

New Algorithm to Enhance the Accuracy of Extracting Steganography Hidden Data

Nourah Alamri

College of Computing and Informatics
Saudi Electronic University
Riyadh, Saudi Arabia
g200000505@seu.edu.sa

Ahmed Abu-Khadrah

College of Computing and Informatics
Saudi Electronic University
Riyadh, Saudi Arabia
abosuliman2@yahoo.com

Abstract— Cybercriminals are employing various techniques to conceal evidence from investigators, allowing them to avoid tracking the traces of the attack or the traces of crimes. Steganography of information was techniques and tactics used to hide the traces of a hacking or electronic attack. Steganography is one of the most severe methods of obfuscating traces to make it harder for investigators to uncover reliable evidence that can be used in court. In this research study, the problem was that the steganography tools that the authors used in previous studies in their research were not accurate in extracting all the hidden data, and their efficiency was poor. The objectives of this research are to evaluate the accuracy of extracting the hidden data by creating different scenarios using python script. Furthermore, develop a new algorithm to enhance the accuracy of extracting veiled data by using Tkinter framework. Finally, to evaluate the performance of the proposed new algorithm by comparing the proposed algorithm with different steganography tools. The proposed algorithm was able to increase the accuracy by 90% and extract the hidden data compared with different tools such as openstego stegspy and stegovirtas.

Keywords- Anti-forensic techniques, Steganography techniques, Steganography detection, Digital investigator.

I. INTRODUCTION

With the growth and development of computer crime and the increasing use of counter forensic tools that overlap and conflict with forensic investigations, various authors have discussed different counter forensic techniques, tools, and methodologies. This chapter reviews the literature and background study of steganography Background.

The various types of crimes and criminal acts include information and communication technology. The widespread use of computers and digital devices without protection can lead different parties to commit crimes [1]. Punishment of criminals based on evidence and digital criminals frequently employ anti-forensic techniques to make it difficult for forensic investigators to collect evidence. Data hiding has made effective developments through continuous research over many years along with the development of Internet technology, and it is one of the most important areas of research that still has a lot of interest [2]. Also, data hiding has many methods in transmitting digital content to multimedia such as audio, video and text [3]. Some of the anti-forensic techniques a malicious attacker uses:

- Cryptography is a widely used anti-forensic technique that includes confidential and sensitive information in a

ciphertext. A malicious attacker uses full-size encryption and a basic file to hide the code or his negative campaigns. Cryptography is a method that helps only the intended recipient read the message and de-blur it [4].

- Overwriting data is one technique attackers use to circumvent forensic investigations and reduce digital footprints so that investigators have difficulty finding forensic evidence. There are many tools that attackers use to overwrite important text, metadata, or entire media on a storage system, which hinders the task of the forensic analyst during the recovery phase.
- Steganography is the process of hiding a secret message inside an audio, video, image, or text file in an unsuspecting manner. Anonymization is often combined with encryption to provide an extra layer of security. Attackers use this technology to hide information from a digital forensic investigator [5].
- Watermarking in the case of watermarking, it is important to hide confidential information in the form of a digital sign, since the watermark is immutable [6]. The

main objective of the digital watermark is to provide copyright authentication and copyright protection [7].

Steganography is a Greek term that translates as "concealed writing." Johannes Trithemius used it for the first time 500 years ago in 1499 in his *Steganographic*, a dissertation on cryptography and Steganography disguised as a book on magic. Steganography is a technique for concealing a message or a file, typically by altering the file's appearance. The steganography method distorts the original images after the hidden data is extracted [8]. It is a well-known technique in anti-forensic operations to disguise files as operating system files within the operational system tree structure to avoid detection. Criminals use Steganography to discourage forensic investigators from collecting and hiding data so it is inaccessible [9].

Steganography is divided into two techniques; Linguistic Steganography and Technical Steganography. [10]. Steganography of text, it is a relatively frequent kind of Steganography in which the information to be concealed is contained within a text file. Its significance has waned in the aftermath of the internet and the proliferation of digital file types. Text stenography with digital files is not often used because text files contain a negligible amount of extra data.

Steganography of video, it is a mechanism for concealing any file with any extension embedded within a carrying Video file. Steganography of protocol introduces data into network protocols such as TCP/IP. The information will be concealed in many fields of the TCP/IP packet's header section that are either optional or never utilized.

