Contents lists available at http://qu.edu.iq

Al-Qadisiyah Journal for Engineering Sciences

Journal homepage: https://qjes.qu.edu.iq

Hide an audio file inside a video for a secure connection

Rasha Muthana [1]*, Nora Ahmed Mohammed [2], and Amir Aboubakr Shaker Mohmoud [2]

[1]*Department of Roads and Transport Engineering, College of Engineering, University of Al-Qadisiyah, Al-Qadisiyah, Iraq*
[2]*Department of Electrical and Communication Engineering, College of Engineering, University of Al- Qadisiyah, Al-Qadisiyah, Iraq*
[3]*School of Digital Forensics and Cyber Security, National Forensic Sciences University Gandhinagar, India*

A R T I C L E I N F O

Article history:
Received 31 May 2024
Received in revised form 16 August 2024
Accepted 05 October 2024

Keywords:
Steganography
RC4 algorithm
Least Significant Bit
Forbidden Zone Data Hiding
Human Auditory System

A B S T R A C T

Steganography is essential in modern cryptography and communications, enhancing the security and confidentiality of sensitive data exchanges. It has become an interesting tool because not only have security requirements for secret messages become stronger, but video has also become more popular. This paper introduces an advanced method combining sequence-to-sequence transformer models for speech recognition, RC4 encryption, and the Least Significant Bit (LSB) technique for data embedding in videos. The approach securely embeds audio messages within video streams, ensuring that even if detected, the data remains inaccessible without the decryption key. Our methodology includes converting audio to text, encrypting it, and embedding the encrypted data into video files, with a subsequent recovery process that preserves the original audio's emotional and tonal qualities. Evaluations using the UCF101 dataset confirm the method's effectiveness in maintaining video quality, with minimal visual distortion, and robust data security. This research provides a secure framework for covert communication, with potential applications in areas requiring high-level data privacy.

1. Introduction

In today's technological age, information security is critical for online communications. It has been one of the most daunting challenges we have faced. One method of confidentiality in audio signals is audio steganography. This technique takes advantage of the masking capabilities of the human ear (HAS), where soft sounds remain undetectable when presented along with loud sounds Encode data into audio signals for purposes, including protection of copyrighted audio content and secure communication. The ever-increasing global reliance on the internet has greatly increased the demand for advanced security. Designers of the Internet are constantly trying to protect it from interference, and they use many methods and policies to achieve this goal. Researchers are also refining the constant maintenance techniques that hackers use when trying to infiltrate and compromise information systems. Now they are coming up with new ways to foil these hackers' plans. Where possible, data should be copied, modified, or accessed illegally without the owner's knowledge to constitute copyright infringement. Consequently, there is a growing need for privacy in digital data sets. This allows owners to confirm copyright ownership, prevent unwanted sensitive information, and easily add information. The main objective of the steganography industry is to hide, store, and embed hidden information in digital data sets. The main goal of steganography is to achieve secure communication in a completely invisible manner so that no one can detect the existence of hidden secrets [1]. The word "steganography" comes from the Greek, which means "writing in disguise." It is a widely used technique in information technology [2]. Throughout history, steganography has used techniques such as hiding messages inside the body of a dog by using invisible ink, shaving a sender's head to write a message or tattoo, or wearing a picture in the hair of an angel [3]. Steganography therefore provides methods for hiding a secondary message from the primary message. The primary message acts as a carrier signal, which can be text, audio, image, video, etc.,

* *Corresponding author.*
E-mail address: rasha.muthna.alfatlawy@qu.edu.iq (Rasha Muthana)

https://doi.org/10.30772/qjes.2024.150139.1250

Nomenclature:

RC4	*Rivest cipher 4*	*DSSS*	*direct sequence spread spectrum*
LSB	*Least Significant Bit*	*KSA*	*Key Scheduling Algorithm*
HAS	*Hearing Assessment scale*	*PRGA*	*Pseudo-Random Generation Algorithm*
FZDH	*Forbidden Zone Data Hiding*	*RGB*	*red, green, blue*
PCA	*Principal Component analysis*	*UCF101*	*University of central Florida 101*
DCT	*Discrete Cosine Transform*	*PSNR*	*Peak Signal-to Noise Ratio*
DWT	*Discrete Wavelet Transform*	*SSIM*	*Structural Similarity Index*
H.264	*Advanced video coding*	*DSSS*	*direct sequence spread spectrum*
HVS	*Human visual system*		
SS	*Spread spectrum*		

