Penerbit Akademia Baru

Journal of Advanced Research in Computing and Applications

Journal homepage:
https://akademiabaru.com/submit/index.php/arca/index
2462-1927

Optimising Text Steganography with Alpha-Based Representative Binary: Analysing Contain Letter Used and Capacity Ratio

Roshidi Din[1,*], Sunariya Utama[1], Hrudaya Kumar Tripathy[2], Jabbar Qasim Almalik[3]

[1] School of Computing, Universiti Utara Malaysia (UUM), 06010 Sintok, Kedah Darul Aman, Malaysia
[2] Kalinnga Institute of Industrial Technology, Bhubaneswar, Odisha, India
[3] Ashur University, Department of Medical Instrumentation Technique Engineering College, Baghdad, Iraq

ARTICLE INFO

Article history:
Received 28 June 2024
Received in revised form 18 July 2024
Accepted 18 August 2024
Available online 18 September 2024

Keywords:
Text steganography; feature-based; stego text; contain letter used; capacity ratio

ABSTRACT

Text steganography has become essential for secure communication in the digital age, offering a method to hide messages within seemingly ordinary text. This paper introduces the Alpha-based Representative Binary technique, an advanced feature-based method for text steganography. It utilizes both capital and small letters to represent 2-bit binary values, embedding more information without altering the visual integrity of the text. By assigning different binary values to uppercase and lowercase letters, this technique optimizes the balance between information density and detectability. Embedding 2-bit values strikes an ideal balance, as it allows for a higher capacity of hidden information while maintaining a low profile. This study evaluates the performance of the Alpha-based Representative Binary technique against existing methods like QUAD and One-Flow-2-bit, focusing on two metrics performance: the total character count in the stego text (contain letter used) and the amount of data that can be embedded relative to the cover text size (capacity ratio). The findings underscore the technique's strengths in maintaining secure and undetectable hidden messages within text. This paper highlights the evolution and effectiveness of feature-based methods, emphasizing the need for continued innovation to address detection risks and content modification challenges. By providing a detailed analysis and potential improvements, this paper aims to contribute to the development of more robust and undetectable text steganography systems. The perceptions gained from this paper is expected as valuable for researchers and practitioners in the field of secure communication.

1. Introduction

Text steganography has emerged as a crucial method for secure communication in the digital age. Unlike cryptography, which scrambles the content of a message to make it unreadable without a key, steganography hides the very existence of the message by embedding it within a seemingly innocuous medium, such as text, image, or audio files [1,2]. This subtlety provides an additional layer of security, as the hidden message does not attract attention. Among the various forms of

[*] *Corresponding author.*
E-mail address: roshidi@uum.edu.my (Roshidi Din)

https://doi.org/10.37934/arca.36.1.4351

steganography, text-based approaches are particularly intriguing due to the everyday use and vast ubiquity of text in digital communications [3,4]. The category of steganography is shown in Figure 1.

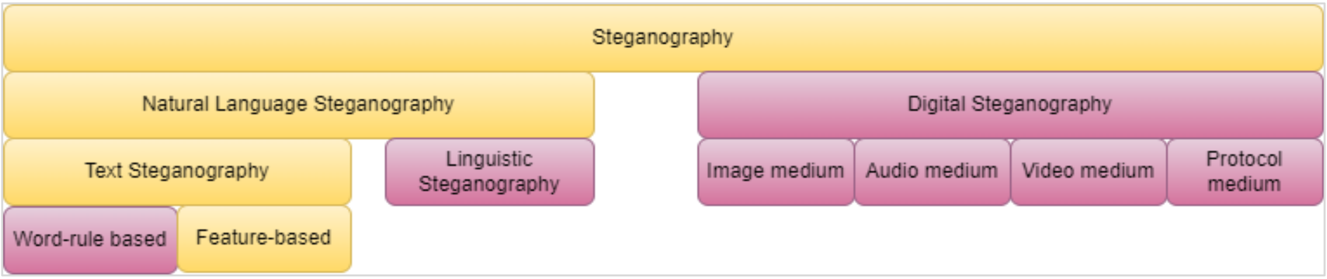

Fig. 1. Classification of steganography field [5]