Steganography of audio, the method of hiding audio information is hiding facts inside an audio signal. Although embedding hidden messages in analog audio is a more complex method than embedding messages in various data with virtual images, it is essential to have instructions and routines that limit access to audio documents and their security. For example, information is entered into audio recordings to secure copyright or to confirm calculated means; in the framework of PC-based sound masking, the hidden message is prepared with an automatic sound [11].

Steganography of digital images, A digital image is a visual representation of an entity. It is an artifact consisting of several image elements called pixels, and each has its digital presentation that explains its luminosity and is stored electronically [12].

The authors in [13] had discussed how to reveal the concealment of digital information. They also stated that the attacker employs this technique to conceal information, such as files, video clips, and messages hidden within other images. As a result, they are not revealed by the forensic investigator while collecting evidence using various tools and the theory of revealing the hidden message inside steganography files. It used

Python code that ran in real-time and detected changes to the file. The goal of this tool is to fetch the absolute path of changed files. This experiment was carried out on a JPG image file, and a tool called Stegovirtas was used in the Docker container to determine whether or not the image contained Steganography. The fact that the tool was good enough to detect Steganography was an advantage of this researcher's experiment.

Exploring Steganography is a task that is not difficult for a digital forensic investigator if he has the knowledge and the right way to gather evidence against malicious attackers. Research conducted by [1] developed five-stage application was used within the framework of the criminal investigation model, which consists of stages: pre-processing, acquisition, preservation, analysis, and presentation; in addition to that, some tools such as FTK Imager, Autopsy, Hideman, and StegSpy have been used to analyze the Steganography of digital evidence. The digital evidence in this paper was obtained through the case scenario results. The research was conducted through the framework of the General Computer Forensic Investigation Model (GCFIM), which contains five stages. The researcher sees the success of implementing the framework, as he was able to find the secret message that was entered through the technique of hiding information in the form of hidden text and the accuracy of the Hideman tool in extracting evidence effectively.

Yari & Zargari has developed steganography tools, like S-tool and OpenStego; his empirical research focused on the use of steganography applications that use the same algorithms to hide information exclusively within an image. The search results were that S-tool in Steganography could not extract data encrypted by OpenStego even though it shares standard features and techniques. Some possible errors appeared: the file may be corrupt, which is not possible because S-tool was able to decrypt the message, or OpenStego may see that the file may be evil; in addition, it cannot be possible that the password used is invalid because it was the same password used for the encryption method. A possible error is an algorithm used that may cause a problem. However, these two tools use the Least Significant Bit substitution method. The results indicate that these tools use a different approach to choose the Least Significant Bit in the substitution method, perhaps random, the last two numbers, or just the last number. OpenStego, which used the same procedure, was unable to recover the embedded image, traces of corruption were challenging to identify, and the method of concealment was unknown [14].

In [13],[1], and [14] the drawback is that the tools they used could not detect all the evidence and were not accurate and effective. It was difficult for the author to extract evidence more efficiently.

This study aims to analyze the attacks that are implemented through several different methods of hiding information. In this

research, the researcher mentioned some techniques and defensive measures that can help and mitigate the harm caused by the malicious attacker. The author [15] mentioned non-digital methods where physical steganography techniques such as human eye blinking in Morse code to spell a secret message are not omitted and developed for centuries. Also, one of the methods is to have a panel switch so that any panel consists of pairs of indices and each pixel in the image corresponds to a specific index in the table, and the color sequence in the palette is not required. So, altering the series allows you to hide a hidden message. This research showed that Steganography is a powerful way to protect information, but it should be considered that it can be a tool used by the attacker. The emission compromising the varying physical nature is invisible and cannot be observed unless specific instruments are used.

Parallelism is a heterogeneous Steganography is a new trend in contemporary streaming media audio steganography, which conceals secret information within the streaming media frame using multiple orthogonal steganography types. Because of the complexity and non-awareness of HPS, detecting it presents a challenge in steganography analysis; therefore, [16] designed a faster and more efficient detection method called Key Feature Extraction and Fusion (KFEF) network based on the attentional mechanism, which is more flexible in the modeling sequence than Recurrent Neural Networks (RNNs). Whereas, Recurrent Neural Networks (RNNs) and Key Feature Extraction and Fusion Network (KFEF) were implemented by using Keras platform, and the researcher applied Steganalysis Feature Fusion Network (SFFN) on the PyTorch platform. For SFFN, Adam's optimizer was used to train the network, and it trained all networks with a GeForce GTX 1080 Ti GPU.