While the secondary, hidden message is called a weighted [4], stop message there is a generic block diagram of steganography. In this research, several unique contributions can be identified compared to other related studies in the field: Integration of Text-to-Audio Conversion with Encryption: The method starts by converting audio to text, which is common in some approaches. However, integrating this with encryption (using RC4) before embedding ensures that the information remains secure throughout the embedding process. This dual-layered approach (text conversion + encryption) enhances the security of the embedded data. Binary Conversion and LSB Embedding: Converting the encrypted text into binary and embedding it into the least significant bits (LSB) of video frames is a technique used in steganography. This method aims to minimize visual distortion, which is crucial for maintaining the quality and integrity of the video while hiding the data effectively. Voice Cloning for Audio Reconstruction: Using a converter model capable of instant voice cloning represents an advancement in the field of text-to-speech synthesis. This allows for the accurate reconstruction of audio data from the embedded text, including nuanced characteristics like emotion, sound, and pitch. Such capabilities enhance the fidelity and usability of the extracted audio. The remaining sections of this paper are organized as follows: In Section 2, we review related work in the field and explore various types of steganography methods. Section 3 details our methodology in depth. In Section 4, we present the results of our study and provide a discussion. Finally, Section 5 offers conclusions and suggestions for future work.

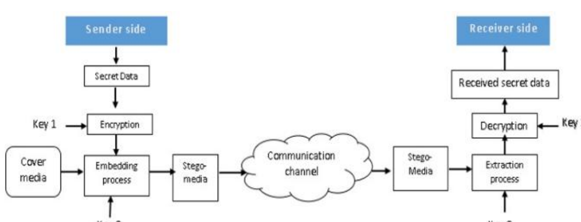

Figure 1. A steganography system's block diagram

2. Related work

Steganography is the technique and study of hiding information in plain sight. Knowledge and experience with steganography greatly influence the protection and use of digital information. Undoubtedly, digital steganography enhances privacy by providing an additional layer of hidden communication but does not replace encryption [5]. Sadiq Sly A.I. Their method has been extensively tested against attack scenarios, showing that the algorithm provides stronger protection and better stealth management when applied to grayscale images [6].

On the other hand, anti-judgment techniques including audio and visual steganography were proposed. The Forbidden Zone Data Hiding (FZDH) algorithm is used to hide the hidden image in the video file, and the phase coding algorithm is used to hide the information in the audio file with the receiver and transmitted PSNR and histogram matching function types meet to maintain safety [7]. Yada, Rambabu Mudusu, A. Nagesh, M.K. Sdanandam presented the hybrid image steganography and audio steganography techniques. The Least Significant Bit (LSB) method was used to encrypt the customer's facial image in the video and in addition RSA algorithm was used to encrypt this encrypted information. Principal component analysis (PCA) was used to identify faces, and a face validation process was performed to confirm the presence of protective mechanisms [8].

2.1 Methods of steganography types

Steganography is the ability to hide a file, message, image, or video inside another file, message, image, or video, it is a way to hide an audio file in video where you would have to convert the video file so that it is audio no file is included as hidden data. There are many steganography techniques and techniques that can be used to embed a file as a transport or overlay. Fig.2 shows

Figure 2. Types of Steganography

2.1.1 Video steganography

Video steganography is one of the best ways to hide data. Essentially this extends the concept of encrypting an image for video files to be like a collection of images that provides video data, which is equivalent to overlaying image data as a way to encrypt a message in the video. Video-dependent steganography is preferable to various other multimedia files due to memory and file space limitations. where it stores only certain files in the video file [9]. There are two basic approaches to video-based steganography: storing frame-to-frame data and translating the frame into the frequency domain and then storing the result. Regardless of the method used, video steganography can be divided into two categories: lossless and lossless steganography.

