, using redundant sections of multimedia data, such as digital photos, audio, music, and messages [1]. Several steganographic algorithms have been suggested over the last several decades using various forms of media as the cover [2]. Steganalysis techniques have also been suggested to identify the stego object, which is consistent [3]. It is more common for an investigator to be interested in discovering the concealed information and discriminating between the cover and stego things [4]. The stego key space, the stego locations, and the selection process for the stego places are only some additional data needed to extract hidden information [5].

An encoded message can be disguised in a cover, such as a clip, image or audio recording [6]. Covert communication might be made possible through the stego, a covert data storage device [7]. A wide range of multimedia security situations, such as privacy protection, has been aided by steganography's use [8]. Non-adaptive and adaptive steganography are two subcategories of steganography that vary in their embedding methodologies [9]. Non-adaptive steganography technologies, on the other hand, tend to alter every part of the cover equally. LSB and LSB Matching (LSBM) are notable examples of this work [10]. As a result, adaptive steganography employs various techniques to hide the hidden message [11].

Algorithm to locate the payload in JPEG-compressed spatial images by LSB matching, which recompresses and decompresses the compressed versions to estimate cover images [12]. Multiple Least Significant Bit (MLSB) steganography's characteristics were studied, and a payload localization technique and stego key recovery procedure was suggested using the best stego subset [13]. Because of their great accuracy, the techniques above may be used to find the concealed payload in various types of encryption, such as LSB matching or MLSB replacement, and even to estimate the number of groups in group parity steganography [14]. Steganography techniques that use JPEG images as a cover are ineffective.

In audio steganography, the hidden message is encoded into the audio [15]. It's a method for safeguarding the transfer of sensitive information or concealing its presence altogether. If the communication is encrypted, it may also protect the secrecy of a secret message [16]. Audio Steganography is a method of transmitting concealed information by altering an audio stream in an undetectable way [17]. This method can conceal a hidden text or audio message inside a host [18]. Host message characteristics, before and during steganography, are identical to those of the stego message itself. Payload localization in pseudorandom scrambled JPEG image steganography has recently been established using co-frequency subimage filtering [19]. The fidelity of the predicted cover images affects the precision of this payload locating approach, which can be enhanced by developing a more accurate estimator [20].

Deep learning-based steganography algorithms are mostly focused on images, whereas audio steganography methods need development. Consequently, GAN-ASF aims to investigate steganography related to the audio domain. There are three components to the proposed training system which are the generator, discriminator, and steganalyzer. The original cover audio is used to create a steganographic cover. By employing the standard message embedding technique LSBM for steganographic cover audio, the concealed information is injected into it, mistaken for cover audio by experts. The site analyzer is taught to make erroneous predictions using this technique. The generator's weight parameters are updated based on misclassification errors.

Steganographic cover audio can be used to smuggle messages after the adversarial training between these three parties has concluded. It is possible to verify that the audio information distribution matches the arrangement of messages using a well-trained generator. However, cover audio containing hidden messages is known as steganographic cover audio. Classic steganography incorporates the secret message in the track's cover audio. The existing methods are fails to identifying the exact embedding location which causes to create the security, integrity and confidentiality issues. Therefore, the exact embedding location must be predicted to improve the security and data protection rate. According to testing data, GAN-ASF can produce cover audio undetectable by existing deep learning steganalyzers. The approach uses the different layers and training model that identifying the exact embedding location. This steganographic process improves the overall data security and protection rate.

2. Literature review and problem statement

[21] described Hybrid Multistage Framework (HMF) for combining encryption and steganography for data manipulation. The main intention of this study is to improving the data manipulation rate. After checking several and everything with the cover picture and various image formats and is appropriate that the text size be 15 % smaller than that of the cover picture in a hybrid multistage encrypted communication architecture that constructs sequential and mumbo jumbo encoding/decoding algorithms with pre-stage text encryption.

Novel linguistic steganalysis framework for Integrating Semantic and Syntactic (SeSy) Features described by [22]. A graph attention network and a transformer-architecture communicative approach would be employed to maintain syntactic information. Experimental evidence shows that the SeSy framework significantly improves on existing advanced steganalysis approaches because it incorporates syntactic information.

Steganography based on Technical and Non-Technical Steganography (STN-TS) for data security demonstrated by [23]. The main intension of this study is to maximizing the robustness and reducing the error rate while creating the steganography. This article examines and analyses steganography algorithms based on characteristics such as PSNR, MSE, and Robustness. The research finishes with suggestions for developing high-quality stego pictures, large payload capacity, and strong steganography approaches based on examining these characteristics and other difficulties.

Frequency Hopped Spread Spectrum (FHSS) techniques for secure audio stenography were discussed by [24]. The system aims to developing the secure data transmission to minimizing the interference and improving security. The transmission and reception of a code determine the changes that take place. Code-division multiple access (CDMA) communications may be made possible by using FHSS to minimize interference and to enable multiple access. A frequency hopping approach was utilized in this technique, where consumers are forced to switch frequencies at a certain time interval.

Multitask Identity-Aware Image Steganography (MIAIS) for restoring secret images initialized by [25]. The MIAIS system is to improving system robustness and increasing the security in steganography. The identification information may be preserved by introducing a basic content loss and designing a minimax optimization to cope with the opposing features. In other words, the robustness findings may be applied to various datasets. We may optionally add a restoration network into our technique to provide a multitasking framework for the hidden picture restoration.

In [26] introduced Genetic Algorithm (GA) based audio steganography for identifying the adopt location. This work uses the steganalysis dataset information for creating the audio steganography. During this process, discrete wavelet transform and discrete cosine transform for making the steganalysis. Then genetic operators such as selection, crossover and mutations are applied to predict the exact embedding location. The genetic algorithm based selected location minimize the unauthorized activities and mean square error rate. However, the system requires additional effort and embedding identification techniques to improve the overall efficiency.

Developed Convolution Neural Networks (CNN) to creating the audio steganography. This system aims to reducing the Iterative Adversarial Attacks while developing the steganography. The network uses the multiple layers that identifying the exact embedding location; during this process non-adaptive and adaptive steganography procedure to eliminate the iterative attacks. This process helps to reduce the overall data security and integrity. However, this method requires additional security methods to improve the authorization and authentication while accessing the data.

