International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 4 (2025)

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 4, 2025

Research Paper

WEB TEXT–BASED COVERLESS INFORMATION HIDING METHOD

[1]N.SANDHYA, [2]S. RAVALI

[1]Assistant Professor, CSE, Tallapadmavathi College of Engineering, Somidi, Kazipet, Hanumakonda – 506003.Email-id: Sandhya.naddunuri@gmail.com.

[2]Research Scholar, H.no: 24UC1D5812, CSE, Tallapadmavathi College of Engineering, Somidi, Kazipet, Hanumakonda – 506003,Email-id: sarigommularavali85@gmail.com.

ABSTRACT

Coverless information hiding has gained attention for its ability to conceal secret information (SI) in carriers without altering them. To address challenges such as low hiding capacity (HC) and mismatches in large-scale text data, this paper proposes a novel method that leverages massive web text resources. First, web crawler technology is employed to collect web texts related to the SI, forming a comprehensive web-text library. Next, relevant texts containing the SI are retrieved, and the optimal text is selected. The position of the SI within the chosen text is then represented using a 2-D coordinate system. Finally, the URL of the selected web text is combined with the location information and sent to the recipient. Experimental results demonstrate improvements in hiding capacity, success rate, and overall security.

Index Terms: Coverless information hiding, secret information (SI), web text retrieval, hiding capacity (HC), 2-D coordinate system, web crawler, information security, text big data.

Received: 19-09-2025 Accepted: 22-10-2025 Published: 29-10-2025

I.INTRODUCTION

Information hiding conceals data within digital carriers such as text, images, videos, and audio. Among these, text-based hiding is the most widely used and can be categorized into three main types: format-based, image-based, and natural language-based methods. Format-based methods embed secret information (SI) by altering word spacing, inserting invisible characters, or modifying document formats (PDF, HTML, Office) [3]–[5]. Image-based methods treat text as binary images, hiding SI by exploiting image features combined with textual patterns [6], [7]. While these approaches offer high hiding capacity (HC), they modify carriers, making them vulnerable to attacks like recomposition, OCR, and steganalysis [8]–[10]. Natural language-based methods use NLP to generate text carriers containing SI, allowing unmodified hiding, but often suffer from incoherence, poor readability, and deviations from linguistic or statistical norms [11]. To overcome these limitations, coverless text information hiding has been proposed [12]. This method retrieves existing carriers containing SI from large text databases shared between sender and recipient, enabling hiding without modifying the carrier. Techniques such as Word Rank Maps [13], hash-based encoding [14], and character encoding [15] have been developed. However, most of these methods suffer from low HC, often supporting only a single keyword per text. Multi-keyword schemes [16], [17], and word-embedding-based approaches [18] improved HC but face declining success rates (SR) as SI length increases. Web text-based methods [20] simplify big data processing but expose location information due to direct encoding. Existing coverless methods still struggle to balance high HC, stable SR, and

security. To address these challenges, this paper proposes a web-text-based coverless information hiding method. The approach leverages massive web pages to hide SI: secret and search keywords are extracted, relevant web pages are retrieved using search engines, and a web spider constructs a web-text library. From this library, optimal web texts containing secret keywords are selected as stego-texts. The location of each keyword is encoded as 2-D coordinates, which are combined with a key, compressed, and embedded into the URL parameters. The URL is then transmitted to the recipient. This method leaves no modification trace in the web text while achieving high HC and stable SR.

II.RELATED WORK

Information hiding and text steganography have been extensively studied over the past few decades. Cox and Miller [1] provided a comprehensive overview of the first fifty years of electronic watermarking, highlighting the evolution of methods for embedding information in digital media. Bennett [2] surveyed linguistic steganography techniques, discussing robustness concerns when hiding information in text and the challenges of preserving semantic coherence. Brassil et al. [3] proposed copyright protection techniques for electronic distribution of text documents, which laid the foundation for secure text-based information hiding.

In the domain of text watermarking, Liu and Tsai [4] introduced a novel approach for robust watermarking in presentation slides using blank space coloring, demonstrating that text carriers could be subtly modified without affecting readability. Koluguri et al. [5] reviewed various text steganography methods and tools, summarizing format-based, image-based, and natural language-based techniques, along with their respective strengths and limitations. Wu and Liu [6] explored data hiding in binary images for authentication and annotation, which inspired subsequent methods combining textual and image-based carriers for secure information embedding.

Collectively, these works highlight the evolution of methods for hiding information in text, ranging from traditional watermarking to advanced steganographic techniques. While early approaches often relied on modifying carriers—resulting in vulnerability to attacks such as OCR, recomposition, and steganalysis—recent research emphasizes coverless methods that leverage unmodified texts or large-scale text databases to improve security, hiding capacity, and imperceptibility. The proposed method in this paper builds upon these foundations by employing web text retrieval to construct stego-texts without modifying carriers, achieving higher hiding capacity and stability in secret information transmission.