While we can accurately retrieve hidden information in lossy steganography, the original video may exhibit flaws or alterations. In contrast, in lossless steganography, both the hidden information and the original video file can be recovered without any errors or modifications [10]. With the rapid development of the Internet and multimedia technologies in recent years, digital video has become an important area for covert data encryption in addition to infinite video sequences providing greater redundancy space for encoding hidden messages for video steganography [11,12]. There are various techniques for video steganography

a. Transform field techniques: In transform field steganography, hidden data is stored in transform coefficients when the video is transformed to other fields, such as frequency or wavelet domain Discrete Cosine Transform (DCT) and Discrete Wavelet Transform are two transforms that are commonly developed for video steganography (DWT).
b. Encryption-based techniques: Some video steganography techniques encrypt the hidden data before storing it in the video. It can be more difficult for an attacker to find and extract private information as a result.
c. Hiding in a compressed area: This method uses the compressed format of the video instead of the raw pixel values in the movie. Most videos use the H.264 format, which is the popular format. This method can be used to change the title, motion vector, quantization coefficient, and other important properties in a compressed video.
d. Hybrid methods: These combine various steganography techniques for increased security and flexibility.

2.1.2 Audio steganography

To perform audio steganography, we can insert hidden messages as digital sound. Because the human ear (HAS) is more sensitive than the human visual system (HVS), storing communication in audio files is more complex than in other file types [14]. There are three important elements of audio steganography: the bearer (which can be an audio file), the message, and the password. These features work together to inject hidden information into the audio. Carrier is another name for a cover-file, which stores private data.

Recent studies have shown that data can be stored inaudibly in envelope audio signals. Audio steganography inserts a hidden message into a digitized audio signal, resulting in a small change in the binary sequence of the audio file corresponding to the hidden message [15] Figure 3 General audio steganography techniques.

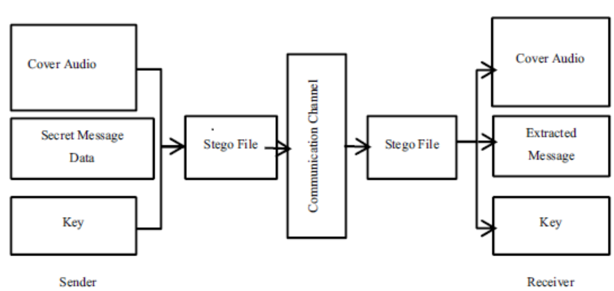

Figure 3. Method of general audio steganography [16]

Hiding in audio (or other media) means experiencing data as noise. The disadvantage of this approach is that data loss compression methods eliminate most undetectable distortions, including moderate low-dB noise. Below is a list of the most common methods used for processing audio files.

a. Least Significant Bit (LSB): Technology is a technology in which the lowest significant pixel or audio file in the pixel is replaced by replacing the private message in the image or audio file with a lsb (least significant bit) code 1 KBps at every 1 KBps in lsb coding in the minimum minutes of the least bits Replacing bits lsb steganography is a simple and effective method of hiding a message in a cover file, such as an image or audio file. without significantly changing the format or quality of the cover file [14].
b. spread spectrum: It is the minimum required to send a broad message. In spread spectrum steganography, the message is encrypted by noise and added to the foreground image, forming a stego image. The associated image is difficult to see with the naked eye [13]. Spread spectrum (SS) is the concept of a redundant copy of the data signal to ensure the recovery of the transmitted signal in the noise mode, using an M-sequence code known as the transmitter and receiver [17]. it was especially plentiful before it was hidden in cover audio. So even if the noise corrupts some values, each target image will still be available to extract the hidden message. While [18] used the traditional direct sequence spread spectrum (DSSS) method to encapsulate sensitive data in MP3 and WAV files.
c. Echo hiding: One method of encrypting information in an audio stream is a form of data encryption. This is intended to be loud without significantly affecting the host signal (cover audio). Echo concealment is more sensitive compared to many data loss suppression methods because it produces changes in cover audio that reflect environmental conditions rather than random noise [19].

As well as there are other techniques to hide audio like Parity Coding, Direct Sequence Spread Spectrum, and Phase Encoding.

3. Methodology

This method of embedding audio into videos first converts the audio to text using a powerful sequential conversion process. The resulting information is then securely encrypted using the RC4 algorithm before being converted to binary and inserted into the least significant bits of the selected video frame using LSB channel, ensuring minimal visual distortion Time after extraction the text will be converted back to original audio, text- speech -A converter model is used, which is capable of instant voice cloning. This allows the audio data to be encoded and retrieved through careful manipulation of voices including emotion, sound and pitch. This process is illustrated in the following Fig. 4.

