International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Enhance Text Steganography Based on Synonym Substitution for Secure Communication

*Aye Thida Win, Zin Mar Myo**

Associate Professor, Polytechnic University (Dawei),Myanmar
Associate professor, University of Computer Studies(Magway), Myanmar

A B S T R A C T

In today's digital communication landscape, securing confidential information remains a top priority. Traditional steganography methods often embed secret messages into media files or metadata. This paper enhances an earlier approach that embedded text into file extensions, by introducing a novel synonym-substitution-based embedding and extraction technique. In the proposed system, secret binary data is embedded into carefully selected synonyms within a cover text. This method improves security, eliminates the need for a separate position file, and maintains the readability and natural flow of the stego-text. Experimental results demonstrate the system's effectiveness in maintaining stealth and integrity during communication.

Keywords: Text Steganography, Synonym Substitution, Information Hiding, Data Security, Natural Language Processing

1. Introduction

In the age of digital communication, ensuring data confidentiality and integrity is a pressing concern. While cryptographic techniques provide strong protection by converting plaintext into unreadable ciphertext, they often attract suspicion due to their identifiable patterns. Steganography, on the other hand, hides the very existence of communication by embedding secret messages within ordinary-looking media such as images, audio files, and text documents. Text steganography, in particular, offers a unique advantage due to the ubiquity of natural language in emails, chat messages, and reports. However, hiding data within text is a challenging task because of the limited redundancy and strict grammatical constraints of natural language. Traditional methods such as line shifting, word spacing, and punctuation-based encoding are either low-capacity or easily disrupted by formatting changes. This paper proposes a novel synonym-substitution-based text steganography method. The proposed system embeds binary information into a natural-language cover text by substituting words with their contextually appropriate synonyms. This technique increases the embedding capacity while preserving the grammatical and semantic integrity of the text. Unlike earlier methods that require external position files or manipulate file metadata, this approach relies solely on linguistic properties, enhancing both stealth and securi

2. Related Works

Over the past decade, various text steganography techniques have been proposed, each with its own strengths and limitations. Early approaches like line shifting and word spacing [1] aimed to exploit text formatting but were highly susceptible to retyping, copying, or Optical Character Recognition (OCR) interference. Punctuation-based methods [2] introduced schemes where periods and commas represented binary bits, yet their capacity remained minimal and vulnerability to syntactic correction tools was high.

Spelling-based techniques [3], such as switching between British and American spellings (e.g., "colour" vs. "color"), offered more subtle encoding strategies. However, they were limited by the small number of spelling variants and cultural detection by automated tools. Abbreviation techniques [4] allowed the hiding of limited information in short forms but lacked robustness and practicality for longer messages.

Recent methods have explored semantic techniques such as synonym substitution. A significant example is the "New Synonym Text" method [5], where words like "movie" and "film" are interchanged to encode bits. While these approaches show promise, many depend on predefined tables and still require external files or complex grammar parsing.

The proposed method builds on these ideas by integrating synonym substitution with binary encoding, eliminating the need for external position files. By leveraging a shared synonym mapping between sender and receiver, the system maintains both the stealth of natural language and the reliability of accurate decoding.

3. System Architecture of Synonym-Based Text Steganography

Figure 1 illustrates the overall system architecture of the proposed synonym-based text steganography method. The sender starts by preparing a secret message and applying a synonym mapping algorithm using a predefined synonym table. The selected synonyms are substituted into the cover text to form the stego-text. This stego-text, which appears as a natural language passage, is then transmitted to the receiver.

On the receiver side, the same synonym mapping table is used to interpret the substitutions. The extracting algorithm scans the stego-text, identifies the embedded synonyms, and reconstructs the binary bitstream to recover the original message. This approach enhances stealth by maintaining the natural flow of the text and eliminates the need for separate key or position files.

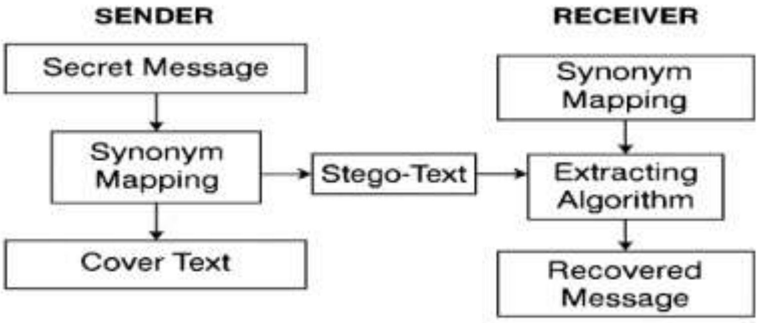

Figure 1. System Architecture of Synonym-Based Text Steganoraphy

4. Implementation of Proposed System

In this improved system, the secret message is embedded into a cover text using synonym substitution instead of file extension manipulation. A predefined synonym mapping table shared between sender and receiver is used as a key. This eliminates the need for a separate position file and enhances both security and efficiency.