Based on Figure 1 displays the classification in steganography that consist of digital steganography as non-text domain and natural language steganography as text domain. This paper focus on natural language steganography specifically in text steganography [6]. Then, one of the notable techniques in text steganography is the feature-based method, which leverages the inherent characteristics of alphabet letters to embed hidden messages [7]. This method capitalizes on the visual or structural uniqueness of letters to encode information. For example, letters can be categorized based on their shapes, such as curved or straight lines, and used to represent different binary values [8]. This approach supports in maintaining the cover text's natural appearance while embedding the hidden message, making it less detectable to unintended recipients [9].

This paper proposed the Alpha-based Representative Binary technique as a significant development in feature-based methods for text steganography. This technique utilizes both capital and small letters to represent binary bits, embedding 2-bit binary values. By assigning different binary values to uppercase and lowercase letters, the Alpha-based Representative Binary method can embed more information without altering the visual integrity of the text. This method also employs a secret table to track the positions and representations of letters, further enhancing the concealment process.

Embedding 2-bit binary values is advantageous over 1-bit, 3-bit, or other bit embedding methods. The 2-bit embedding strikes a balance between information density and detectability [10,11]. Embedding only 1-bit per letter limits the amount of information that can be concealed, reducing efficiency [12]. On the other hand, embedding 3-bit or more per letter can significantly alter the cover text's structure, increasing the risk of detection [10]. The 2-bit method allows for a higher capacity of hidden information while maintaining a low profile, thus providing an optimal solution for secure text steganography [13].

The Alpha-based Representative Binary technique it is essential to refine these techniques to ensure the performance against such modifications, enhancing the security and reliability of hidden communications. This evaluation aims to assess the performance of the Alpha-based Representative Binary technique that compare with other techniques of 2-bit implementation in term of contain letter used and capacity ratio performance [4,14-16].

2. Literature Review

The concept of steganography dates back to ancient times, with historical instances of hidden messages embedded in seemingly innocuous objects [17]. In the digital era, steganography has evolved significantly, leveraging advanced techniques to embed messages within digital media. Text steganography, a subset of this field, focuses on concealing information within text documents [14]. Various methods have been developed to achieve this, each with its strengths and weaknesses.

Feature-based methods have been a focal point in text steganography research. Alanazi and Gutub [8] highlighted the importance of using the unique characteristics of letters to hide messages. This approach ensures that the embedded message remains visually inconspicuous, thereby reducing the risk of detection. Majeed *et al.,* [15] supported this view in emphasizing that leveraging the inherent properties of alphabet letters is a key advantage of feature-based methods.

The QUAD technique is one such method that categorizes letters based on their shapes. Letters with curves, horizontal lines, vertical lines, and diagonal lines are assigned different binary values [18,19]. This categorization allows for a systematic approach to embedding hidden messages. Kumar and Pabboju [19] noted that this method effectively balances the distribution of binary bits across different letters, making the hidden message less detectable.

The One-Flow-2-bit technique introduces another layer of complexity by considering the writability of letters in a single flow. Letters are categorized based on their ability to be written without lifting the pen and the presence of vertical or horizontal lines [20]. Azeem *et al.,* [21] discussed the effectiveness of this method in enhancing the concealment of messages, as it takes advantage of the natural flow of writing.

The Alpha-based Representative Binary technique, a more recent development, represents a significant advancement in feature-based methods. This technique's use of both capital and small letters to represent binary bits, adding an extra layer of security. By differentiating between uppercase and lowercase letters, Alpha-based Representative Binary can embed more information while maintaining the natural appearance of the text [15]. This method also employs a secret table to track the positions and representations of letters, further enhancing the concealment process.

However, the effectiveness of text steganography methods, including Alpha-based Representative Binary, is often challenged by the potential for detection through content modification. Alkhudaydi and Gutub [22] pointed out that retyping, editing, or other changes to the cover text could disrupt the hidden message, making it detectable or leading to its loss. It emphasizing the need for robust the development techniques that can withstand modifications in conceal the text.