3.1 Speech Recognition using encoder-decoder Transformer

This work used a sequential Transformer model for speech recognition to convert audio files to text as the first step for Audio file embedding in Video Steganography using LSB method [26], [27]. The Robust Speech Recognition via Large -scale weak supervision brings a new language to recognition approach that Capitalizes on large-scale poorly monitored data, which differs from traditional methods that rely entirely on written training. This system makes efficient use of raw or semi-encrypted data, providing scalability and cost-effectiveness. Taking advantage of significant data sets of anonymous speech and minimal labeled data, the model achieves state-of-the-art, successful performance in harsh environments such as noise and echo The system robustness and scalability make it a promising solution for diverse linguistic knowledge applications. The method shown in Fig.5 illustrates this approach in detail, using a series-to-sequence transformer model trained on different speech processing tasks to build several steps of a traditional speech processing pipeline is replaced by one example.

3.2 The encryption/decryption phase

In this research, the RC4 algorithm is used as the main cryptographic tool [27, 28]. Recognized as the most developed software stream cipher for modern applications and widely used in various industries, RC4 has key cryptographic features called Key Scheduling Algorithm (KSA) and Pseudo-Random Generation Algorithm (PRGA). -The enduring popularity of RC4 in the community can be attributed to its simple- yet robust design, making it desirable for a variety of security applications The details of the KSA and PRGA algorithms are shown in the following Fig. 6.

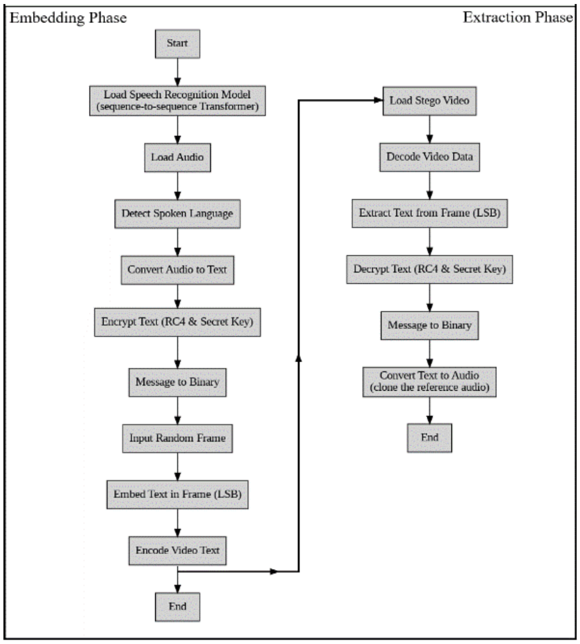

Figure 4. The methodology of the Cryptographic Embedding of Audio Data within Video Streams

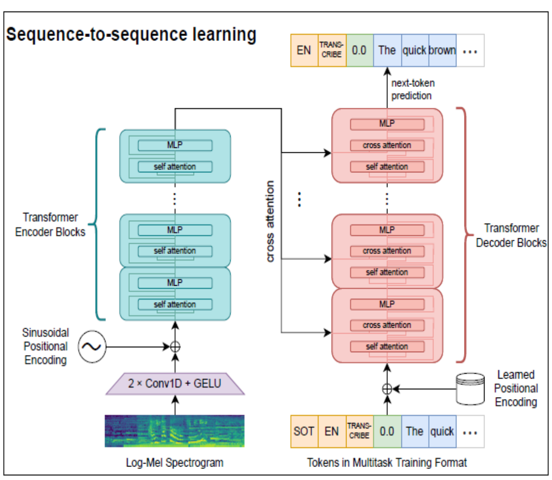

Figure 5. Speech Recognition using a sequence-to-sequence Transformer model [26]

Algorithm KSA
Initialization:
For $i = 0, \ldots, N-1$
$S[i] = i$;
$j = 0$;
Scrambling:
For $i = 0, \ldots, N-1$
$j = (j + S[i] + K[i])$;
$\mathrm{Swap}(S[i], S[j])$;

Algorithm PRGA
Initialization:
$i = j = 0$;
Output Keystream Generation Loop:
$i = i + 1$;
$j = j + S[i]$;
$\mathrm{Swap}(S[i], S[j])$;
$t = S[i] + S[j]$;
Output $z = S[t]$;

