DOI: http://dx.doi.org/10.26483/ijarcs.v13i2.6817

Volume 13, No. 2, March-April 2022

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

SECURE DATA HIDING MECHANISM USING TEXT COVER MESSAGE

Hema S
Ph.D Research Scholar (Part Time), PG & Research Department of Computer Science, Govt. Arts College (Autonomous), Salem-7, India

Dr.Kangaiammal A
Assistant Professor, Department of Computer Applications, Govt. Arts College (Autonomous), Salem-7, India

Abstract: Steganography is the art of hiding secret information within other information that is digitally concealed. It protects the confidentiality of two communicating parties. Information hiding is a highly strong and adaptable strategy that may be used to protect vital information in a variety of situations. There are several steganography techniques available, each with its own set of advantages and disadvantages. This study presents a new data hiding mechanism that hides file information in text cover messages to safeguard user data from vulnerabilities. It also utilize a deduplication strategy that employs the suggested DSHA algorithm to reduce repeated data storage. Experimental results reveal that the proposed data hiding mechanism using the DSHA reduces the time it takes to save the file and also reduces the time to retrieve it. Furthermore, indicating that this technology will be a good alternative for Steganography. This novel technique ensures security while maintaining confidentiality.

Keywords: Steganography; Data Hiding Mechanism; Stego-object; Text Cover Message; ECC

I. INTRODUCTION

Data security is one of the most important aspects of today's fast-paced, contemporary technological world. As a result, the data should only be used by authorized individuals and not by any other unauthorized individuals. Nowadays, tens of thousands of messages and bits of data are continually being transmitted from one location to another over the Internet. What is required is that the correct data be sent confidentially to the correct recipient and that the information be understood only by that recipient.

Cryptography and steganography are well-known and commonly utilized approaches in information security. Steganography and cryptography are both critical components of information security [12]. To deliver secret communications, the first cryptographic technology was devised. In cryptography, the message was encrypted in another message in such a way that only the sender and recipient understood how to decode it. Only authorized individuals were able to decode the message using a cryptographic key. The restriction of cryptography is that if attackers find secret content in the communication, the message is more likely to be decoded by them.

Steganography techniques were developed to address the inadequacies of cryptographic systems. Steganography is the technique of hiding information in a closed medium (such as an image, audio, video, or text) and making it invisible. There are more serious concerns about data security. Improved security is required to prevent unauthorized access to data. Taking all of this into account, this paper provides a new data hiding technique that hides file information in text cover messages that are used to protect data.

The rest of the paper is structured as follows: Section 2 provides an overview of Steganography techniques. Section 3 describes the proposed data hiding approach. Section 4 discusses the outcomes of the suggested approach. The conclusion and future work are discussed in the fifth part.

II. STEGANOGRAPHY OVERVIEW

Steganography is a technique for hiding the existence of embedded information and an art of concealing information [1]. It protects the message in a more advanced way than cryptography. Furthermore, it only conceals the message's content, not its existence.

The majority of people nowadays use the medium to transmit data in the form of text, photos, video, and audio. To securely transfer secret data, multimedia items such as audio, video, and images are employed as cover sources[2].

The secret message and cover media are used in steganography. The data that has to be hidden is the message, and the carrier that hides the message is the cover media. It produces Stego-object after the Embedding Process. This Stego-object may be retrieved as the cover media and secret Message using the extracting algorithm.

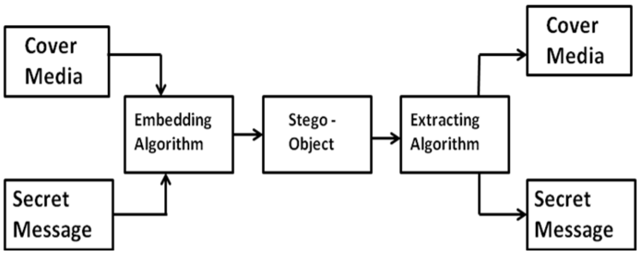

Fig 1. Steganography Diagram [3]

Types of Steganography

There are a variety of stenographic techniques that may be used to obtain security depending on the type of cover medium. [4].