Furthermore, Yaghobi and Sajedi [7] proposed the use of more appropriated algorithms for letter categorization, aiming to improve the distribution of binary bits and reduce detection risks. Additionally, innovations in managing the secret table used in the embedding process have been suggested to enhance the robustness of the hidden message against modifications.

In short, the literature highlights the importance of developing robust and effective text steganography methods. Feature-based techniques, including the QUAD, One-Flow-2-bit, and Alpha-based Representative Binary methods, have shown promise in concealing hidden messages within text. However, challenges related to detection and content modification remain.

This paper aims to compare the existing technique of 2-bit feature-based technique by providing analysis in term to two metrics of the Alpha-based Representative Binary technique. By comparing it with other feature-based methods and exploring potential improvements in order to contribute to the development still natural appearance and readability of the stego text after embedded by hidden message in the text.

3. Methodology

The methodology of feature-based in text steganography encompasses a range of innovative techniques designed to discreetly embed hidden messages within ordinary text, enhancing secure communication. Collectively, these methods strive to balance the natural appearance of the cover text with the efficiency and security of the embedded information.

3.1 QUAD technique and One-Flow-2-Bit technique

The QUAD and One-Flow-2 -bit technique divides letters into four categories based on their physical characteristics in Table 1.

Table 1
QUAD and One-Flow-2-Bit techniques group classification

Bit	QUAD Technique		One-Flow-2 -bit Technique	
	Type of letters	Letter chosen	Type of letters	Letter chosen
00	Curved letters	C, D, O, Q, S, U	Not writable in one flow, no vertical /horizontal line	Q, X
01	Letters with middle horizontal line	A, B, E, F, H, P, R	Not writable in one flow, with vertical /horizontal line	A, B, D, E, F, H, K, T
10	Letters with one vertical straight line	I, J, K, L, T, Y	Writable in one flow, no vertical /horizontal line	C, G, O, S, V, W
11	Letters with diagonal line	M, N, V, W, X, Z	Writable in one flow, with vertical /horizontal line	I, J, L, M, N, P, U, Y, Z

The QUAD technique categorizes letters into four groups based on their physical characteristics, such as curvature, presence of horizontal lines, vertical lines, and diagonal lines, each assigned a unique 2-bit binary value. Similarly, the One-Flow-2-bit technique classifies letters based on writability and the presence of vertical or horizontal lines, also assigning distinct 2-bit binary values [18-20].

3.2 Alpha-based Representative Binary Technique

The Alpha-based Representative Binary technique represents all letters using capital and small letters in determining the group classification in embedding technique that shows in Table 2.

Table 2
Alpha-based Representative Binary group classification

Represents		Letters												
Number Embed		1	2	3	4	5	6	7	8	9	10	11	12	13
01	Capital	A	B	C	D	E	F	G	H	I	J	K	L	M
10	Small	a	b	c	d	e	f	g	h	i	j	k	l	m
Number Embed		14	15	16	17	18	19	20	21	22	23	24	25	26
01	Capital	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
10	Small	n	o	p	q	r	s	t	u	v	w	x	y	z

Based on Table 2 this technique uses a secret table based on number embed to map the letters and their positions in the stego text. The Alpha-based Representative Binary technique leverages the use of capital and small letters to represent binary values, with capital letters representing '01' and small letters '10', enhancing the complexity and capacity of hidden messages. This technique employs a secret table to track the positions and representations of letters in the stego text, ensuring that the hidden message remains concealed and undetectable. Each of these methods aims to maintain the natural appearance of the cover text while embedding information, balancing efficiency and security in text steganography

3.3 Performance Metric

This paper utilizes two metrics of evaluation in order to discover some performances in Alpha-based Binary Representative technique that compares with QUAD technique and One-Flow-2-Bit technique. The two metric that used to evaluates those techniques are contain letter used and capacity ratio.