Figure 6. The details of RC4 Algorithm (KSA and PRGA) [27]

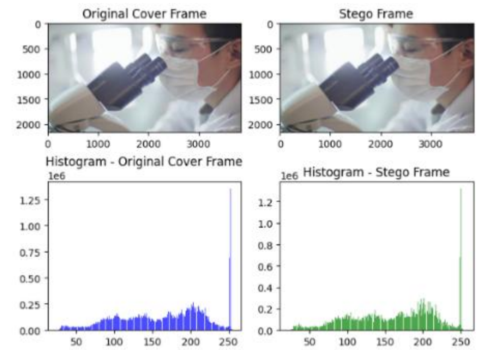

Figure 7. The example of the output frames

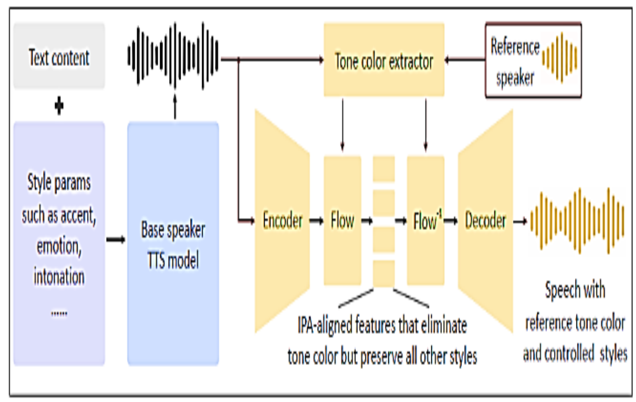

Figure 8. A visual representation of the Open Voice framework [29].

3.3 Least significant bit (LSB) technique

This processing step uses the Least Significant Bit (LSB) method to encode information into the front-end video. The LSB method involves replacing each pixel's red, green, blue, and relatively trivial RGB values in the selected frame of the video to essentially encode the binary representation of the encrypted data, using the Whisper model (respectively). to-sequence Transformer example). The transcribed audio data is encrypted before being converted to binary form and this binary data is then inserted into the less important RGB pixel values of the video frames. The LSB conversion is invisible to the human eye, ensuring that it is not visible in the front-end video. The process ends with the creation of a “stage video” file (steganography video) that stores encrypted information in the LSBs of the video frames. This robust approach allows for covert communication using video overlays and visual effects, an example of output frames featuring a combination of audio text transcription, encryption, and LSB steganography for securing the fuzzy data embedding is shown in the following Fig. 7.

3.4 A Transformer-based instant voice cloning framework

This paper used Open Voice [29], a simple open-source model, and converted output text (generated by the decoding process) to the original input audio, which is a pioneering instant voice cloning technique in text language using an efficient Transformer system. Transcends traditional techniques by utilizing the transformation capabilities of Open Voice Transformer programs, efficiently converting output text from clone voice decoding programs to original input audio Addressing critical challenges in the field, Open Voice empowers users to deliver carefully controlling voices is not past the limit of phonological character cloning More impressively, the model achieves zero-shot cross-linguistic voice cloning without the need for extensive multilingual training data sets, an impressive improvement due to Transformer versatility turns out, serving as the backbone of the MyShell.ai instant voice cloning backbone and demonstrating its prowess in a variety of real-world applications Fig. 8 provides a visual representation of the Open Voice framework. In this case, an original speaker model is used to control both form and speech, while a transformer is used to retain the phonological characteristics of the reference speaker in synthetic speech.

4. Results and discussion

4.1 Speech recognition using encoder-decoder transformer

The speech recognition phase of the proposed method used a sequence-sequence transformer model trained on a large dataset of audio transcripts from the Internet. Extensively studied in previous research, this model scaled to 680,000 hours of multilingual multitasking observations was able to It is worth mentioning, that these models have previously shown competition with fully observed results, even at zero-shot transfer settings, eliminating the need for fine-tuning the close-to-human accuracy and complexity of the prototype highlights how audio files are converted to text. This study on a large and heterogeneous data set establishes the sequence transformer as a robust solution for the speech recognition task and the models and heuristics presented form a valuable foundation for analysis in complex language processing. This comprehensive review of the speech recognition model sets the stage for the integration in the subsequent stages of the proposed method, contributing to the success of the overall audio-text transcription process in video steganography in the system.