A. Text Steganography

It entails hiding information within text files. There are several methods for embedding secret data in text files.

Because text steganography can only store text files, it uses less memory. It allows for rapid file transfers or communication between computers.

B. Image Steganography

The method of hiding a hidden message in an image file is known as image Steganography. Pixel intensities are a critical aspect in masking data in image Steganography.

C. Audio Steganography

Audio Steganography refers to the process of encoding hidden information into digital sound. The secret message is encoded in sound files such as WAV, AU, and MP3.

D. Video Steganography

Video steganography is a technique for concealing hidden information in a video. When compared to image steganography, video steganography is significantly more secure and efficient.

III. Literature Review

[5] illustrates a basic and new technique to steganography using transliteration. For data hiding strategies, this study uses Bengali digital text. The suggested method's fundamental concept is to take use of a unique property of Bengali phonetic keyboard layouts to hide secret information in the form of bits. The results demonstrate that the proposed method has a lot of potential as a steganography methodology. This approach is easily adaptable and applicable to any language with a non-roman alphabet.

In this paper [6], the author introduced a new method to improve the cognitive-imperceptibility steganography method, which is unable to control the semantic expression of the steganographic text and has potential security risks. To find out how powerful the code is for extracting semantic information, this paper compares three different models: the Cadet Recurrent Unit (GRU) model, the Transformer model, and the Topic-Aware mode. The experimental results suggest that the proposed approaches can improve the produced steganographic text's cognitive-imperceptibility, hence improving the privacy and security of text generative steganography.

The author focused on linguistic-based steganography in this research [7], proposing a new data hiding strategy that uses Unicode, i.e., Zero Width Character (ZWC) and Zero Width Joiner, to conceal mysterious information in the cover text (ZWJ). To hide the hidden bits in the carrier, the proposed approach employs ZWC for non-[1]-connected characters and ZWJ for connected letters. It also guaranteed that the cover text remained undetectable. The simulation results show that the proposed method achieves a high level of cover medium capacity, security, and robustness.

In this paper [8], the secret message is compressed using Huffman coding. Then, for further security, add some additional bits and choose particular lines with unique qualities known as "host lines" or "frequency host lines." One of the most essential features of this suggested system is that the sender and receiver may interact using printed pages alone, eliminating the need to exchange electronic information. In addition, as compared to numerous previous text steganography systems, this suggested method may embed more information in a cover-text while maintaining the same file size for the cover. In addition, the parties communicating in this suggested approach employ a stego-key for further security.

A text steganographic technique based on colour coding, permutation, and numeration systems was suggested in this study [9]. The suggested strategy uses a permutation algorithm for the first method and numeration methods for the second to embed the secret message in the cover text by colouring it. The stego-text is subsequently mailed to the recipient. Their model performs a better hiding process in terms of hiding capability, according to the findings of their results.

This research [10] focuses on hiding information in Joint Photographic Experts Group (JPEG) photos, which are subsequently uploaded and distributed over Face book. Face book’s input file types are tampered with in an attempt to solve the compression issue. Although their concept needs some preparation and testing to verify that lengthy text messages and small image files can be hidden in JPEG cover files and transmitted over Face book.

The proposed method [11] introduces two additional levels of security to the standard LSB steganography. The first level is that because only the green and blue colours are used, instead of three colors red, green, and blue in the standard LSB, and thus increases the complexity of an attacker, when he/she tries to retrieve the secret message. The second level exploits the new bit inversion technique that reverses the bits of the image pixels after applying the standard LSB. From the experiment, the technique proposed in this paper does not affect both the efficiency and quality of the stego-image and improves the level of protection.

Yan et al. [13] presented a deduplication strategy for cloud storage of encrypted data. The goal of this study is to defend data owners' privacy. conventional deduplication algorithms are unlikely to work with encrypted data. To solve this problem, they presented a strategy based on data ownership challenges and Proxy Re-Encryption (PRE) to cope with duplicate encrypted data stored in the cloud. With the help of an authorised party, they used Elliptic Curve Cryptography (ECC) to authenticate data ownership. Even when the data holders are offline, this technique can facilitate data exchange with deduplication and achieve outstanding performance.