Light-weight generative neural networks are recommended for developing smart steganography. The network consists of three network models such as encoder, decoder and discriminator for identifying the secrete message. The algorithm helps to identifying the exact embedding location and the efficiency of the system evaluated using the steganalysis dataset. Then the introduced system manages the data security and robustness. Based on the analysis, there are some drawbacks such as efficiency ratio, accuracy ratio, performance ratio, signal-to-noise ratio, and error rate. Hence this paper GAN-ASF to achieve the audio steganography based on deep learning techniques.

3. The aim and objectives of the study

The aim of the study is identifying optimal message embedding location in audio steganography using generative adversarial networks.

To achieve this aim, the following objectives are accomplished:

– to maximizing the steganography quality by identifying the exact embedding location in audio by utilizing the Generative Adversarial Neural Network;

– to minimizing the deviation between the original and audio steganography by utilizing generator and discriminator to predict the embedding location which minimize the error rate;

– to enhance the transmitted data confidentiality, integrity and security by embedding the cover audio with the original information.

4. Materials and methods

4. 1. Object and hypothesis of the study

Dataset description: there are 33 npy files with a npy file size of (8000, 50*time). There are 8000 speech samples in total, with a sample size of (frames, QIM parameters). An AMR frame takes up time multiplied by a factor of 50. Thirty-three QIM parameters make up AMR; the first five are LPC parameters, the middle two are FCB parameters (Pulse Pairs), and the last two are ACB parameters (four integer pitch delay variables plus four fractional pitch delay variables). The dataset is collected from Multiple Voip Steganography Datasets. Here, the discussed system implemented with the help of FPGA related spatial domain. The collected files are processed by Generative Adversarial Networks (GAN) to implementing the quality steganography. The main hypothesis of this study is to improving the overall security in steganography-based information transmission. Let's assume that the transferred audio messages are sensitive and secured one. Therefore, the identified embedding locations are highly ensuring the security to the transferred data. The transferred data able to adapt with the transmission medium that ensures the security in the network.

4. 2. Generative adversarial network

The fundamental objective of GAN is to construct, with real-world samples, a generator that can create new samples with a data distribution identical to the genuine samples. It is possible to think of the generator as a kind of transformer since it takes arbitrary noise and converts it into actual samples. A discriminator to separate the produced samples from the actual data is used to improve the generator's performance. To summarise, an equilibrium is ultimately attained in the training process via the constant war game between generator and discriminator. This is feasible if the discriminator cannot discriminate between produced and actual samples. Naive GAN's training process is not stable, and the vanishing gradient issue might arise. Researchers have developed a new generation of GANs to make it easier to train the GANs. Fig. 1 shows the Generative Adversarial Network-based Audio Steganography Framework.

The steganography training framework has three main components: generator, discriminator, and a specially trained steganalyzer. Lin-Net is the platform on which the trained steganalyzer has been developed. To find the masked message, Lin-Net has already been trained. Cover audio is steganographically encoded using a generator that creates cover audio that is almost impossible to distinguish from the original. Additionally, cover audio should sound as close to the original as feasible when used for a steganographic cover.

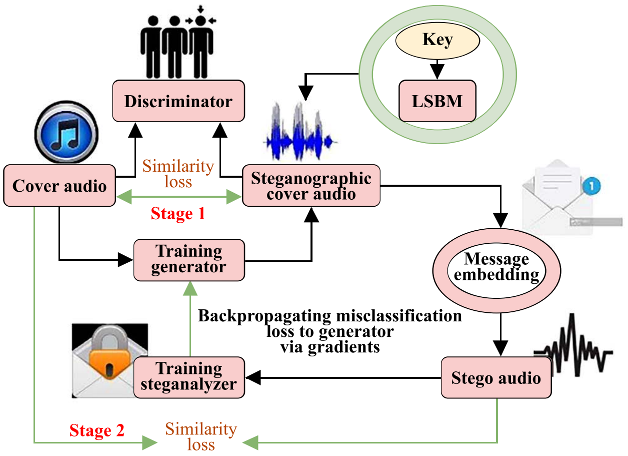


Fig. 1. Generative adversarial network-based audio steganography framework

By using the time-honoured LSBM message embedding approach, it is possible to create steganographic cover audio that can be mistakenly identified as such by a trained steganalyzer. Loss in forecasting is reported back to the generator, and weight parameters are adjusted accordingly. To be clear, let's promote steganalyzer misbehaviour because GAN-ASF aims to trick the deep learning-based steganalyzers. A well-trained generator may be obtained after enough adversarial training. Ultimate steganography, the most advanced kind of steganography, is used by GAN-ASF and involves the use of a specially trained generator. In the next step, let's employ the conventional steganography technique known as LSBM to encrypt a hidden message into the steganographic cover audio, creating an undetectable stego.

4. 3. Loss function

To ensure that the discriminator has a superior discriminant capacity. Consequently, let's separate the adversarial training process into two stages: the first stage includes just the discriminator and generator, and the second stage includes the trained steganalyzer.

Stage 1. The loss function for stage 1 can be stated as follows:

$$K_{s1} = K_C - K_{h1}. \tag{1}$$

As presented in equation (1), where K_{s1} denotes the stage 1 loss of the GAN framework, K_c and K_{h1} are the two losses calculated using binary cross-entropy as their basis. The generator's loss at stage 1 K_{h1} can be determined using this equation:

$$K_{h1} = \mathbb{D}_z\left[\log\left(1 + C\left(h(z)\right)\right)\right], \tag{2}$$

As presented in equation (2), D_z is the expectation operator applied to the input audio clips, where $C(h(z))$ denotes the actual audio. The discriminator loss K_c can be calculated as follows:

$$K_C = \left\{\mathbb{D}_z\left[\log C\left(h(z)\right) - \log\left(1 + C(z)\right)\right]\right\}. \tag{3}$$

As presented in equation (3), the sound that has been encoded using steganography can be distinguished from unencrypted sound $C(z)$ by using this discriminator.