4.2 Encryption/Decryption and data embedding phases

The encryption and decryption part of the proposed method used the well-established RC4 algorithm, a stream cipher known for its simple and robust Key Scheduling Algorithm (KSA) and Pseudo-Random Generation Algorithm (2009). PRGA) are key factors that contributed to the enduring popularity of RC4 in the cryptographic community for a more detailed description of the presentation, see Figure 6, which shows the RC4 algorithm and its KSA and PRGA methods.

At the same time, the data embedding phase used the Least Significant Bit (LSB) method to seamlessly embed the encrypted information into the masked video frame This changes the least significant bits in RGB values in each pixel in the selected frame to encode the binary representation of the encrypted data LSB conversions in progress subtly ensured that there was little visual impact on the video in front, and culminated in the creation of a stage video file. An example of the output frames obtained by this steganographic method is shown in Figure 6.

4.3 Evaluation on UCF101 dataset and quantitative metrics

For a more comprehensive analysis, the study used the UCF101 data set [30], which is known to contain a variety of human activity data [31,32]. A strategic selection of 40 random videos from this data set provided a robust testing ground for testing the proposed methods. The integration of realistic user-uploaded videos with the UCF101, including dynamic camera movements and chaotic backgrounds, helped generate a follow-up evaluation of the model's performance complete in different circumstances. In addition, the proposed steganography method was quantitatively evaluated in 40 random videos from the UCF101 dataset. Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) were calculated for each video before and after the steganographic data embedding process.

Table 1. Effectiveness of the results

Video no	PSNR	SSIM	Video no	PSNR	SSIM
1	41.70107	0.99670	21	42.80049	0.99699
2	43.17023	0.99617	22	43.15288	0.98796
3	40.65104	0.99362	23	44.16538	0.99717
4	42.74064	0.99333	24	41.98324	0.99764
5	43.02833	0.99650	25	42.07368	0.99773
6	41.27595	0.99571	26	42.32896	0.99802
7	43.03433	0.99582	27	39.99358	0.99234
8	42.79062	0.99533	28	41.92000	0.99481
9	42.16290	0.99626	29	41.50550	0.99517
10	40.08923	0.99261	30	41.45172	0.99283
11	41.53259	0.99318	31	42.83902	0.99641
12	42.69118	0.99606	32	42.30342	0.99649
13	43.40457	0.99619	33	39.78585	0.98679
14	42.29921	0.99716	34	42.31341	0.99720
15	42.43825	0.99647	35	43.04999	0.99804
16	40.42876	0.99120	36	42.80855	0.99763
17	40.79143	0.98925	37	41.27898	0.99870
18	41.27899	0.99267	38	41.49838	0.99880
19	41.72762	0.98958	39	41.40537	0.99882
20	42.84116	0.99709	40	43.23830	0.98966

4.4 Length consistency measurement

Corresponding lengths were measured to ensure the accuracy of the data. This required the length of the original text (before the steganographic process) and the length of the text obtained after the procedure. Notably, the number of leaves remained the same in both cases, providing confidence in the safe input and output of the video images This convergence builds according to the proposed method is strong in emphasizing the basic information storage of the steganographic system.

5. Conclusion

The research methodology, in conclusion, provides a complete and practical method for securely and discretely encapsulating audio data in videos. Framework Text-to-speech synthesis for audio reception, encryption using the RC4 algorithm, sequence-to-sequence converter models for speech recognition, and data encoding using LSB techniques. The evaluation on the UCF101 dataset demonstrates the effectiveness of the method in properly encoding information and maintaining video quality. Quantitative measures, including PSNR and SSIM, confirm the robustness of the method in preserving visual fidelity. Furthermore, the stability of text length before and after the steganography processing again verifies the reliability of the method in preserving the original data during insertion and deletion. Future research could focus on ways to improve the proposed method. First, exploring additional encryption algorithms or combinations of encryption methods can help make data security more robust. Moreover, based on sensory and visual characteristics, machine learning algorithms can be trained to identify optimal placements for video images this approach can lead to more efficient and distinctive audio data storage.

Authors' contribution

All authors contributed equally to the preparation of this article.

Declaration of competing interest

The authors declare no conflicts of interest.