- *Contain letter used:* It calculates the total character count in the stego text by determining the number of embedded letters displayed in the system's text area [16,23].
- *Capacity Ratio*: This metric evaluates the amount of data that can be embedded within the cover text relative to its size [4,18]. The formulation of the ratio is shown in Eq.(1).

$$\text{Capacity ratio} = \frac{\text{Amount of binary embedded}}{\text{Total letter of cover text}} \times 100\% \tag{1}$$

These metrics provide the comprehensive assessment to determine the performance of feature-based in text steganography techniques while optimizing the embedded information that calculates based on contain letter used and capacity ratio.

4. Results

The evaluation of the Alpha-based Representative Binary technique, the QUAD technique, and the One-Flow-2-Bit technique was conducted using the metrics of contain letter used and capacity ratio.

4.1 Contain Letter Used

This The Alpha-based Representative Binary technique demonstrated a higher total character count compared to the QUAD and One-Flow-2-Bit techniques. This indicates that the Alpha-based method is more effective in embedding a greater number of hidden letters within the cover text while maintaining the text's readability and natural appearance. Table 3 shows the total character count in the stego text for each technique.

Table 3
Contain letter used for each technique

Stego text (ST)	QUAD	One-Flow-2-bit	Alpha-based
ST 1	132	235	29
ST 2	130	373	59
ST 3	223	516	88
ST 4	342	652	118
ST 5	323	787	145
ST 6	397	925	173
ST 7	469	1060	184
ST 8	544	1193	184
ST 9	646	1335	184
ST 10	753	1447	184

The data in Table 3 reveals that the Alpha-based Representative Binary technique consistently results in a lower character count in the stego text across all samples (ST 1 to ST 10) compared to the QUAD and One-Flow-2-Bit techniques. The Alpha-based method is more efficient in embedding hidden messages requiring the contain letter used that only used fewer characters to achieve the same level of information concealment. The visualization the comparison the three techniques regarding the contain letter used performance is shown in Figure 2.

The comparative analysis of the techniques, as depicted in Figure 2 reveals that the Alpha-based Representative Binary technique outperforms the QUAD and One-Flow-2-Bit techniques in contain letter used. The lower total character count in letter used indicate that the Alpha-based method is more effective in embedding hidden messages in generate stego text.

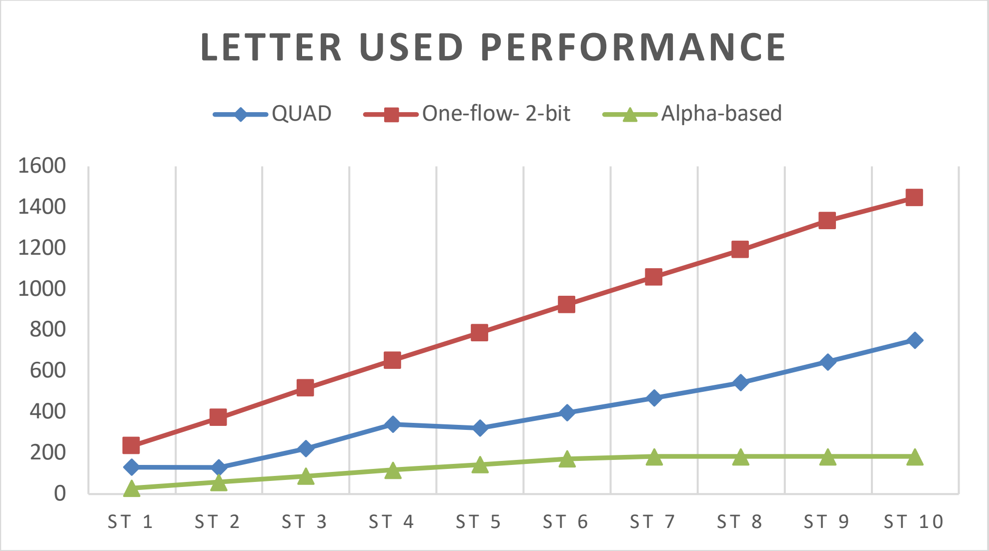

Fig. 2. Contain letter used comparison

4.2 Capacity Ratio

The capacity ratio for each technique was calculated using the formula provided in Eq.(1). Table 4 illustrates the capacity ratio for the Alpha-based Representative Binary technique, the QUAD technique, and the One-Flow-2-Bit technique. In text steganography, a lower capacity ratio indicates better performance, as it suggests that a smaller proportion of the cover text is used to embed the hidden message, making the stego text appear more natural and less detectable.