Stage 2. The loss function for stage 2 is shown as follows:

$$K_{s2} = \beta K_{GAN} - \gamma K_S. \tag{4}$$

As presented in equation (4), the loss of the GAN framework in stage 2 is represented by K_{s2} and the similarity loss function K_S is used to quantify the similarity between steganographic cover audio β and g indicates the original audio. The loss of the generator and discriminator K_{GAN} is stated as:

$$K_{GAN} = K_C - K_{h2}. \qquad (5)$$

As presented in equation (5), binary cross-entropy is used to calculate the losses K_c and K_{h2}. The generator's objective is to generate steganographic audio identical to the original in terms of aural quality. An audio steganalyzer receives the embedded audio and uses it to determine the likelihood that it has been covered. The generator's loss Kh_2 can be determined as follows:

$$K_{h2} = \mathbb{D}_z\left[\log\left(1+C\left(h(z)\right)\right)\right] - \\ -\mathbb{D}_z\left[\log\left(1+T\left(E\left(h(z)\right)\right)\right)\right]. \qquad (6)$$

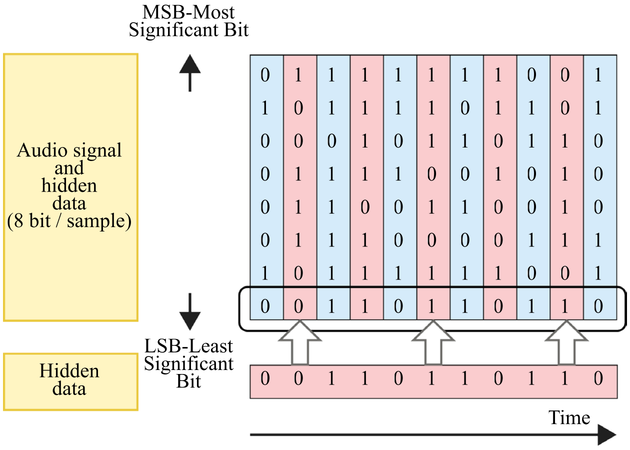
The similarity loss K_S is defined as:

$$K_S = \mathbb{D}_z\left[h(z)+z_1\right] - \mathbb{D}_z\left[E\left(h(z)\right)+z_1\right]. \qquad (7)$$

As presented in equations (6), (7), where z indicates the actual audio. D_z is the expectation operator applied to the original audio clips. C, $h(z)$ and T signify the discriminator, steganalyzer, and generator, respectively. $E(h(z))$ denotes the classical information technique. The K_1 norm is used to quantify the loss of steganographic cover audio compared to the original, K_1 norm steganographic cover audio has a somewhat superior perceptual quality than K_2 norm steganographic cover audio of similarity loss.

Fig. 2 shows that one bit of embedded data overwrites the 8 b/sample LSB. One of the first forms of information concealing, the LSB (Least Significant Bit), goes back to the early days of computer programming. The audio on the cover is compressed to the smallest bit feasible to carry each bit of the message. In this case, 16 kbps of data is encoded into the stream for a 16 kHz sampled audio file. The LSB methodology provides a huge capacity for data embedding and is rather easy to utilize or combine with other approaches. The security of this method is low since it is vulnerable to simple assaults, yet it has a low resistance to noise addition. Stego-audio data will be lost if it is boosted, filtered, noise-added, or compressed using lossy techniques. It's also possible for an attacker to see the message by simply eliminating the LSB plane in question. A simple LSB method has included a speech message in wireless communication. The least error replacement approach was used while embedding four bits per sample to maximize concealing capacity while decreasing error in stego audio. Four samples later, the embedding mistake has spread. The second approach described here discovered a hidden data channel capable of 176.3 kbps in a 44.1 kHz transmission.

A new embedding layer has been added to the LSB method's resilience to distortion and noise, increasing its depth from four to six and eight layers. There is just one bit of the original 16-bit sample that is altered by the message's bit. The other bits can be changed to generate a new sample closer to the original to reduce embedding error. The embedding of the LSB has been shifted to the eighth bit, preventing it from hiding in the host signal's quiet or almost silent sections. Compared to standard LSB techniques, the embedding happens in the eighth bit, which improves the method's durability. There will be a reduction in concealing capability due to the need to preserve the audio stream's audio detection accuracy by leaving certain samples unalterable. Because embedded bits are located in the sixth or eighth position from the stego audio, message retrieval remains one of the fundamental drawbacks of the LSB and its variation. Then the embedding procedure shown in Fig. 3.


MSB-Most Significant Bit
Audio signal and hidden data (8 bit / sample)
Hidden data
LSB-Least Significant Bit
Time

0	1	1	1	1	1	1	1	0	0	1
1	0	1	1	1	1	0	1	1	0	1
0	0	0	1	0	1	1	0	1	1	0
0	1	1	1	1	0	0	1	0	1	0
0	1	1	0	0	1	1	0	0	1	0
0	1	1	1	0	0	0	0	1	1	1
1	0	1	1	1	1	1	1	0	0	1
0	0	1	1	0	1	1	0	1	1	0

0 0 1 1 0 1 1 0 1 1 0

Fig. 2. One bit of embedded data overwrites the 8 b/sample Least Significant Bit

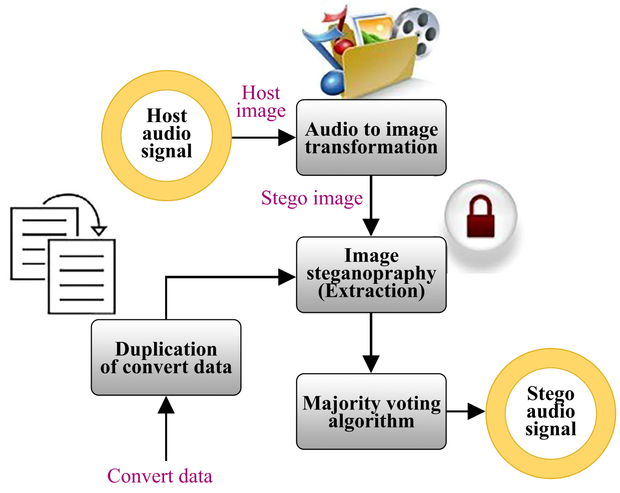

Host audio signal
Host image
Audio to image transformation
Stego image
Image steganopraphy (Extraction)
Duplication of convert data
Convert data
Majority voting algorithm
Stego audio signal

Fig. 3. Procedure for embedding

Fig. 4 show the procedure for extracting. Audio steganography is a method of concealing a message inside the sound of another sound source. This approach uses a filter to reduce the output signal's dynamic range, primarily outside the HAS dynamic range. Following this filtering step, the sub-signal is transformed into an image in two dimensions (2D) to maintain the visual-auditory link. As a result, audio steganography is reduced to an image steganography problem. One of this method's primary features is its large embedding capacity and resistance to MP3 compression. Data obfuscation techniques similar to these have been suggested. Audio input has been transformed into a visual representation utilizing wavelet audio to image transform. A basic form of image steganography is used to hide the data in the final image. Afterwards, the image is converted back into an audio stream.