III.EXISTING SYSTEM

In information security, coverless information hiding has gained attention as a promising technique. However, current methods face a key limitation: they typically allow only a single Chinese character to be hidden within each natural text. This severely restricts the overall hiding capacity, making these approaches impractical for real-world applications. While some methods attempt to hide multiple keywords along with their counts to improve capacity, they still exhibit significant drawbacks. These include limited discussion of constraints beyond the single-character limit, unclear practical limitations such as vulnerability to detection or restricted data transmission, and insufficient technical details on the underlying mechanisms. Additionally, the scope of existing systems is often narrow, failing to address trade-offs related to security, text coherence, and resistance to attacks.

DISADVANTAGES

- Existing methods hide only a single character per text, limiting efficiency and practicality.

- Low hiding capacity increases detection risk and restricts data transmission.

IV.PROPOSED SYSTEM

The proposed system presents a novel coverless information hiding method that exploits the vast number of web pages on the Internet to conceal secret information (SI). The process begins by extracting both search and secret keywords, followed by retrieving web pages associated with the SI using mature search engines. Web spider technology is then employed to fetch web texts from these pages, constructing a comprehensive web-text library.

From this library, sets of web texts containing the secret keywords are identified, and the optimal text is selected as the stego-text. The exact location of each secret keyword within the chosen text is represented using a coordinate system. This location information is combined with a key, compressed, and embedded into the parameters of a URL, which serves as the transmission medium.Key advantages of this method include the absence of any modification traces in the web text, ensuring high security and undetectability. By leveraging the immense volume of web data, the system achieves high hiding capacity (HC) and maintains a stable hiding success rate (SR). The use of coordinates and compression enables efficient embedding of secret keywords within URLs, enhancing steganographic efficiency. Furthermore, the integration of search engines and web spider technology ensures the system is practical, scalable, and suitable for real-world applications.

ADVANTAGES

- The carrier medium remains unchanged, ensuring high security and undetectability.
- Utilization of the vast web data significantly increases the hiding capacity (HC).

V.SYSTEM MODEL

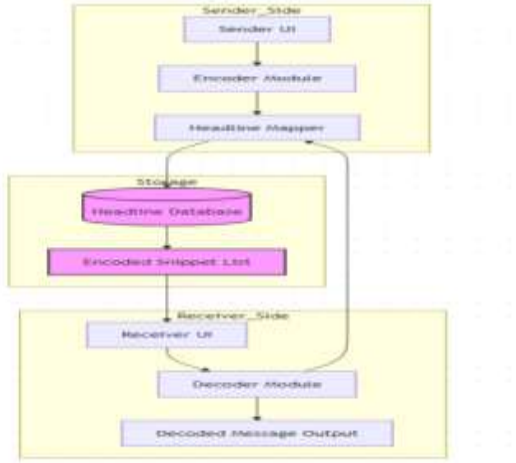

Fig.System Model

VI.IMPLEMENTATION

The system is structured into several interdependent modules to ensure secure and efficient coverless information hiding. The WebText Corpus Generator collects clean and diverse sentences from the web, which are stored in the Mapping Database linking each sentence to specific binary sequences. The Encoding Engine converts secret messages into natural language sentences, while the Decoding Engine retrieves the hidden data accurately. A Synonym and Paraphrase Expansion Module increases variability, making the messages appear natural and unpredictable. The User Interface Module, built with Flask, allows users to encode and decode messages easily, with real-time feedback and basic authentication. Finally, the Security Module protects the mapping dictionary and ensures resistance against detection and steganalysis.

VII.RESULTS

Fig a: Login Page

Fig b: Sender Pane

Fig c: receiver decryption interface

Fig 4 : transmission decoded

VIII.CONCLUSION

In conclusion, the WebText-based coverless information hiding technique offers a significant advancement in secure and covert communication. By mapping secret information to naturally occurring web sentences without modifying any data, it overcomes many limitations of traditional steganography. The system ensures message integrity, readability, and strong resistance to detection, while providing applications in journalism, intelligence, corporate security, and censorship circumvention. Despite challenges like mapping management and scalability, future improvements with AI-driven language generation and cryptographic integration promise enhanced robustness and adaptability. This approach represents a practical and innovative solution for privacy and security in today's heavily monitored digital environment.

IX.FUTURE SCOPE:

The WebText-based coverless information hiding system has promising future directions, including integrating advanced NLP and deep learning models like GPT and BERT to generate diverse, contextually relevant sentences, enhancing stealth. Expanding the corpus to support multiple languages can enable global covert communication and increase resistance to detection. Adaptive learning techniques could allow the system to evolve with user feedback and language trends, optimizing encoding and decoding efficiency. These advancements would further improve the system's robustness, scalability, and applicability in real-world scenarios.