The model he proposed can extract the main feature of exceptions very effectively due to Steganography and incorporation of extracted features of various steganography algorithms used in HPS. The experimental results showed that the proposed researcher's method helped significantly improve the classification accuracy in detecting low imputation rates and short section samples. The disadvantage is that if an intangible action is detectable in the packet-length frequency pattern extracted from these methods, the attacker may be able to discover the current techniques by utilizing this flaw.

In this paper [17] they classify the network steganography method based on packet length by using the idea of embedding the message in different sizes of successive packet pairs. In his research, he tested the proposed method against the current attack of random network type traffic and for various cases of message distribution over the entire reference traffic. The results of the test carried out by the researcher were that in the case of random reference traffic with two different strategies for distributing the message over the entire reference or at its beginning, this proposed method has excellent resistance against

the current attack. The drawback is if an intangible action is detectable in the packet-length frequency pattern extracted from these methods, possibly the attacker can discover the existing ways using this defect.

The study's goal [18] was to develop a method of discovering network steganography through the use of deep learning technology. The advantage of this study conducted by the researcher was found several techniques to hide the network. The results of this study appeared through the use of deep learning technology in training other models to support the classification process. The results proved that this method has scientific meaning and many practical meanings. This suggestion proposed by the researcher helped to show better results than other models at all different scales. The researcher points out that it is possible to think of improving the ability to detect abnormal bundles in the future. The drawback is there was no improvement in the ability to see strange packets.

Melo et al. proposed an undetectable scheme by increasing the complexity of data hiding in the TCP, the initial sequence number, using dynamic identifiers. To complicate the concealment of information from malicious attackers and the complexity of their access to critical data. This research showed that the proposed scheme outperformed the current system with a simple detection accuracy of 0.52%. Seven different classifications were used to compare the proposed hidden Initial Sequence Number results through the dynamic identifier and the original secret channel [19]. The drawback is the proposed solution was used, but the accuracy of the tool was not substantial, it revealed a meager rate of 0.52%

Brodzki & Bieniasz propose a new method of network masking by using TCP retransmission, which was implemented by creating a secret communication channel in the client-server model. Statistical analysis was conducted to demonstrate the technique's great advantages in detection. The drawback is to need to further research methods of detection and defense against advanced mechanisms used by malicious attackers in their attack [20].

The authors [21] improve the validity and efficiency of the image data hiding approach. An approach to data masking is proposed by producing various transformed face images from different datasets. After this stage, a face image that is encrypted with a secret message is sent to the receiver. The receiver uses compelling deep learning models to retrieve the private message by recognizing the parents of the distorted face images. In addition, the researcher has designed two new architectures for a convolutional neural network (CNN) (such as MFR-Net V1 and MFR-Net V2) to perform face recognition variably and achieve high accuracy compared to other networks. Experimental results are generated by MATLAB and are implemented by Pytorch, and an NVIDIA RTX 3090 GPU is used for acceleration. The results that appeared in the research

were that the proposed scheme is characterized by high retrieval capacity, high accuracy, and good durability. The drawback is that the research lacks development and no better representation of the secret message. Various types of mutant face images of the same parents are used, and this misses the study from improvement in aspects of parameter-based face transformation and face alignment.

Several authors have studied and discussed counter forensic techniques. Also, the authors reviewed and discussed some methodologies, tools, and approaches, ranging from using a few other tools to detection methodologies for counter-forensic methods. However, although many authors have discussed anti-forensic tools, they have not been able to effectively and accurately collect digital evidence, which will be discussed in this research by using proposed algorithm to enhance the accuracy of steganography tools to help the forensic investigator conduct investigations more effectively.

Computer criminals have come to use their ways of committing a crime, and digital forensics has been instrumental in catching those criminals who believe their heinous act will leave no evidence behind. Unfortunately, however, these criminals have become familiar with computer forensic tools that make recovering evidence for a forensic investigator difficult or nearly impossible.

Also, cybercriminals are more familiar with digital forensics tools and techniques, which results in more sophisticated anti-forensics methods. With the rising usage of anti-forensics tools and tactics, digital forensics investigators find it more challenging to conduct their investigations efficiently. Anti-forensics seems to be a substantial impediment to digital forensics, and the issue is rising exponentially. To overcome this expanding barrier, digital forensics investigators must stay current on anti-forensics tools, tactics, and remedies.