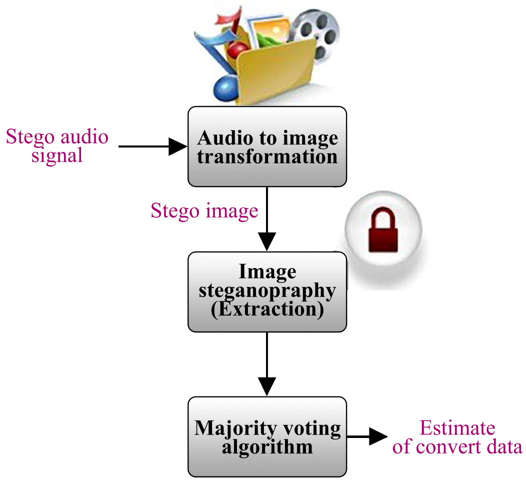

Stego audio signal
Audio to image transformation
Stego image
Image steganopraphy (Extraction)
Majority voting algorithm
Estimate of convert data

Fig. 4. Procedure for extracting

A time-domain processing approach is used to watermark the information. The watermark can be detected using this approach without needing the original signal. The audio signal's frequency and amplitude are manipulated during the watermarking process, which reduces the signal's audibility. Normal signal processing processes such as filtering, resampling, and so on cannot remove the embedded watermark. However, this method is vulnerable to sophisticated assaults, such as the inability to identify watermarks in a signal that has experienced a temporal shift. It's more probable that attackers will alter the original samples when using these techniques. Steganography utilizes audio based on the idea of music edge detection. The major purpose of this method is to resist Time Scale Modification (TSM) procedures. By selectively extending the acoustic signals in areas with very little temporal data, the TSM algorithms strive to preserve the music's edges. There are watermarks included in this work because of its TSM attribute. In addition to being resistant to typical assaults, this approach is also resistant to time-scale alteration methods.

Stencilling speech into an MP3 file using this approach is impossible since it operates in the temporal domain. The basic premise of this approach is first to identify the speech's quiet pauses. Then, secret information is hidden by varying the duration of 12 intervals. MP3-resistant voice steganography methods have been developed. However, the embedded data has been lost in a noisy environment because these approaches are used in a modified domain. The approach can overcome this obstacle and concealing data in voice signals at a high rate. Although the quality of speech is lowered, with its big payload and great privacy, audio steganography has the least potential to conceal data as a temporal encoding method. This technique is known as «automated audio generation steganography». The secret bits stream that has to be encoded is used to generate high-quality audio coverings automatically. Suppose each sample point's conditional probability distribution space can be suitably coded. In that case, it is possible to pick the associated signal output by the bitstream and thereby achieve a hidden message embedding.

According to this method of data concealment, secret data may be disguised in cover audio of the same size. It can be accomplished using a fractal coding technique, which utilizes a high compression ratio to produce an acceptable reconstructed signal. Up to 30 % more information may be hidden in audio steganography, making this the most important advancement. Interval and variable low-bit coding methods have been used in modern steganography for wireless communication. Configurable low-bit coding allows for the embedding speed and capacity to be varied by a time for hidden messages to be inserted into an audio file at a predetermined threshold, an entirely new method of hiding information in audio files. An additional audio channel is generated to conceal the information. The secret key is a threshold level that encodes a secret message signal in the audio stream. Fig. 5 shows the optimum cover JPEG image of the estimate approach based on sub-image cover probability.

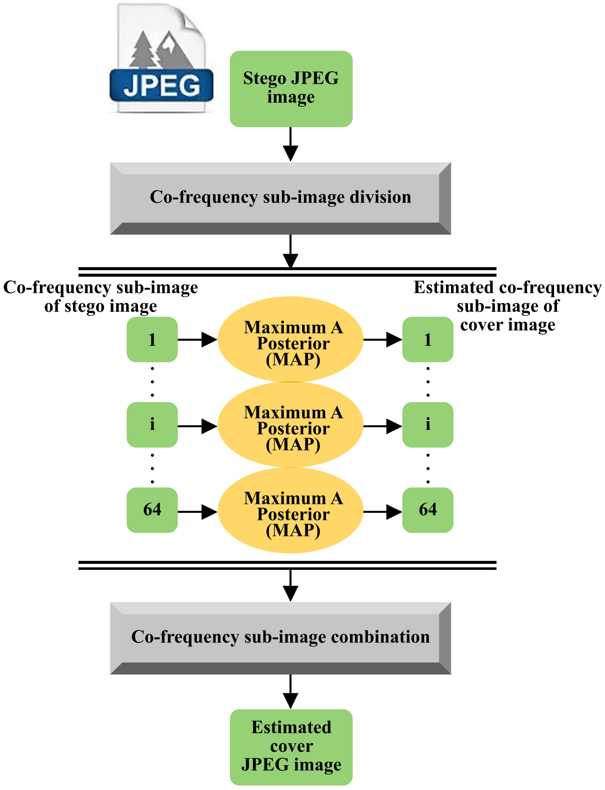

JPEG
Stego JPEG image
Co-frequency sub-image division
Co-frequency sub-image of stego image
Estimated co-frequency sub-image of cover image
1
Maximum A Posterior (MAP)
1
i
Maximum A Posterior (MAP)
i
64
Maximum A Posterior (MAP)
64
Co-frequency sub-image combination
Estimated cover JPEG image

Fig. 5. The optimum cover JPEG image of estimate approach based on sub-image cover probability

In theory, it is possible to compute the probability of all possible coverings and look for an equation (11). In contrast, the cover image has too many coefficient values to explore it thoroughly. It is possible to use the Viterbi algorithm to best predict the cover JPEG image by integrating the co-frequency sub-images. First-order Markov models at various frequencies can provide more precise estimates of the cover co-frequency sub-images. There are a large number of zero coefficients in several frequency ranges. 'Non-zero coefficients are thus not statistically significant. Co-frequency sub-images can be estimated using first-order Markov models blended across locations.

4. 4. Payload localization based on cover co-frequency sub-image

The coefficient at the position (k, l) of the r^{th} stego image has a residual value $S_r\,(k, l)$ defined:

$$s_r(k,l)=\left|T_r(k,l)+D_r(k,l)\right|. \quad (8)$$

As presented in equation (8), a stego picture's cover image is denoted by D_r, where D_r is the cover image for T_r.