REFERENCES

1. I. J. Cox and M. L. Miller, "The first 50 years of electronic watermarking," *EURASIP J. Adv. Signal Process.*, vol. 2002, Dec. 2002, Art. no. 820936.
2. K. Bennett, "Linguistic steganography: Survey, analysis, and robustness concerns for hiding information in text," Dept. Linguistic, Purdue Univ., West Lafayette, IN, USA, CERIAS Tech. Rep. 13, 2004.
3. J. T. Brassil, S. Low, and N. F. Maxemchuk, "Copyright protection for the electronic distribution of text documents," *Proc. IEEE*, vol. 87, no. 7, pp. 1181–1196, Jul. 1999.
4. T.-Y. Liu and W.-H. Tsai, *Robust Watermarking in Slides of Presentations by Blank Space Coloring: A New Approach*, Lecture Notes in Computer Science, vol. 5510, Springer-Verlag, Berlin, 2009, pp. 49–64.
5. A. Koluguri, S. Gouse, and P. B. Reddy, "Text steganography methods and its tools," *Int. J. Adv. Sci. Tech. Res.*, vol. 2, no. 4, pp. 888–902, 2014.
6. M. Wu and B. Liu, "Data hiding in binary image for authentication and annotation," *IEEE Trans. Multimedia*, vol. 6, no. 4, pp. 528–538, Aug. 2004.

7. Z. H. Xia, S. H. Wang, X. M. Sun, and J. Wang, "Print-scan resilient watermarking for the Chinese text image," *Int. J. Grid Distrib. Comput.*, vol. 6, no. 6, pp. 51–62, 2013.
8. X. G. H. Luo, "A steganalysis method based on the distribution of space characters," in *Proc. IEEE Int. Conf. Commun., Circuits Syst.*, vol. 1, Jun. 2006, pp. 54–56.
9. H. Kwon, Y. Kim, and S. Lee, "A tool for the detection of hidden data in Microsoft compound document file format," in *Proc. Int. Conf. Inf. Sci. Secur. (ICISS)*, Jan. 2008, pp. 141–146.
10. L. Goyal, M. Raman, P. Diwan, and M. K. Vijay, "A robust method for integrity protection of digital data in text document watermarking," *Int. J. Innov. Res. Sci. Technol.*, vol. 1, no. 6, pp. 14–18, Nov. 2014.
11. P. Wayner, *Disappearing Cryptography: Information Hiding: Steganography Watermarking*, 2nd ed., San Francisco, CA, USA: Morgan Kaufmann, 2009.
12. X. Chen, H. Sun, Y. Tobe, Z. Zhou, and X. Sun, "Coverless information hiding method based on the Chinese mathematical expression," in *Proc. 1st Int. Conf. Cloud Comput. Secur.*, Nanjing, China, 2015, pp. 133–143.
13. J. Zhang, J. Shen, L. Wang, and H. Lin, "Coverless text information hiding method based on the word rank map," in *Proc. Int. Conf. Cloud Comput. Secur.*, Nanjing, China, Nov. 2016, pp. 145–155.
14. J. Zhang, Y. Xie, L. Wang, and H. Lin, "Coverless text information hiding method using the frequent words distance," in *Proc. 3rd Int. Conf. Cloud Comput. Secur.*, Oct. 2017, pp. 121–132.
15. X. Chen, S. Chen, and Y. Wu, "Coverless information hiding method based on the Chinese character encoding," *J. Internet Technol.*, vol. 18, no. 2, pp. 313–320, Mar. 2017.
16. Z. Zhou, Y. Mu, N. Zhao, Q. M. J. Wu, and C.-N. Yang, "Coverless information hiding method based on multi-keywords," in *Proc. Int. Conf. Cloud Comput. Secur.*, Nanjing, China, Nov. 2016, pp. 39–47.
17. Z. Zhou, Y. Mu, C.-N. Yang, and N. Zhao, "Coverless multi-keywords information hiding method based on text," *Int. J. Secur. Appl.*, vol. 10, no. 9, pp. 309–320, 2016.
18. L. Xiang, J. Yu, C. Yang, D. Zeng, and X. Shen, "A word-embedding-based steganalysis method for linguistic steganography via synonym substitution," *IEEE Access*, vol. 6, pp. 64131–64141, 2018.
19. X. Chen and S. Chen, "Text coverless information hiding based on compound and selection of words," *Soft Comput.*, pp. 1–8, Jun. 2018.
20. S. Shi, Y. Qi, and Y. Huang, "An approach to text steganography based on search in Internet," in *Proc. Int. Comput. Symp. (ICS)*, Dec. 2016, pp. 227–232.
21. R. Mihalcea and P. Tarau, "TextRank: Bringing order into text," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2004, pp. 404–411.
22. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013. [Online]. Available: https://arxiv.org/abs/1301.3781
23. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.
24. Sogou Corpus, 2016. [Online]. Available: http://download.labs.sogou.com