Investigators must be vigilant for evidence distorting techniques such as zero-foot printing, data concealing, etc. All these instances necessitate some type of obfuscation, which should include the removal of a criminal's trail stamp.

One of the significant problems the digital investigator encounters is Data concealment using cryptography, which conceals the data by giving it cryptic codes and Steganography as an additional method of data concealment.

The problem was that the steganography tools that the authors used in previous studies in their research were not accurate in extracting all the veiled data, and their efficiency was poor.

In this paper a new algorithm is developed to enhance the accuracy of extracting veiled data by using Tkinter framework. In addition to that, the performance of the proposed algorithm is evaluated by comparing the proposed algorithm with different steganography tools.

II. PROPOSED MODEL

In this research, a scenario is created to explain the mechanism of hiding the secret message inside the image and the importance of hiding information. Through the proposed scenario, several tasks can be performed such as hiding multiple format files inside an image, hiding a picture within a picture and hiding a secret message inside a picture and then extracting these hidden files through steganography tools. After creating the scenario, a new algorithm is created through Python and compares its result with the previous studies. Figure 1 shows the overall methodology for this research.

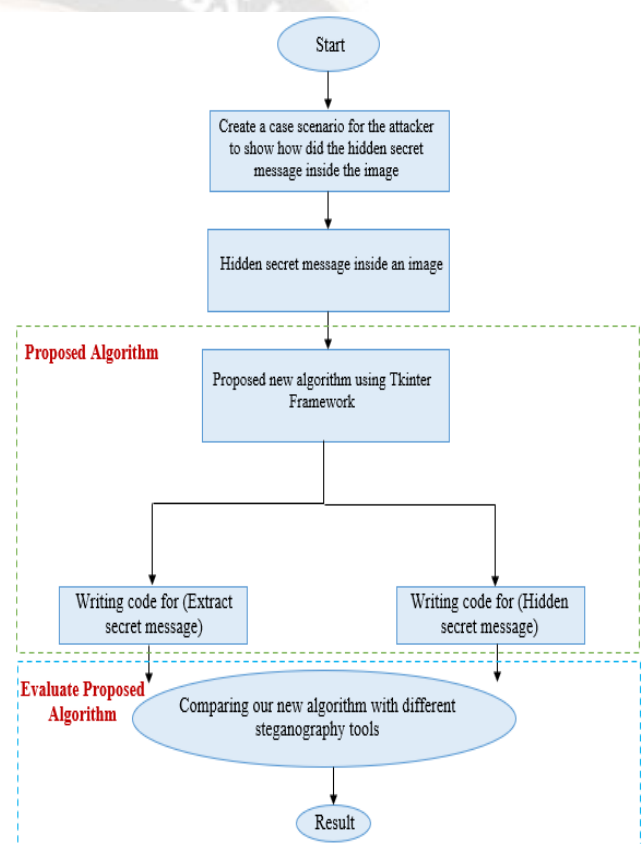


Figure 1: Overall methodology

The attacker broke into the IT company. The FBI found something had happened; then he wrote notes. The FBI called a forensic investigator to investigate, extract the hidden secret message and collect forensic evidence that will help arrest this hacker, as shown in Figure 2. As in this scenario, the attacker broke into the IT company. The FBI found something had happened; then he wrote notes. The FBI called a forensic investigator to investigate, extract the hidden secret message and collect forensic evidence that would help arrest this hacker.