Detection R is capable of distinguishing between stego poses and non-stego postures $\widehat{s}_r(k,l)$ are stated as:

$$\widehat{s}(k,l)=\frac{\sum_{r=1}^{R}\widehat{s}_r(k,l)}{R}=\frac{\sum_{r=1}^{R}\left|T_r(k,l)+\breve{D}_r(k,l)\right|}{R}. \tag{9}$$

As presented in equation (9), investigators are capable of deciphering stego imagery $\widehat{s}_r(k,l)$ can calculate the mean of estimated residuals r in the same place $T_r(k, l)$ of several cover images $\breve{D}_r(k,l)$.

According to the averaged estimated residuals, the investigator can be able to discern between stego and non-stego positions more reliably than by guessing at random based on the number of images inserted along the same path as follows,

$$g(k,l)=\begin{cases}1,\widehat{s}(k,l)\le td;\\0,\widehat{s}(k,l)\ge td.\end{cases} \tag{10}$$

As presented in equation (10), a non-stego position is denoted by $g(k, l)$. A stego position is indicated by $\widehat{s}(k,l)$. And the threshold for making a choice is represented by *td*.

Image co-frequency Markov Model (MM).

Co-frequency sub-image estimates for T_r^e should be calculated using the cover co-frequency sub-image estimation D_r^e with the highest posterior probability q, in terms of statistical inference $\breve{D}_r^e$ is given in (11):

$$\breve{D}_r^e = arg\max_{D_r^e} q\left(D_r^e|T_r^e\right); \tag{11}$$

$$\breve{D}_r^e = arg\max_{D_r^e} q\left(T_r^e|D_r^e\right)q\left(D_r^e\right).$$

The optimum estimate of the cover co-frequency sub-image k is therefore translated into a issue of the greatest estimation of the posterior probability $q\left(T_r^e|D_r^e\right)$ as follows:

$$q\left(T_r^e|D_r^e\right)=\prod\nolimits_k q\left(T_r^e(k)|D_r^e(k)\right); \tag{12}$$

$$q\left(D_r^e\right)=\prod\nolimits_k q\begin{pmatrix}D_r^e(k)|D_r^e(k+1),\\D_r^e(k+2),\ldots,D_r^e(k+m)\end{pmatrix}. \tag{13}$$

As presented in equations (12) and (13), quantized Discrete cosine transform (DCT) coefficients for stego co-frequency sub-images $D_r^e(k)$ are connected to their quantized DCT coefficients to cover co-frequency sub-images $T_r^e(k)$ only if m is a positive integer; the cover co-frequency sub-image $q\left(D_r^e\right)$ On the other hand, it is modelled using an ordered Markov model.

4. 5. First-order MM-based optimal cover JPEG image estimate

The hidden MM can be used to describe the co-frequency sub-image t_{mk}, and the Viterbi algorithm is a prominent way to solve the hidden MM issue t_{1k}. The Viterbi technique can be used to find the best cover co-frequency sub-image d_{1k}. Calculation of probable initial cover element values is carried out through the Viterbi algorithm u (d_{1k}) as follows:

$$u\left(d_{1k}\right)=q\left(t_{1k}|d_{1k}\right)q\left(d_{1k}\right). \tag{14}$$

Then, the potential values for the following cover elements $u(d_{mk})$ are calculated as follows:

$$u\left(d_{mk}\right)=u\left(d_{m+1,k}\right)q\left(d_{mk}|d_{m+1,k}\right)q\left(t_{mk}|d_{mk}\right)^{d_{m+1,k}}. \tag{15}$$

As presented in equations (14) and (15), the possible value q for the m^{th} cover element in the image is represented by $d_{m+1,k}$.

F5 steganography has a coefficient variable transition probability of q, as follows:

$$q\left(t_k|d_k\right)=\begin{cases}1+\dfrac{p}{2}, t_k=d_k+1 \text{ and } t_k<0;\\1+\dfrac{p}{2}, t_k=d_k-1 \text{ and } t_k>0;\\\dfrac{p}{2}, t_k=d_k \text{ and } t_k\ne 0;\\1, t_k=d_k=0\ ;\\0, \text{others}.\end{cases} \tag{16}$$

As presented in equation (16), each node's score t_k is calculated sequentially, a it is then connected to the node that has the highest value d_k.

4. 6. F5 payload position without matrix encoding

The following equation determines the difference between the given stego image and the estimated cover Joint photographic experts group JPEG image $s_r(k, l)$:

$$s_r(k,l)\begin{cases}0,\ \mathrm{mod}(k,8)=0 \text{ and } \mathrm{mod}(l,8)=0;\\\left|T_r(k,l)+\breve{D}_r(k,l)\right|, others;\end{cases} \tag{17}$$

As presented in equation (10), using the specified stego JPEG pictures $\breve{D}_r(k,l)$ and estimated cover JPEG images $T_r(k, l)$ Residuals can be calculated for each point k, and then averaged.

5. Results of message embedding location in audio steganography

5. 1. Accuracy of the message embedding location

There are various ways to hide messages in digital media, such as audio, picture, or video. Steganography plays an important part in real-world applications since privacy is a major issue for everyone in today's society. Audio signals are the most often used cover signals due to their larger size and greater capacity to conceal more information. With higher redundancy and a quicker data transmission rate, digital audio signals are suited for cover usage. This study examines how these techniques may be used to communicate information in audio and spoken signals in audio steganography-based methodologies. This paper discusses many strategies for digital audio steganography and their algorithms and concepts, and a compendium is offered. The research includes a critical analysis of the current techniques, which may be used by the community to choose which one best meets their needs.

LSB embedding is a method for concealing text inside a picture. Secret communications are encrypted twice using the new approach. The metadata is first constructed, and the header information is inserted into the first few bytes of the cover image. Second, after hashing commonly occurring phrases, the secret message is encoded in the cover image using an upgraded technique. More computations must be performed to store the same amount of secret data in a smaller area.

As found in equation (18) and Fig. 6 shows the accuracy ratio based on the dataset [29]. Once encoded sound is transferred digitally, a comparison of the original ($Y_0(m)$) signal and its decoded version $Y_d(m)$) is as simple as comparing apples to oranges, and assuming the transmission medium and any file storage maintain their integrity, the accuracy should be remarkable. Problems with data retrieval accuracy and encryption key security have been documented, as have the sluggishness and complexity of computations. Encryption of a text file using a hybrid system is described in this work as a straightforward, low-complexity solution that ensures the security of the message while maintaining its dependability. Using the LSB approach, the technique takes a bit from the symbol's bitstream to rotate the preceding and subsequent bits in the bitstream:

$$Y_d = \{y(m)\} 0 < m < m_{\max}. \tag{18}$$

Securing data transfer, encrypting data, and using steganography are critical. In this study, steganography and cryptography have been combined to produce a secure hybrid stego-system. To begin, a cypher text is generated by encrypting a text message using a novel approach based on bits cycling. An upgraded LSB approach is utilized in the second step to disguise the text bits in a wave format. Data may be effectively secured using a hybrid approach. SNR and error were used to assess the suggested system's performance.