Table 4
Capacity ratio for each technique

Stego text (ST)	QUAD (%)	One-Flow-2-bit (%)	Alpha-based (%)
ST 1	0.022	0.018	0.002
ST 2	0.032	0.029	0.004
ST 3	0.044	0.04	0.006
ST 4	0.054	0.051	0.009
ST 5	0.065	0.062	0.011
ST 6	0.074	0.072	0.013
ST 7	0.085	0.082	0.0145
ST 8	0.095	0.093	0.0145
ST 9	0.107	0.104	0.0145
ST 10	0.117	0.115	0.0145

The data in Table 4 shows that the Alpha-based Representative Binary technique consistently achieves the lowest capacity ratio across all samples (ST 1 to ST 10) compared to the QUAD and One-Flow-2-Bit techniques. This indicates that the Alpha-based method uses a smaller proportion of the cover text to embed hidden messages, resulting in a more efficient and less detectable steganographic technique. The illustration the comparison the three techniques regarding the capacity ratio performance is shown in Figure 3.

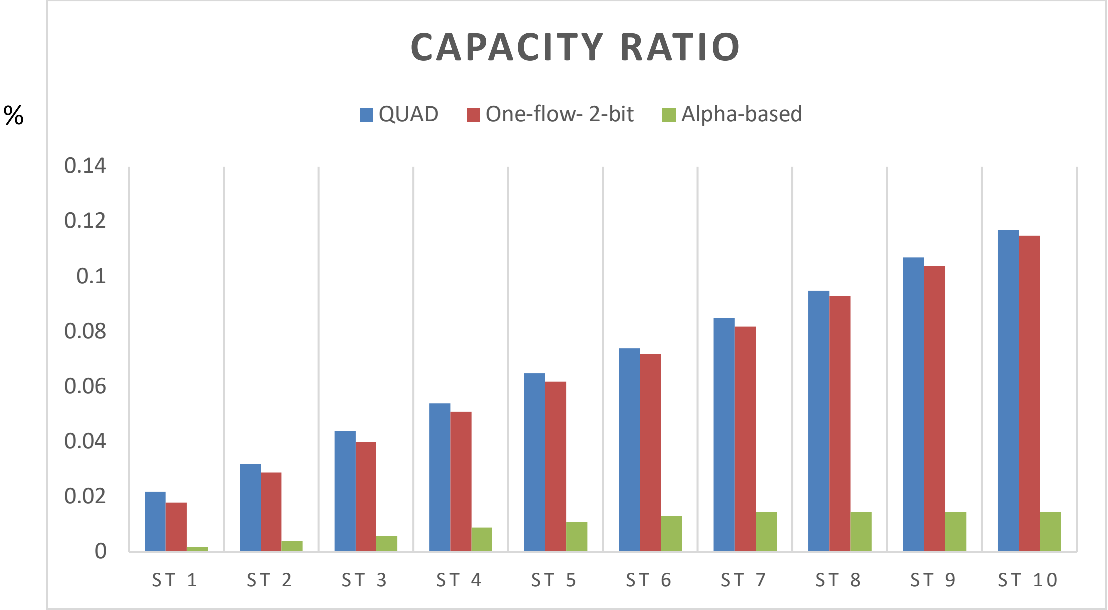

Fig. 3. Comparison capacity ratio performance

In Figure 3 shows the Alpha-based Representative Binary has lowest capacity ratio rather two techniques as feature-based 2-bit technique in text steganography. The lower capacity ratios achieved by the Alpha-based Representative Binary technique reflect its superior performance in maintaining the natural appearance and readability of the stego text while effectively embedding the necessary information.