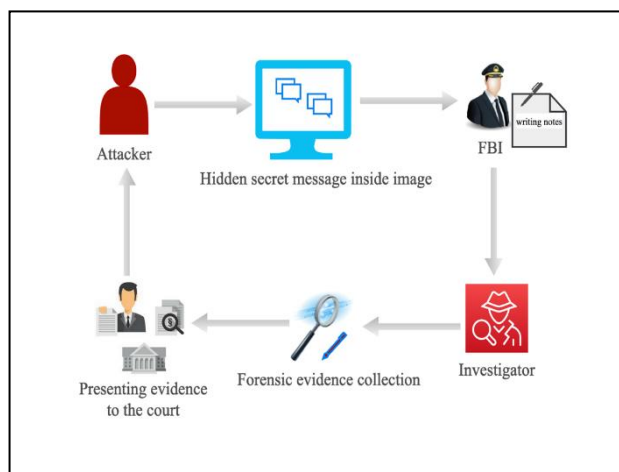


Figure 2: Proposed Case Scenario

In this research, the proposed new algorithm was created by the Python Tkinter framework. Python has a lot of GUI frameworks, but Tkinter is the only framework that's built into the Python standard library. Tkinter has several strengths. It's cross-platform, so the same code works on Windows, macOS, and Linux. Visual elements are rendered using native operating system elements, so applications built with Tkinter look like they belong on the platform where they're run. However, Tkinter is lightweight and relatively painless to use compared to other frameworks. This makes it a compelling choice for building GUI applications in Python. Stegano tool is used to hide the secret message inside an image via Python script. The last stage in the proposed algorithm is to show the results, after extracting the secret message hidden inside the image by takinter and Stegano tool, then the result of proposed algorithm is compared with previous studies conducted by other researchers. The comparison is by using the previous authors' tools such as StegSpy and OpenStego, which could not reveal all the hidden data, and then compared it with the proposed algorithm.

III. RESEARCH RESULTS AND DISCUSSIONS

With crimes increasingly occurring in cyberspace, it was necessary to build a new field called computer forensics for methodically and technically gathering electronic evidence, investigating, and presenting results to law authorities' enforcement to prove the crime. While significant work is being done in computer forensics, one of the primary challenges now facing the industry is ensuring the integrity and trustworthiness of digital evidence obtained during forensic investigations. At every point of the analysis, there is frequently a danger of digital evidence being altered or tampered with. This issue is referred to as anti-forensics since it casts doubt on the legitimacy of digital evidence.

The various types of crimes and criminal acts include information and communication technology. The widespread

use of computers and digital devices without protection can lead different parties to commit crimes. Punishment of criminals based on evidence and digital criminals frequently employ anti-forensic techniques to make it difficult for forensic investigators to collect evidence. Some of the anti-forensic techniques a malicious attacker uses: steganography is the process of hiding a secret message inside an audio, video, image, or text file in an unsuspecting manner. Anonymization is often combined with encryption to provide an extra layer of security.

In the Figure 3, after writing the code to hide the secret message and the image to hide the secret message inside, a new image called 1.png in which the secret message is located. But when clicking on it, the secret message does not appear, this confirms that the proposed algorithm was successfully implemented in hiding the secret message effectively and accurately.

First, prepare the environment and the editor is used to build the Tkinter framework by using Code Editor PyCharm to write python code for hidden secret message inside an image and extract hidden this secret message. The proposed algorithm consists of two parts, the first part is to write a code to hide the secret message inside the image, and the second part is to extract the secret message from the image. The results show image and secrete message as shown in Figure 4. The username and password are extracted from the image. This indicates that the proposed algorithm was highly effective in extracting the secret message from the image.