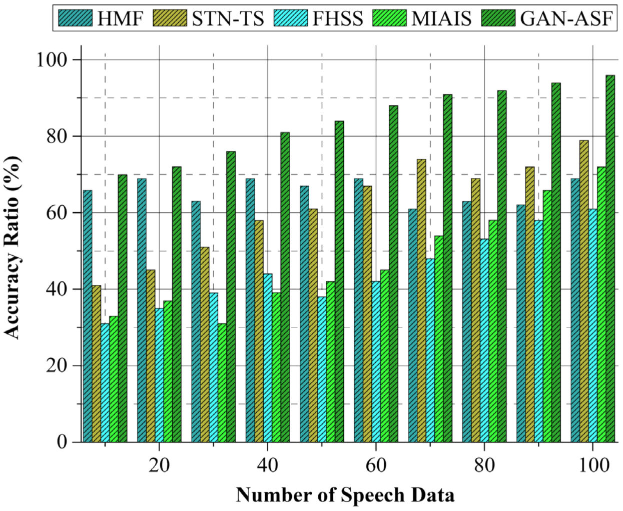


Fig. 6. Accuracy Ratio (%)

Fig. 7 explores the performance ratio (%) based on the dataset [26]. The presented findings demonstrated that the suggested strategy outperformed machine learning-based steganalysis techniques. Compared to prior audio data concealing approaches, the suggested method has similar perceptual performance and stability while including a larger quantity of data. Stego images with random noise have been added to the cover image to imitate the embedding process on the cover image. An adversarial attack on deep learning-based network steganalysis has been used to gain perturbation, as indicated in the article. Secret messages were finally encoded using the adaptive steganography technique. The suggested technique outperformed deep learning-based steganalysis methods by a wide margin. Fig. 8 describes the efficiency ratio based on the dataset [26].

The suggested AS technique is tested extensively to establish its efficacy and superiority over other AS methods. The suggested AS technique is tested extensively to establish its efficacy and superiority over other audio steganography methods. The best approach to make cover audio with great perceived quality is to utilize the platform. Even at high embedding levels, a steganographic cover audio created by the developer of the proposed steganography approach has excellent perceptual quality and good undetectability performance.

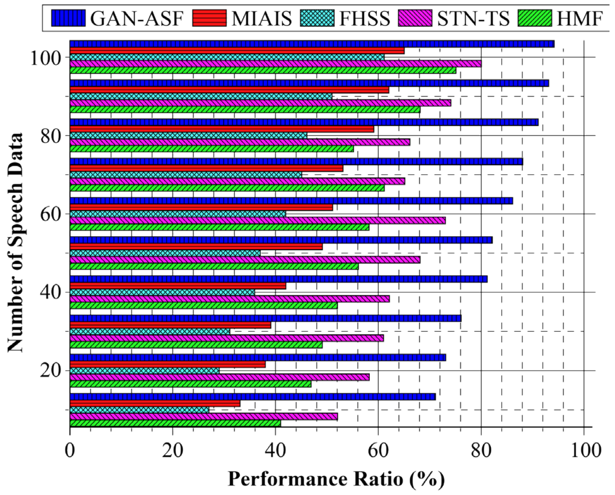


Fig. 7. Performance Ratio (%)

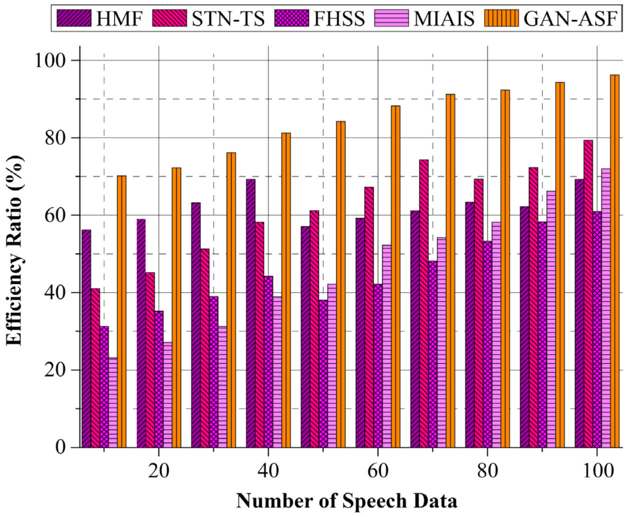


Fig. 8. Efficiency Ratio (%)

5. 2. Reducing the difference between the cover and original audio

The purposeful and incidental attack has an issue with these devices, which is why they're not as secure as other devices. A solution has been presented that would conceal the message using the method. It is one of the most popular and oldest strategies for concealing information. The message bits are contained in the least significant bit of the host signal, which has less impact on the host signal than the original signal. Security and communication system modification resistance is weak for audio files that might hide vast quantities of data. Fig. 9 illustrated that the error rate analysis of GAN-ASF based audio steganography process:

$$\text{Error} = \frac{\sum_1^m [y-x]^2}{N * M}, \tag{19}$$

where X is a stego signal, Y represents the original signal, M and N denotes the number of rows and columns in the input signals.

Fig. 9 and (19) show error rate has been calculated based on the dataset [26]. In place of the traditional three random keys used in LSB, this audio steganography approach uses Lifting Wavelet Transform (LWT) and LSB. Three randomly generated keys improve LSB's resilience. In addition, our method utilized LWT instead of other ways to prevent the rounding error of the approximate values as LWT is transformed.