Funding source

This study didn't receive any specific funds.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] M. M. Amin, M. Salleh, S. Ibrahim, M. R. Katmin and M. Z. I. Shamsuddin, "Information hiding using steganography," 4th National Conference of Telecommunication Technology, 2003. NCTT 2003 Proceedings., Shah Alam, Malaysia, 2003, pp. 21-25. https://ieeexplore.ieee.org/document/1188294

[2] K. N. Bangera, N. S. Reddy, Y. Paddambail and G. Shivaprasad, "Multilayer security using RSA cryptography and dual audio steganography", in 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2017. https://ieeexplore.ieee.org/document/8256645

[3] El-Khamy, S.E., Korany, N.O. & El-Sherif, M.H. "A security enhanced robust audio steganography algorithm for image hiding using sample comparison in discrete wavelet transform domain and RSA encryption". Multimed Tools Appl 76, 24091–24106 (2017). https://doi.org/10.1007/s11042-016-4113-8

[4] M. M. S. ,. M. A. R. Sazeen T. Abdulrazzaq, "A novel steganography approach for audio files," SN Computer Science, vol. 1, pp. 1-13, 2020. https://doi.org/10.1007/s42979-020-0080-2

[5] D. Artz, "Digital steganography: hiding data within data.," IEEE Internet computing, vol. 5, no. 3, pp. 75-80, 2001. https://doi.org/10.1109/4236.935180

[6] S. A. M. H. A.-H. A. M. N. Al-Taweel, "Image in image steganography technique based on arnold transform and LSB algorithms," International Journal of Computer Applications, vol. 181, no. 10, pp. 32-39, 2018. https://www.ijcaonline.org/archives/volume181/number10/altaweel-2018-ijca-917652.pdf

[7] H. L. A. W. O. Nyo, "Secure data transmission of video steganography using Arnold scrambling and DWT," International Journal of Computer Network and Information, vol. 9, no. 6, p. 45, 2019. https://doi.org/10.5815/ijcnis.2019.06.06

[8] R. A. N. M. S. Mudusu, "Enhancing Data Security Using Audio-Video Steganography," International Journal of Engineering & Technology, vol. 7, no. 2.20, pp. 276-279, 2018. https://doi.org/10.14419/ijet.v7i2.20.14777

[9] G. K. ,. P. S. Shivani Gupta, "Video Steganography Using Discrete Wavelet Transform and Artificial Intelligence," International Journal of Trend in Scientific Research and Development (IJTSRD), vol. 3, no. 4, pp. 1210- 1215, May-Jun 2019. https://www.ijtsrd.com/papers/ijtsrd25067.pdf

[10] R. Patel and M. Patel, "Steganography over video file by hiding video in another video file, random byte hiding and LSB technique," 2014 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, India, 2014, pp. 1-5. https://doi.org/10.1109/ICCIC.2014.7238343

[11] S. L. Y. W. H. Z. S. L. Yunxia Liu*, "Video Steganography: A Review," Neurocomputing, vol. 335, pp. 238-250, 2019. https://doi.org/10.1016/j.neucom.2018.09.091

[12] A. S. K. ,. M. G. M. M. Mennatallah M. Sadek, "Video steganography: a comprehensive review," Multimedia Tools and Applications, vol. 74, pp. 7063-7094, 2015. https://doi.org/10.1007/s11042-014-1952-z

[13] T. W. P., R. E. S. Aryfandy Febryan, "Steganography Methods on Text, Audio, Image and Video: A Survey," International Journal of Applied Engineering Research, vol. 12, no. 21, pp. 10485-10490, 2017. https://www.ripublication.com/ijaer17/ijaerv12n21_04.pdf

[14] S. B. Navneet Kaur*, "Audio Steganography Techniques-A Survey," Journal of Engineering Research and Applications, vol. 4, no. 6, pp. 94-100, June 2014. https://doi.org/10.1007/978-981-10-3773-3_56

[15] D. A. A. Eng. Saad Nasser Al.Azzam, "Survey steganography applications," Al-Salam Journal for Engineering and Technology, vol. 2, no. 1, pp. 69-75, 2023. https://doi.org/10.55145/ajest.2023.01.01.008

[16] M. M. S. ,. M. A. R. Sazeen T. Abdulrazzaq, "A novel steganography approach for audio files," SN Computer Science, vol. 1, pp. 1-13, 2020. https://doi.org/10.1007/s42979-020-0080-2