Bloom filter-based data deduplication for cloud data storage was introduced by Jang et al. (2018). It is a quick way for detecting and removing duplicate data. A Bloom filter-based data deduplication technique combines a Bloom filter's quick duplication detection with a source-based deduplication mechanism that doesn't require any additional storage. The suggested technique reduced deduplication time when compared to other state-of-the-art current deduplication algorithms, according to the evaluation findings.

In [14], Xu J et al. presented a secure client-side deduplication (CSD) system that strives to preserve users' sensitive data privacy. Not only does this technology protect data from outside threats, but it also protects it from nosy cloud storage servers. They improved the convergent encryption approach and allowed for a one-time leak of a target file before their system started to run. Due to its one-time leaking, this technique may be used to deduplicate files with very low min-entropy.

IV. PROPOSED METHODOLOGY

In the proposed methodology, the data deduplication process is initiated first. To begin with, the Distributed Storage Hash Algorithm (DSHA) is recommended for detecting and removing duplicated data, as well as shortening the hash value and storing it in the appropriate storage node.

Following the deduplication procedure, a novel steganography approach is implemented in two ways based on the above discussion: The data embedding and data extraction techniques of the suggested method are described in depth. The suggested method's framework is depicted in Figure 2.

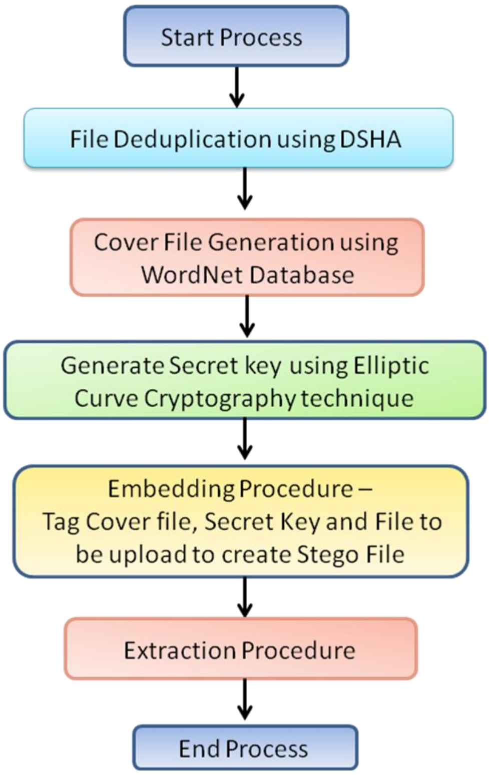
Start Process
File Deduplication using DSHA
Cover File Generation using WordNet Database
Generate Secret key using Elliptic Curve Cryptography technique
Embedding Procedure – Tag Cover file, Secret Key and File to be upload to create Stego File
Extraction Procedure
End Process

Fig 2. Proposed Method

The proposed method described in Fig-2 uses cover text to hide file information. To produce the cover text, the proposed steganography requires pre-define specific parameters. It also generates a secret key to hide the file data, which must be shared with the information extraction algorithm in order to extract the secret data effectively. Following algorithm depicts the suggested method's information hiding algorithm. This approach can produce a high-quality Stego_File with a rapid running time and a good embedding capacity.

Embedding Algorithm:

```
Step 1: Start the process
Step 2: Generate randomly N number of cover words from
            WordNet database. Denote the String Set as
                    SS = {ss1,ss2, . . . ,ssN};

            Set q = 0;
                    while q ≤ N do
            Select randomly the q-th word from WordNet
            database.
            Update String Set (SS) with new cover word
                    q = q +1;
             end while
Step 3: Generate Cover Text File using String Set (SS)
Step 4: Generate a secret key from the cover file that contains
            cover words using Elliptic Curve Cryptography.
Step 5: Attach the cover file to the secret key.
Step 6: Create a Stego-File by attaching the cover file that
            contains the secret key to the file to be hidden.
Step 7: Upload the Stego-File.
Step 8: End the process.
```

Information extraction, in contrast to information hiding, is the retrieval of embedded secret information from a Stego-File. Information embedding and information extraction are essentially the same procedure. For the hiding and extraction processes, they must all utilise the same secret key.