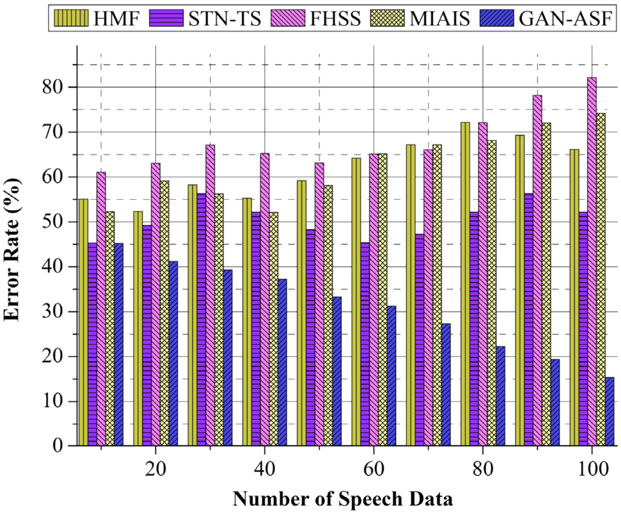


Fig. 9. Error Rate (%)

5. 3. Quality and security analysis of Audio-stageography

As an audio steganographer one of the goals is to make any modification to the carrying sound small that it cannot be detected by the hearing threshold or by digital monitoring. Calculations evaluating the original and encrypted signals were performed to determine the influence of message encoding on audio signal quality. All forms of audio steganography impact our proposed method used Signal to Noise Ratio (SNR). This is important since it helps determine how much distortion the hidden data introduces and how much data can be concealed. Matlab is used to build the approach, and numerous audio signals were used to test it. Text, audio, and picture files may be utilized to encode a secret message:

$$SNR = 10^{*} \log_{10} \frac{\sum_{j=1}^{m} Y^{2}(m)}{\sum_{j=1}^{m} \left[Y^{2}(m) - X(m) \right]^{2}}, \qquad (20)$$

where X is a stego signal, Y represents the original signal, which denotes the number of rows and columns in the input signals.

Fig. 10 and (19) show signal to the noise ratio has been described based on the dataset [26]. The concealed message was quantified as adding noise to a previously recorded audio clip. In the noise-controlled rehearsal room, SNR data is acquired through experiments comparing the encoded and original signal at various sensitivity variable settings and the single sample per character encoding. Thus the introduced GAN based audio steganography process successfully identifying the message embedding location which is difficult to identifying by third party or unauthorized users compared to the existing methods. Then the generated audio cover satisfies the quality and security factors.

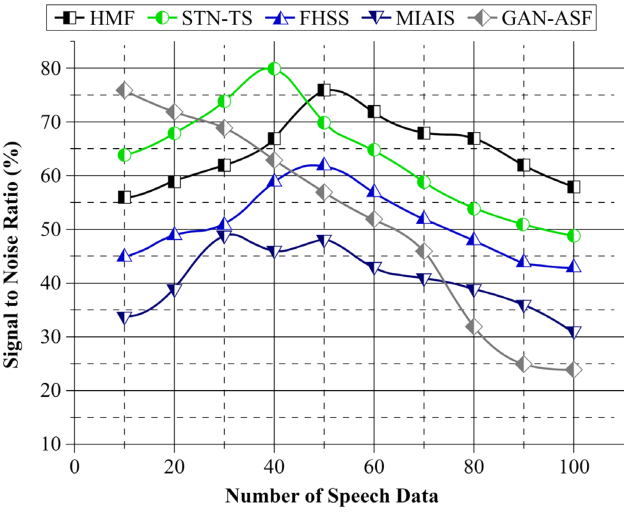


Fig. 10. Signal to Noise Ratio (%)

6. Discussion of experimental results of Generative Adversarial Network based Audio Steganography

The research work uses the Generative Adversarial Networks (GAN) with Least Significant Bit Matching (LSBM) approach is utilized for creating the effective audio steganography. Here, the cover audio is successfully investigated by applying the GAN approach that utilizes the different modules such as Generator, discriminator and encoder to extracting the audio features such as variations, modulations, pitch etc. These features are helps to identifying the exact message embedding locations in the cover audio. During the process, LSBM technique incorporated with GAN that improves the overall audio steganography efficiency up to 96.2 %. The steganography efficiency is discussed in Fig. 6, 8 which clearly shows that introduced algorithm recognize the embedding location with maximum rate for different speech data. After identifying the embedding location, the quality of the signal is similar to the original signal which is difficult to identifying by the intermediate user. Therefore, the quality and security of the cover audio is successfully managed compared to the existing systems. The effectiveness of the introduced GAN based audio steganography approach is compared with several existing methods. However, this study requires the optimization techniques to improve the steganography performance. The integration of neural model with optimization approach reduces the difficulties in embedding location identification process. This optimization process reduces the deviations between the computing outputs and the efficiency of the system illustrated in Fig. 9. In large data volume process, system meets the optimization problems that reduce the entire audio steganography efficiency. Along with this, high-dimension of feature handling process creating the computation complexity which is described in Fig. 10. The Fig. 10 shows that text embedded audio signals are more quality because of the optimal selection of embedding location. It can be overcome by applying the feature selection techniques. Therefore, the optimization and feature selection techniques are requiring to improve the overall audio steganography process. Once the exact embedding location is identified the quality of audio has been maintained with high security.

7. Conclusions

1. The research work uses the generator, discriminator and steganalysis for analyzing the message embedding location. This process improves the overall efficiency of message embedding location identification in audio steganography. The GAN network has multiple layers that successfully identifying the appropriate location in audio signal. The AS and voice processing methodologies have been published in the literature and consolidated. Techniques that use audio steganography mostly deal with audio and spoken signals for covert communication. A thorough examination of steganography methods considers the three most important facets: capacity, security, and resilience. Various methodologies in this subject are compared and grouped based on their operational commonalities, which may be valuable to researchers in this field. The effective utilization of the neural layers improves the steganography efficiency up to 96.2 % of accuracy and 94.5 % of performance ratio.

2. The audio steganography process uses the neural network to predict the embedding location identification process. During the analysis, network uses the fine-tuned network parameters, which leads to minimize the deviation between the steganography location outputs. Here, the neural networks parameters are continuously updated while the deviations are occurred. During this process, backpropagation learning algorithm is applied for reducing the error value. The introduced method minimizes the error value up to 15.7 % compared to other method.