[17] B. A. K. A. M. H. H. Fatiha Djebbar, "Comparative study of digital audio steganography techniques," EURASIP Journal on Audio, Speech, and Music Processing, vol. 1, pp. 1-16, 2012. https://doi.org/10.1186/1687-4722-2012-25

[18] R. V.-M. ,. L. N. d. R. ,. P. S. Hernandez-Garay, "Steganographic communication channel using audio signals," in 12th International Conference on Mathematical Methods in Electromagnetic Theory, Odessa, UKraine, 2008. https://doi.org/10.1109/MMET.2008.4581017

[19] A. L. W. B. Daniel Gruhl, "Echo Hiding," First International Workshop Cambridge, UK, pp. 296-315, 1996. https://doi.org/10.1007/3-540-61996-8_48

[20] J. E. ,. M. O. T. Morkel, "An Overview of Image Steganography," InISSA, vol. 1, no. 2, pp. 1-11, 2005. https://martinolivier.com/open/stegoverview.pdf

[21] Cheddad, A., Condell, J., Curran, K., & Mc Kevitt, P. (2010). "Digital image steganography: Survey and analysis of current methods". Signal processing, 90(3), 727-752. https://doi.org/10.1016/j.sigpro.2009.08.010

[22] M. M. J. B. J. F. T. K. Ingemar Cox, "Digital watermarking and steganography". Morgan kaufmann., 2007. https://papersim.com/wp-content/uploads/Ref.pdf

[23] P. P. ,. P. J. V. ,. B. H. Inas Jawad Kadhim, "Comprehensive survey of image steganography: Techniques, Evaluations, and Evaluations, and trends in future research," Neurocomputing, vol. 335, pp. 299-326, 2019. https://doi.org/10.1016/j.neucom.2018.06.075

[24] R. S. ,. Z. S. ,. K. H. Mohammed Abdul Majeed, "A Review on Text Steganography Techniques," Mathematics, vol. 9, no. 21, p. 2829, 2021. https://doi.org/10.3390/math9212829

[25] D. B. M. Y. Yee Lip Por, "Information hiding: A new approach in text steganography," in In Proceedings of the International Conference on Applied Computer and Applied Computational Science, World Scientific and Engineering Academy and Society (WSEAS 2008), Hangzhou, China, 2008. https://www.academia.edu/1491402/Information_Hiding_A_New_Approach_in_Text_Steganography

[26] Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023, July). "Robust speech recognition via large-scale weak supervision". In International Conference on Machine Learning (pp. 28492-28518). PMLR. https://doi.org/10.48550/arXiv.2212.04356

[27] Basu, R., Ganguly, S., Maitra, S., & Paul, G. (2008). "A complete characterization of the evolution of RC4 pseudo random generation algorithm". Journal of Mathematical Cryptology, 2(3), 257-289. https://doi.org/10.1515/JMC.2008.012

[28] Jindal, P., & Singh, B. (2015). RC4 encryption-A literature survey. Procedia Computer Science, 46, 697-705. https://doi.org/10.1016/j.procs.2015.02.129

[29] Qin, Z., Zhao, W., Yu, X., & Sun, X. (2023). "OpenVoice: Versatile Instant Voice Cloning". arXiv preprint arXiv:2312.01479. https://doi.org/10.48550/arXiv.2312.01479

[30] Soomro, K., Zamir, A. R., & Shah, M. (2012). "UCF101: A dataset of 101 human actions classes from videos in the wild". arXiv preprint arXiv:1212.0402. https://doi.org/10.48550/arXiv.1212.0402

[31] Cho, Hyunjong, Hyungtae Lee, and Zhuolin Jiang. "Evaluation of LC-KSVD on UCF101 action dataset." THUMOS: ICCV workshop on action recognition with a large number of classes. Vol. 7. 2013. https://www.crcv.ucf.edu/THUMOS14/index.files/NotebookPapers13/Paper%2026%20(Supplementary).pdf

[32] Ramesh, M., & Mahesh, K. (2019). "Sports video classification with deep convolution neural network: a test on UCF101 dataset". International Journal of Engineering and Advanced Technology, 8(4S2), 2249-8958. DOI: 10.35940/ijeat.D1007.0484S219