Extraction Algorithm:

```
Step 1: Start the process.
Step 2: Use Stego-File for the Extraction Process.
Step 3: Take the secret key out of the Stego-File.
Step 4: By using a secret key, they extract hidden file data
           from Stego-File.
Step 5: Complete the procedure.
```

V. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we present the experimental results and analysis to verify the performance of the proposed method.

A. Experimental Setup

Table 1. Experimental Setup

Hardware Setup	Processor - Intel(R)Core(TM)i5-5200U CPU @ 2.20GHz, hard disk drive is 1TB, Memory: 8 GB RAM
Software Setup	Operating System : 64-bit Windows 10 home version 1803Programming Environment: Net Beans IDE 7.0.1, Java: 1.7.0; Database Server: 127.0.0.1 via TCP/IP, Software: MYSQL 5.5.27

The effectiveness of the recommended method has been assessed by using Java as a front-end and a MYSQL as back-end. The experimental setup is shown in Table 1. A range of file formats and sizes are used in the experiment.

B. Performance Analysis

To hide the file information, the first step is to generate a cover-text file. In this suggested method, two crucial elements are employed to construct the cover-text file. The WordNet database is one, while the Elliptic Curve Cryptography (ECC) technique is the other. The random number generating method is used to produce cover-words from the WordNet database. The cover-text file is organised around the cover-words that have been generated. The secret key is then produced in order to improve the security of the data that is associated to the cover-text file. The ECC technique is used to construct this secret key, which is based on the cover-words used to build the cover-text. Table 2 summarizes the outcome of the test results.

File to be hidden	Size of the file (KB)	Generated cover file	Secret Key
Comp1	5441KB	surgical procedure of stopping the flow of blood (as with a hemostat)a former French coin of low denomination; often used of any small amount of money circulating printed notices as a means of advertising type genus of the Unavailable; extinct large herbivorous ungulates somewhat resembling elephants; from the Eocene in Wyoming election again an event that might have been arranged although it was really accidental	3046022100da71ee
Visual1	9825KB	a detailed description of design criteria for a piece of work naming explicitly(patent law) a document drawn up by the applicant for a patent of invention that provides an explicit and detailed description of the nature and use of an inventions restriction that is insisted upon as a condition for an agreement	3045022100e98107
prideplus	8921KB	appear at county fairs and carnivals as a stunt flier and parachute jumper our the country making political speeches, giving lectures, or presenting plays a sodium salt of carbonic acid; used in making soap powders and glass and paper sweet drink containing carbonated water and flavouring a belt of parks or rural land surrounding a town or city a stone tool from the Pale age	3045022024a42b77
approval1	9982KB	cloth covering that forms the part of a garment below the waist a garment hanging from the waist; worn mainly by girls and women(Fungi) a remnant of the partial veil that in mature mushrooms surrounds the stem like a collar informal terms for a (young) woman avoid or try to avoid fulfilling, answering, or performing (duties, questions, or issues)pass around or about; move along the border form the edge of extend on all sides of simultaneously; encircle blockage consisting of an object designed to fill a hole tightly a wad of something chewable as tobacco blatant	3045022100a57191
keystore	7546KB	entirely or fully to a suitable or appropriate extent or degree favorably; with approval to a great extent or degree with great or especially intimate knowledge with prudence or propriety with skill or in a pleasing manner in a manner affording benefit or advantage in financial comfort without unusual distress or resentment; with good humor resulting favorably wise or advantageous and hence advisable mite or tick	304502201553dc3c

Table 2. Outcome of the Test Results

To evaluate the efficiency of the suggested strategy, we conducted extensive tests to produce a Stego-File. From the test results, we have confirmed that the proposed method securely hides the file information. The use of a secret key further strengthens the security of this procedure.

Writing Performance:

The writing performance of the proposed algorithm is analysed using DSHA in combination with the Data Hiding Mechanism in the following method. The files to be saved are listed in the table below.