3. Here, the embedding process is performed by applying the Least-Significant Bit Matching (LSBM) techniques are applied to encrypt the message that helps to maintain the data security, integrity and confidentiality. In addition, the method ensure the audio quality which is evaluated in the experimental analysis, the system gets 24.3 % of SNR value and 94.8 % of efficiency ratio.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

References

1. Zhao, J., Wang, S. (2022). A stable GAN for image steganography with multi-order feature fusion. Neural Computing and Applications. doi: https://doi.org/10.1007/s00521-022-07270-w
2. Asimopoulos, D. C., Nitsiou, M., Lazaridis, L., Fragulis, G. F. (2022). Generative Adversarial Networks: a systematic review and applications. SHS Web of Conferences, 139, 03012. doi: https://doi.org/10.1051/shsconf/202213903012
3. Li, X., Yao, R., Lee, J. (2022). Research on Digital Steganography and Image Synthesis Model Based on Improved Wavelet Neural Network. Computational Intelligence and Neuroscience, 2022, 1–12. doi: https://doi.org/10.1155/2022/7145387
4. Najm, O. A., Nori, A. S. (2022). Steganography Method of the Bigger Size in WebP Image Using M2PAM Algorithm for Social Applications. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 13 (2), 595–610. URL: https://turcomat.org/index.php/turkbilmat/article/view/12359
5. Xiang, L., Wang, R., Yang, Z., Liu, Y. (2022). Generative Linguistic Steganography: A Comprehensive Review. KSII Transactions on Internet and Information Systems, 16 (3), 986–1005. doi: https://doi.org/10.3837/tiis.2022.03.013
6. Mallika, Ubhi, J. S., Aggarwal, A. K. (2022). Neural Style Transfer for image within images and conditional GANs for destylization. Journal of Visual Communication and Image Representation, 85, 103483. doi: https://doi.org/10.1016/j.jvcir.2022.103483
7. Qureshi, K. N., Kaiwartya, O., Jeon, G., Piccialli, F. (2022). Neurocomputing for internet of things: Object recognition and detection strategy. Neurocomputing, 485, 263–273. doi: https://doi.org/10.1016/j.neucom.2021.04.140
8. Phipps, A., Ouazzane, K., Vassilev, V. (2022). Securing Voice Communications Using Audio Steganography. International Journal of Computer Network and Information Security, 14 (3), 1–18. doi: https://doi.org/10.5815/ijcnis.2022.03.01
9. Wang, Y., Huang, L., Yee, A. L. (2022). Full-convolution Siamese network algorithm under deep learning used in tracking of facial video image in newborns. The Journal of Supercomputing, 78 (12), 14343–14361. doi: https://doi.org/10.1007/s11227-022-04439-x
10. Bao, Z., Xue, R. (2021). Survey on deep learning applications in digital image security. Optical Engineering, 60 (12). doi: https://doi.org/10.1117/1.oe.60.12.120901
11. Zhang, D., Ma, M., Xia, L. (2022). A comprehensive review on GANs for time-series signals. Neural Computing and Applications, 34 (5), 3551–3571. doi: https://doi.org/10.1007/s00521-022-06888-0
12. Nguyen, T. D., Le, H. Q. (2022). A secure image steganography based on modified matrix encoding using the adaptive region selection technique. Multimedia Tools and Applications, 81 (18), 25251–25281. doi: https://doi.org/10.1007/s11042-022-12677-7
13. Hemeida, F., Alexan, W., Mamdouh, S. (2021). A Comparative Study of Audio Steganography Schemes. International Journal of Computing and Digital Systems, 10 (1), 555–562. doi: https://doi.org/10.12785/ijcds/100153
14. Chen, L., Wang, R., Yan, D., Wang, J. (2021). Learning to Generate Steganographic Cover for Audio Steganography Using GAN. IEEE Access, 9, 88098–88107. doi: https://doi.org/10.1109/access.2021.3090445
15. Rakshit, P., Ganguly, S., Pal, S., Le, D.-N. (2021). Securing Technique Using Pattern-Based LSB Audio Steganography and Intensity-Based Visual Cryptography. Computers, Materials & Continua, 67 (1), 1207–1224. doi: https://doi.org/10.32604/cmc.2021.014293
16. Hameed, A. S. (2021). A High Secure Speech Transmission Using Audio Steganography and Duffing Oscillator. Wireless Personal Communications, 120 (1), 499–513. doi: https://doi.org/10.1007/s11277-021-08470-8

17. Ying, K., Wang, R., Lin, Y., Yan, D. (2021). Adaptive Audio Steganography Based on Improved Syndrome-Trellis Codes. IEEE Access, 9, 11705–11715. doi: https://doi.org/10.1109/access.2021.3050004
18. Abdulkadhim, H. A., Shehab, J. N. (2022). 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), 320. doi: https://doi.org/10.11591/ijece.v12i1.pp320-330
19. Ganwani, P., Gupta, L., Jain, C., Kulkarni, R., Chaudhari, S. (2021). LSB Based Audio Steganography using RSA and ChaCha20 Encryption. 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT). doi: https://doi.org/10.1109/icccnt51525.2021.9580177
20. Mahmoud, M. M., Elshoush, H. T. (2022). Enhancing LSB Using Binary Message Size Encoding for High Capacity, Transparent and Secure Audio Steganography – An Innovative Approach. IEEE Access, 10, 29954–29971. doi: https://doi.org/10.1109/access.2022.3155146
21. Osman, O. M., Kanona, M. E. A., Hassan, M. K., Elkhair, A. A. E., Mohamed, K. S. (2022). Hybrid multistage framework for data manipulation by combining cryptography and steganography. Bulletin of Electrical Engineering and Informatics, 11 (1), 327–335. doi: https://doi.org/10.11591/eei.v11i1.3451
22. Yang, J., Yang, Z., Zhang, S., Tu, H., Huang, Y. (2022). SeSy: Linguistic Steganalysis Framework Integrating Semantic and Syntactic Features. IEEE Signal Processing Letters, 29, 31–35. doi: https://doi.org/10.1109/lsp.2021.3122901
23. Dhawan, S., Gupta, R. (2020). Analysis of various data security techniques of steganography: A survey. Information Security Journal: A Global Perspective, 30 (2), 63–87. doi: https://doi.org/10.1080/19393555.2020.1801911
24. Jeyalilly, M., Kannan, S., Muthukumaravel, A. (2020). New Innovative Secure Audio Stenography Using Frequency Hopped Spread Spectrum Techniques in Mobile Computing. Computing. Malaya Journal of Matematik, S (2), 4564–4568. URL: https://www.malayajournal.org/articles/MJM0S201177.pdf
25. Cui, J., Zhang, P., Li, S., Zheng, L., Bao, C., Xia, J., Li, X. (2021). Multitask Identity-Aware Image Steganography via Minimax Optimization. IEEE Transactions on Image Processing, 30, 8567–8579. doi: https://doi.org/10.1109/tip.2021.3107999
26. Multiple Voip Steganography Datasets. URL: https://www.kaggle.com/datasets/wujunyan/amr-steg