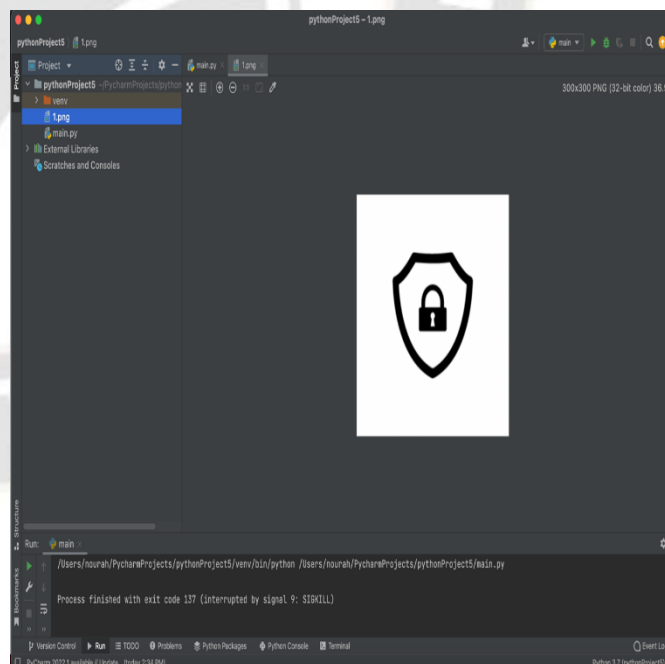


Figure 3: Extract result after hidden secret message

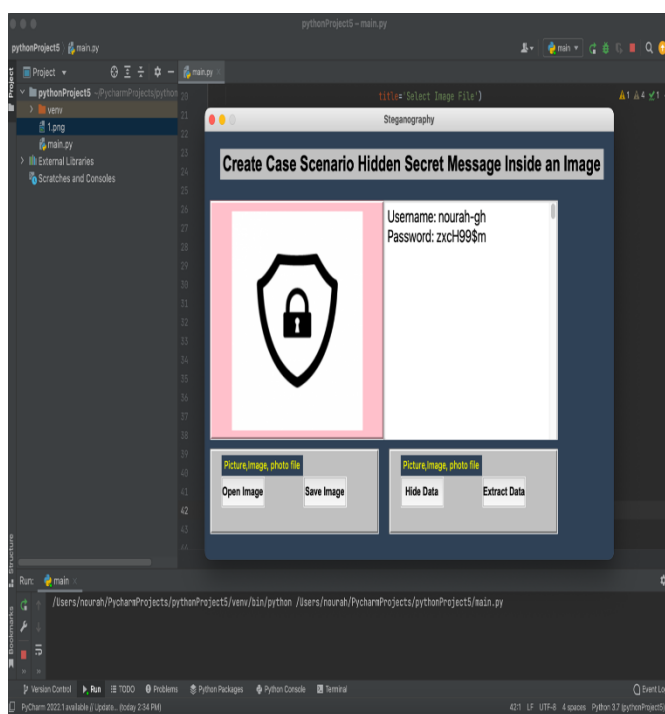


Figure 4: Extract final result after hidden secret message using Tkinter

In this research, a new algorithm is created through Python and the results are compared with the previous studies. Several authors have studied and discussed counter forensic techniques. Also, the authors reviewed and discussed some methodologies, tools, and approaches, ranging from using a few other tools to detection methodologies for counter-forensic methods. However, although many authors have discussed anti-forensic tools, they have not been able to collect digital evidence effectively and accurately, which are discussed in this research by using our proposed algorithm to enhance the accuracy of steganography tools to help the forensic investigator conduct investigations more effectively.

After using the proposed algorithm with Tkinter tools, started to compare the results with other researchers and show that the proposed algorithm was better, and the results are more accurate and effective. The comparison is created by using nine images. Table I shows the results of proposed algorithm and other tools which are used by other previous study.

Table I: Compare proposed algorithm result with other previous study.

Researcher	Accuracy	Tool
Proposed Algorithm	90 %	Tkinter framework
[14]	70 %	Openstego
[1]	0 %	Stegspy
[13]	0 %	Stegovirtas

IV. CONCLUSIONS

Steganography of information is the techniques and methods used to hide the effects of hacking or cyber-attack. One of the most dangerous techniques of antiquities obfuscation is steganography, making it difficult for investigators to uncover

reliable evidence that can be used in court. Computer criminals have come to use their ways of committing a crime, and digital forensics has been instrumental in catching those criminals who believe their heinous act will leave no evidence behind. Unfortunately, however, these criminals have become familiar with computer forensic tools that make recovering evidence for a forensic investigator difficult or nearly impossible. In this research study, the problem was that the steganography tools that the authors used in their previous studies in their research were not accurate in extracting all the hidden data, and their efficiency was poor. The objectives of this research are to evaluate the accuracy of extracting hidden data by creating different scenarios using a Python script. Moreover, developed a new algorithm to enhance the accuracy of extracting the veiled data using the Tkinter framework. The result of this research was that the proposed algorithm was able to increase the accuracy by 90% and extract the hidden data from 9 different images 9 times more accurate and effective. Also, this result was better than other researchers, first researcher who used the first tool openstego which gave 70% accuracy and second researcher used stegspy tool which had 0% accuracy and could not extract hidden data from different images also third researcher he used stegovirtas tool, but this tool could not detect all file types and did not have strong security.

ACKNOWLEDGMENT

Thanks to ALLAH whose help us in this work and thanks to the Saudi Electronic University for sponsoring this work.