Table 3. Files that are used to store

File Name	File Size	Time taken by Data Hiding Mechanism using SHA-1(ms)	Time taken by Data Hiding Mechanism using DSHA(ms)
Pay1.pdf	855KB	6.0025	5.2187
recreation.pdf	1414KB	5.9023	5.2219
collectionrep.docx	1112KB	5.2401	5.0015
infocollect.pdf	5778KB	6.1732	5.3013
add.docx	6561KB	6.1347	5.1242

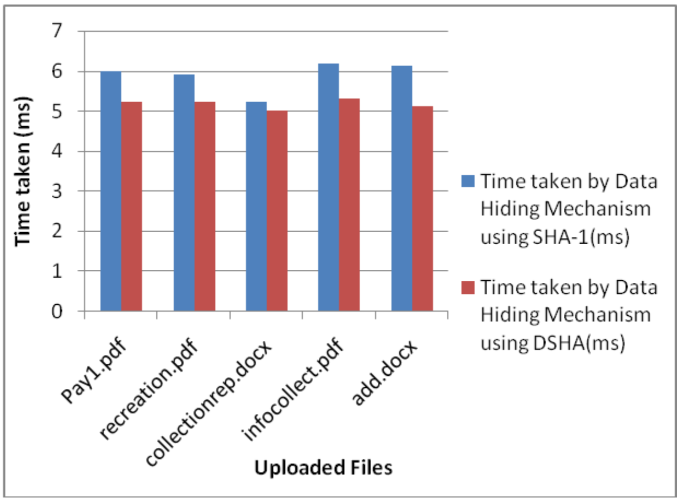

Fig 3. Writing performances of SHA-1 and DSHA algorithms

Figure 3 illustrates the suggested DSHA's optimal performance. DSHA stores each file significantly faster than SHA-1 (as shown in the figure above).

The DSHA algorithm took 56.5915 milliseconds to save 12 distinct files ranging in size from 1000 to 1500 KB, whereas SHA-1 took 67.1648 milliseconds. To store 10 files ranging in size from 2000 to 2500KB, DSHA took 60.6826 seconds.

SHA-1, on the other hand, took 71 seconds to complete. DSHA, on the other hand, took 30.6762 seconds to save 5 files with sizes ranging from 3000 to 3500KB. SHA-1, on the other hand, took 42.7375 seconds to complete. In the same way, files of various sizes were examined. Of all the performance analyses, it is clear that the DSHA data writing process provides better performance than SHA-1 because it makes use of multiple storage nodes.

Reading Performance:

In this process each storage node is identified by unique ID and also the file data is saved at the storage node on the basis of its hash value.

Table 4. Files that are used to download

File Name	File Size	Time taken by Data Hiding Mechanism using SHA-1 (ms)	Time taken by Data Hiding Mechanism using DSHA (ms)
Pay1.pdf	855KB	5.3151	4.3182
recreation.pdf	1414KB	4.8524	3.6019
collectionrep.docx	1112KB	4.7401	3.7125
infocollect.pdf	5778KB	5.3572	3.8013
add.docx	6561KB	5.3362	4.1342

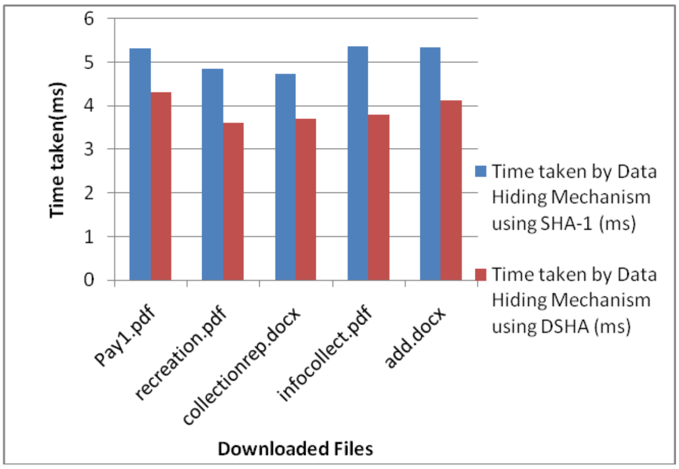

Fig 4. Reading Performance of SHA-1 Algorithm and DSHA Algorithm

When a file is downloaded, the hash value of the file is utilized to identify the proper storage node. Because the storage node is associated with a DSHA hash value, it minimizes the time it takes to retrieve the file, which is faster than SHA-1.