5. Conclusions

This paper presents a comprehensive evaluation of the Alpha-based Representative Binary technique compared to the QUAD and One-Flow-2-Bit techniques in the realm of feature-based text steganography. The evaluation focuses on two critical performance metrics: contain letter used and capacity ratio. The Alpha-based Representative Binary technique consistently outperformed the other two methods across all samples. It demonstrated a lower total character count, indicating higher efficiency in embedding hidden messages while maintaining the natural appearance and readability of the stego text. This efficiency not only enhances the stego text's readability but also minimizes the likelihood of detection, making it appropriate choice for secure text steganography.

Furthermore, the Alpha-based Representative Binary technique achieved the lowest capacity ratio in all evaluated samples. This result signifies that the Alpha-based method utilizes a smaller proportion of the cover text for embedding hidden messages, thereby producing a more natural and less detectable steganographic text. The study highlights the technique in maintaining the integrity of the cover text while embedding a significant amount of hidden information. Consequently, the Alpha-based Representative Binary technique stands out as efficient method for secure text steganography, balancing the demands of information density and concealment the hidden message.

Acknowledgement

This research was PhD work that funded by the Universiti Utara Malaysia (UUM) under the Geran *Penjanaan Penyelidikan* 2024 (SO Code: 21657)

References

[1] Ditta, Allah, Muhammad Azeem, Shahid Naseem, Khurram Gulzar Rana, Muhammad Adnan Khan, and 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* 34, no. 5 (2022): 2180-2191. https://doi.org/10.1016/j.jksuci.2020.07.010

[2] Din, Roshidi, Reema Ahmed Thabit, Nur Izura Udzir, and Sunariya Utama. "Traid-bit embedding process on Arabic text steganography method." *Bulletin of Electrical Engineering and Informatics* 10, no. 1 (2021): 493-500. https://doi.org/10.11591/eei.v10i1.2518

[3] Qasim, Alaa Jabbar, Roshidi Din, and Farah Qasim Ahmed Alyousuf. "Review on techniques and file formats of image compression." *Bulletin of Electrical Engineering and Informatics* 9, no. 2 (2020): 602-610. https://doi.org/10.11591/eei.v9i2.2085

[4] Al-Nofaie, Safia, Adnan Gutub, and Manal Al-Ghamdi. "Enhancing Arabic text steganography for personal usage utilizing pseudo-spaces." *Journal of King Saud University-Computer and Information Sciences* 33, no. 8 (2021): 963-974. https://doi.org/10.1016/j.jksuci.2019.06.010

[5] Utama, Sunariya, and Roshidi Din. "Performance review of feature-based method in implementation text steganography approach." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 28, no. 2 (2022): 325-333. https://doi.org/10.37934/araset.28.2.325333

[6] Din, Roshidi, Ahmad Hamid Shakir, Sarmad Hamzah Ali, Alaa Jabbar Qasim Almaliki, Sunariya Utama, and Jabbar Qasim Almaliki. "Exploring Steganographic Techniques for Enhanced Data Protection in Digital Files." *International Journal of Computational Thinking and Data Science* 1, no. 1 (2024): 1-9. https://doi.org/10.37934/CTDS.1.1.19

[7] Yaghobi, Shabnam Rahber, and Hedieh Sajedi. "Text steganography in webometrics." *International Journal of Information Technology* 13, no. 2 (2021): 621-635. https://doi.org/10.1007/s41870-020-00572-z

[8] Alanazi, Norah, Esam Khan, and Adnan Gutub. "Functionality-improved Arabic text steganography based on unicode features." *Arabian Journal for Science and Engineering* 45 (2020): 11037-11050. https://doi.org/10.1007/s13369-020-04917-5

[9] Sajjad, Asma, Humaira Ashraf, N. Z. Jhanjhi, Mamoona Humayun, Mehedi Masud, and Mohammed A. AlZain. "Improved Video Steganography with Dual Cover Medium, DNA and Complex Frames." *Computers, Materials & Continua* 74, no. 2 (2023). https://doi.org/10.32604/cmc.2023.030197