REFERENCES

- [1] M. Hajar Akbar, U. Ahmad, D. Yogyakarta, and I. Sunardi, "Analysis of Steganographic on Digital Evidence using General Computer Forensic Investigation Model Framework," IJACSA) International Journal of Advanced Computer Science and Applications, vol. 11, no. 11, 2020.
- [2] C. Kim, C.-N. Yang, J. Baek, and L. Leng, "Survey on Data Hiding Based on Block Truncation Coding," Applied Sciences, vol. 11, no. 19, 2021, doi: 10.3390/app11199209.
- [3] P. Maniriho, L. Jovial Mahoro, Z. Bizimana, E. Niyigaba, and T. Ahmad, "Reversible difference expansion multi-layer data hiding technique for medical images," International Journal of Advances in Intelligent Informatics, vol. 7, no. 1, pp. 1–11, 2021, doi: 10.26555/ijain.v7i1.483.
- [4] S. Haimour, M. AL-Mousa, and R. Marie, "Using Chaotic Stream Cipher to Enhance Data Hiding in Digital Images," International Journal of Advanced Trends in Computer Science and Engineering, vol. 9, no. 4, 2021.
- [5] P. Rajba and W. Mazurczyk, "Information Hiding Using Minification," IEEE Access, vol. 9, 2021.
- [6] K. Joshi, R. Saini, R. Yadav, and R. Nandal, "A New Blended Approach of Data Hiding Using XNOR Operation," Journal of Scientific Research, vol. 66, no. 2, p. 2022, 2022, doi: 10.37398/JSR.2022.660216.

- [7] J. Bala, S. Rai, and R. Shweta, "Robust Digital Watermarking for Digital Images Based On Dwt-Svd," ROBUST DIGITAL WATERMARKING FOR DIGITAL IMAGES BASED ON DWT-SVD International Journal of Economic Perspectives, vol. 16, no. 3, pp. 87–97, 2022.
- [8] S. Kumar and M. Rai, "A Review of Enhanced Reversible Data Hiding on Encrypted Images Using Selective Pixel Flipping Method," INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY & SCIENCE, vol. 9, no. 5, 2021.
- [9] M. Dalal, M. and Juneja, "Steganography and Steganalysis (in digital forensics): a Cybersecurity guide," a Cybersecurity guide. Multimedia Tools and Applications, vol. 80, 2021, doi: 10.1007/s11042-020-09929-9.
- [10] A. Yahya, Steganography techniques for digital images. Springer International Publishing, 2018.
- [11] S. Mishra et al., "Audio Steganography Techniques: A Survey," In Advances in Computer and Computational Sciences: Proceedings of ICCCS 2016, vol. 6, 2018, doi: 10.1007/978-981-10-3773-3_56.
- [12] K. Farhan Rafat, "Nondeterministic Secure LSB Steganography for Digital Images," 2020, doi: 10.1109/ICCWS48432.2020.9292369.
- [13] A. Hammoudi, "Detecting Digital Steganography," journal of Palestine Technical University, vol. 1, Jan. 2020.
- [14] I. Yari and S. Zargari, "An Overview and Computer Forensic Challenges in Image Steganography,," IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2017. .
- [15] S. Romanova, A., & Toliupa, "Steganography methods used in attacks on information and communication systems," Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 8, 2018.
- [16] H. Wang, Z. Yang, Y. Hu, Z. Yang, and Y. Huang, "Fast Detection of Heterogeneous Parallel Steganography for Streaming Voice,," in Proceedings of the 2021 ACM Workshop on Information Hiding and Multimedia Security, 2021, p. 1255.
- [17] V. Sabeti and M. Shoaie, "Network Steganography Based on PVD Idea,," in International Conference on Computer and Knowledge Engineering (ICCKE), 2018, pp. 1332–1345.
- [18] C. Do Xuan and L. Van Duong, "A New Approach for Network Steganography Detection based on Deep Learning Techniques," IJACSA) International Journal of Advanced Computer Science and Applications, vol. 12, no. 7, p. 2021, 2021.
- [19] P. M. B. Melo, A. M. Sison, and R. P. Medina, "Enhanced TCP Sequence Number Steganography using Dynamic Identifier,," in 2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), 2019, p. 1233.
- [20] A. M. Brodzki and J. Edrzej Bieniasz, "Yet Another Network Steganography Technique Based on TCP Retransmissions," in In 2019 5th International Conference on Frontiers of Signal Processing (ICFSP), 2019, p. 1443.
- [21] Y.-H. Li, C.-C. Chang, G.-D. Su, K.-L. Yang, M. S. Aslam, and Y. Liu, "Coverless Image Steganography Using Morphed Face Recognition based on Convolutional Neural Network," EURASIP Journal on Wireless Communications and Networking, 2022, doi: 10.21203/rs.3.rs-981238/v1.