VI CONCLUSION

Over the past few years, steganography has become an emerging technology in data hiding. This paper discussed a novel technique of data hiding mechanism that conceals file information in text cover messages to safeguard user data from vulnerabilities. It also includes a deduplication approach that employs the DSHA algorithm to limit the storage of duplicate data. This innovative method assures security while protecting privacy. The test results obtained for the proposed system are very appealing and provide hope that this method will be a good choice for steganography. As a future work, combining the concepts of cryptography and steganography can provide additional protection for confidential information

VII REFERENCES

[1] Morkel, T., Eloff, J. H., and Olivier, M. S, “An Overview of Image Steganography”, In ISSA, Vol. 1, No. 2, 2005.

[2] Arya, A., and Soni, S, “A Literature Review on Various Recent Steganography Techniques”, International Journal on Future Revolution in Computer Science & Communication Engineering, ISSN: 2454-4248 Vol. 4 Issue: 1,2018.

[3] Alwahbani, Samah, and Elshoush, Huwaida, “Hybrid Audio Steganography and Cryptography Method Based on High Least Significant Bit (LSB) Layers and One-Time Pad—A Novel Approach”, PP. 431-453, 10.1007/978-3-319-69266-1_21,2018,

[4] Dr. Rajkumar L Biradar and Ambika Umashetty , ”A Survey Paper on Steganography Techniques”, International Journal of Innovative Research in Computer and Communication Engineering ,Vol. 4, Issue 1, January 2016.

[5] Khairullah, M. A novel steganography method using transliteration of Bengali text. J. King Saud Univ.-Comput. Inf. Sci. 2019, 31, 348–366.

[6] Yang, Z.; Xiang, L.; Zhang, S.; Sun, X.; Huang, Y. Linguistic generative steganography with enhanced cognitive-imperceptibility. IEEE Signal. Process. Lett. 2021, 28, 409–413.

[7] Allah Ditta, Muhammad Azeem, Shahid Naseem, Khurram Gulzar Rana, Muhammad Adnan Khan, Zafar Iqbal,”A secure and size efficient algorithm to enhance data hiding capacity and security of cover text by using Unicode”,Journal of King Saud University - Computer and Information Sciences,2020.

[8] Behrooz Khosravi, Behnam Khosravi, Bahman Khosravi, Khashayar Nazarkardeh, A new method for pdf steganography in justified texts, Journal of Information Security and Applications, Volume 45,2019, Pages 61-70, ISSN 2214-2126, https://doi.org/10.1016/j.jisa.2019.01.003.

[9] Sadié, Juvet K., Leonel Moyou Metcheka, and René Ndoundam. "Two high capacity text steganography schemes based on color coding." arXiv preprint arXiv:2004.00948 (2020).

[10] J. Hiney, T. Dakve, K. Szczypiorski and K. Gaj, "Using Facebook for Image Steganography," *2015 10th International Conference on Availability, Reliability and Security*, 2015, pp. 442-447, doi: 10.1109/ARES.2015.20.

[11] M. Jain and S. K. Lenka, "Secret data transmission using vital image steganography over transposition cipher," *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, 2015, pp.1026-1029, doi: 10.1109/ICGCIoT.2015.7380614.

[12] Neetha Francis, 2015, Information Security using Cryptography and Steganography, 53 (IJERT) NSRCL – 2015 (Volume 3 – Issue 28).

[13] Z. Yan, W. Ding, X. Yu, H. Zhu and R. H. Deng, "Deduplication on Encrypted Big Data in Cloud," in *IEEE Transactions on Big Data*, vol. 2, no. 2, pp. 138-150, June 1 2016.

[14] Xu, J., Chang, E.C., Zhou, J.: Weak leakage-resilient client-side deduplication of encrypted data in cloud storage. In: 8th ACM SIGSAC Symposium, pp. 195–206,2011.