[10] Gurunath, R., and Debabrata Samanta. "A new 3-bit hiding covert channel algorithm for public data and medical data security using format-based text steganography." *Journal of Database Management (JDM)* 34, no. 2 (2023): 1-22. https://doi.org/10.4018/JDM.324076

[11] Gurunath, R., and Debabrata Samanta. "A new 3-bit hiding covert channel algorithm for public data and medical data security using format-based text steganography." *Journal of Database Management (JDM)* 34, no. 2 (2023): 1-22. https://doi.org/10.4018/JDM.324076

[12] Xiang, Lingyun, Shuanghui Yang, Yuhang Liu, Qian Li, and Chengzhang Zhu. "Novel linguistic steganography based on character-level text generation." *Mathematics* 8, no. 9 (2020): 1558. https://doi.org/10.3390/math8091558

[13] Scientific, Little Lion. "TEXT STEGANOGRAPHY BASED ON UNICODE CHARACTERS AS MARKER IN INDONESIAN EXCEL FILE." *Journal of Theoretical and Applied Information Technology* 102, no. 12 (2024).

[14] Wu, Ning, Zhongliang Yang, Yi Yang, Lian Li, Poli Shang, Weibo Ma, and Zhenru Liu. "STBS-Stega: Coverless text steganography based on state transition-binary sequence." *International Journal of Distributed Sensor Networks* 16, no. 3 (2020): 1550147720914257. https://doi.org/10.1177/1550147720914257

[15] Majeed, Mohammed Abdul, Rossilawati Sulaiman, Zarina Shukur, and Mohammad Kamrul Hasan. "A review on text steganography techniques." *Mathematics* 9, no. 21 (2021): 2829. https://doi.org/10.3390/math9212829

[16] Fateh, Mansoor, and Mohsen Rezvani. "An email-based high capacity text steganography using repeating characters." *International Journal of Computers and Applications* 43, no. 3 (2021): 226-232. https://doi.org/10.1080/1206212X.2018.1517713

[17] Thabit, Reema, Nur Izura Udzir, Sharifah Md Yasin, Aziah Asmawi, Nuur Alifah Roslan, and Roshidi Din. "A comparative analysis of arabic text steganography." *Applied Sciences* 11, no. 15 (2021): 6851. https://doi.org/10.3390/app11156851

[18] Muhammad, Mohd Hilal, Hanizan Shaker Hussain, Roshidi Din, Hafiza Samad, and Sunariya Utama. "Review on feature-based method performance in text steganography." *Bulletin of Electrical Engineering and Informatics* 10, no. 1 (2021): 427-433. https://doi.org/10.11591/eei.v10i1.2508

[19] Aditya Kumar, K., and Suresh Pabboju. "Text Steganography: Design and Implementation of a Secure and Secret Message Sharing System." In *Advances in Decision Sciences, Image Processing, Security and Computer Vision: International Conference on Emerging Trends in Engineering (ICETE), Vol. 1*, pp. 470-479. Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-24322-7_58

[20] Kouser, Saeeda, and Aihab Khan. "A NOVEL FEATURE EXTRACTION APPROACH: CAPACITY BASED ZERO-TEXT STEGANOGRAPHY." *International Journal on Information Technologies & Security* 9, no. 3 (2017).

[21] Azeem, Muhammad, Cai Yongquan, Khurram Gulzar Rana, Zeeshan Shaukat, and Allah Ditta. "A secure and size efficient approach to enhance the performance of text steganographic algorithm." In *2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, pp. 402-407. IEEE, 2019. https://doi.org/10.1109/ICMTMA.2019.00095

[22] Alkhudaydi, Malak, and Adnan Gutub. "Securing data via cryptography and arabic text steganography." *SN Computer Science* 2, no. 1 (2021): 46. https://doi.org/10.1007/s42979-020-00438-y

[23] Majeed, Mohammed Abdul, Rossilawati Sulaiman, and Zarina Shukur. "New Text Steganography Technique based on Multilayer Encoding with Format-Preserving Encryption and Huffman Coding." *International Journal of Advanced Computer Science and Applications* 13, no. 12 (2022). https://doi.org/10.14569/IJACSA.2022.0131222