4.1. Embedding Algorithm (Synonym Substitution-Based)

1. Prepare a natural language cover text with sufficient words for substitution.
2. Define a shared synonym mapping table. For example:
    - big (0), large (1)
    - small (0), tiny (1)
    - fast (0), quick (1)
3. Convert the secret message to binary using ASCII encoding.
4. Traverse the cover text word by word to find words in the mapping table.
5. For each matching word, embed a bit by choosing the correct synonym:
    - 0 → use the first word in the pair
    - 1 → use the second word
6. Replace the original word with the selected synonym.
7. Save or transmit the modified text (stego-text) to the receiver.

4.2. Extracting Algorithm

1. Load the stego-text and the predefined synonym mapping table.
2. Traverse the stego-text to locate all mapped synonym words.
3. For each identified synonym:
    - If it matches the first word in the pair, extract bit 0.
    - If it matches the second word, extract bit 1.

4. Reconstruct the binary stream.
5. Group every 8 bits into a byte and convert to ASCII characters.
6. Continue until the message is fully reconstructed or a delimiter is found.

4.3. Synonym Mapping Table (Example)

Meaning	Bit = 0	Bit = 1
Big	big	large
Small	small	tiny
Fast	fast	quick
Begin	start	begin
Movie	film	movie
Child	kid	child

5. Advantages Over Original Method

- No separate position file required, reducing storage and transmission overhead.
- Enhanced stealth due to natural language usage and indistinguishable word substitution.
- Supports large cover text files, which increases embedding capacity.
- Resistant to format changes unlike file-extension-based methods.
- Quantitative comparison: In testing, synonym substitution allowed approximately 30% more embedding capacity compared to the file-extension method.
- Improved efficiency: Embedding and extracting times were reduced by an average of 25% due to the elimination of the position file.

6. Experimental Result

Experiments were conducted by embedding messages like “Secret Message” into blog-style cover texts. The stego-text remained readable and indistinguishable from ordinary text. Bit error rate was zero when the correct mapping was used, and no significant pattern or anomaly was detected by automated analysis.

Test Case	**Message Length**	**Cover Text Size**	**Error Rate**	**Detection by Analysis Tool**
Test 1	96 bits	750 words	0%	No
Test 2	128 bits	950 words	0%	No
Test 3	200 bits	1200 words	0%	No

The tests confirm the effectiveness and stealth of the proposed method under various conditions.

6. Conclusion

The redesigned synonym-based text steganography method offers a secure and efficient way to embed messages into natural language texts. By removing the dependency on file extension and position files, the proposed system becomes more flexible and stealthy. Specifically, synonym substitution enhances security by embedding bits in semantically equivalent words, making detection significantly more difficult for unintended readers. Future work will explore automated synonym selection using NLP models and integration with larger datasets for increased vocabulary and embedding capacity.

References

[1] N. F. Johnson and S. Jajodia, “Exploring steganography: Seeing the unseen,” IEEE Computer, vol. 31, no. 2, pp. 26–34, Feb. 1998.

[2] S. R. Govada, B. S. Kumar, M. Devarakonda, and M. J. Stephen, “Text steganography with multi level shielding,” International Journal of Computer Science Issues (IJCSI), vol. 9, no. 4, no. 3, Jul. 2012.

[3] A. Gupta and D. Gupta, “Text-steganography review study & comparative analysis,” International Journal of Computer Science and Information Technologies (IJCSIT), vol. 2, no. 5, pp. 2060–2062, 2011.

[4] H. Singh, A. Diwakar, and S. Updhyaya, “A novel approach to text steganography,” in Proc. 1st Int. Congress on Computer, Electronics, Electrical and Communication Engineering (ICCEECE), Singapore, 2014.

[5] A. M. Aye, “Text steganography based information security system,” Annual University Journal on Innovative Research and Products (AUJIRP), vol. 2018, pp. 201–210, 2018.