Manuscript ID – 103

An English Sentence Dictionary based Secure Text Steganographic Technique for Message-Data Confidentiality

COCOLE-2023

Presented by

Dr. Alekha Kr Mishra

NIT Jamshedpur

Authors

Akash Kumar Dey

Geeta Gayatri Behera

Alekha Kumar Mishra

Outline

- Introduction
- Classification of Steganography
- Related Works
- Motivation
- Proposed Work
- Results
- Conclusion

Introduction[13]

- Steganography is the practice of hiding messages within non-secret text or data.
- It is derived from a Greek word called steganographia, meaning “concealed writing”.
- In contrast to cryptography it hides actuality and existence plaintext instead of trasforming to another form.
- Steganography aims to achieve secure communication unnotices to avoid suspicious eavesdropping on data transmission
- Current era of security is dominated by cryptographic methods , nevertheless steganogrphy is still an active area of reasearch

Steganography Building Blocks[4]

- The message to be hidden is called secret message
- The secret message embedded within a file or message called cover message/text.
- The cover message with an embedded secret message is known as a stego file.
- The key used in the hiding process is known as stego key, which is also used to retrieve hidden content.

Classification of Steganography [11]

- The steganographic techniques are broadly classified into :
  - Text steganography : uses text as a cover file
  - Image steganography : hiding a text or an image inside another image
  - Audio steganography: hiding the secret message into the audio
  - Video steganography: hiding secret information inside videos

Related Works

Author	Year	Technique used to hide text	Remarks
Acharjee *et al.* [2]	2016	bit-level XOR operation	No use of key therefore is not secured and robust.
Banerjee *et al.* [3]	2021	use face geometry biometric authentication and handwriting based text steganography technique	Overhead time to find face reference points and mapping to table
Chaudhary *et al.* [4]	2016	Encode text based on the shape of the capital alphabets. S.imilar encoding approach for Hindi language alphabets	overhead time claimed to be way lessser than the other techniques.
Gharavi and Rajaei [5]	2018	Curvelet transform features	average Mean Square Error and PSNR of their experiment lies around 12.00 and 38 respectively

Related Works contd...

Author	Year	Technique used to hide text	Remarks
Karthikeyan *et al.* [7]	2017	combines the Least Signification Bit (LSB) Technique along with standard Data Encryption Standard (DES) cipher	Mean Square Error and Peak Signal-to-Noise Ratio (PSNR) of their experiment lies around .0035 and 52 respectively.
Kataria *et al.* [8]	2013	reordering of two XORed characters from the message based on a key	minimizes the time overhead requirement.
Manish Kumar *et al.* [9]	2022	Combines _x005F_x005F_x005F_x005F_x005F_x005F_x005F_x005F_x005F_x005F_x005F_x0002_ing LSB technique and the Advanced Encryption Standard (AES) encryption	average Mean Square Error and PSNR of their experiment lies around 0.0019922 and 75.1375 respectively
Wu *et al.* [14]	2019	uses the half frequency crossover rule, which utilizes the natural language characteristics	achieved embedding rate of 2.78.

Motivation

- Most of the reported techniques focus on type of cover message format
- The LSB technique with DES and AES are reported to have better plaintext vs. covertext ratio for image steganographic techniques
- Dictionary data stucture takes O(1) time for Insert, Delete and Search operations.
- None of the text based steganographic techniques has used dictionary as the media to be used as repository of cover files

Proposed Work - Dictionary Based Text Steganography(DBTS)

- DBTS uses a dictionary of collection of English Sentences that are used as repository of covertext.
- The English sentences are stored via key as the string of alphabets absent in it.
- The list of sentences that contain every letter of the plaintext is selected from dictionary, and one of those sentences is randomly selected to conceal the hidden plaintext.
- By identifying potential hiding places for the plaintext characters, a stegokey is created for the plaintext.

Dictionary Based Text Steganography(DBTS) - Advantages

- The plaintext characters in the covertext are randomly distributed based on stegokey indexes.
- The uncovering procedure simply uses stegokey to search for the symbol contained in the covertext at the designated index.
- The uncovering process does not need the dictionary used during the encoding process.
- DBTS takes significantly lesser amount of time due to use of dictionary data structure.

Steps involved – An overview

- **Step 1:** Build a sentence dictionary using English sentence dataset [1].
- **Step 2:** For a given plaintext, select the candidate covertext message from the pool of sentences retrieved from the dictionary.
- **Step 3:** The algorithm uses the plaintext to identify the location of the symbols of plaintext in the selected covertext message.
- **Step 4:** The stegokey based on the hidden location of the plaintext in the selected covertext.
- **Step 5:** Decoding does not rely on the sentence dictionary used during message concealment. Only the key is required to successfully retrieve the secret plaintext from the covertext.
- **Note:** The generation of the stegokey is a crucial aspect of this technique, based on the candidate sentence chosen to hide the secret plaintext.

Algorithm 1 : Building Dictionary

Algorithm 1: The Algorithms for Building the English Sentence Dictionary.

```
Input: sentenceFile
Output: D
D= new Dict();
for each s in sentenceFile do
    find alphabets absent in s;
    add absent alphabets to abs_key_set;
    if D[abs_key_set] is empty then
        D[abs_key_set]=s;
    else
        add s to D[abs_key_set] sentence list;
output D;
```

Algorithm 2 -Encoding plaintext into covertext

Algorithm 2: The Algorithms for Encoding plaintext.

Input: plaintext, D

Output: covertext

```
select the sentences from D that contains all alphabets in plaintext;
add sentences to the selectedList;
choose a candidate sentence s at random from selectedList;
for each c in plaintext do
    indices=findIndexes(c,s);
    index=random(indices);
    add index to the stegokey;
covertext = s;
output covertext, stegokey;
```

Algorithm 3: Decoding plaintext from covertext

Algorithm 3: The Algorithms for Extracting plaintext.

Input: covertext, stegokey
Output: plaintext

```
for each k in key do
    val=covertext.getValue(k);
    add val to plaintext;
output plaintext;
```

Time complexities

- The time complexity of the Algorithm 1 is O(n), where n is the number of English sentences in the dataset.
- The Algorithm 2 traverses through the input plaintext alphabets to pick hiding sentence from the dictionary. Therefore, it has the time complexity of O(k), where k is the number of alphabets in the input plaintext and k << n.
- The decoding or extraction process uses the stegokey to retrieve the original plaintext, therefore it has constant time complexity O(1).

Results : Some example Test outputs

Plaintext = " deploy troops on secondfront tonight"

Run-1

Selected Covertext = It made clear that you pressured a foreign government to interfere in our political process on your behalf, you violated your oath of office and betrayed our nation

key = (5, 6, 23, 9, 20, 19, 2, 1, 12, 20, 20, 23, 26, 2, 20, 41, 2, 26, 6, 8, 20, 41, 5, 35, 12, 20, 41, 1, 2, 1, 20, 41, 39, 40, 15, 1)

Run-2

Selected Covertext = For this project I see myself as becoming a Japan and Europe go-between

key = (52, 13, 9, 27, 1, 24, 3, 4, 2, 1, 1, 9, 7, 3, 1, 39, 3, 7, 13, 14, 1, 39, 52, 28, 2, 1, 39, 4, 3, 4, 1, 39, 6, 40, 5, 4)

Run-3

Selected Covertext = My metabolism is such that no matter how much I eat I don't put on weight. Just now, this second, you've made enemies of people throughout the world.

key = (55, 5, 61, 10, 9, 2, 3, 6, 36, 9, 9, 61, 12, 3, 9, 28, 3, 12, 5, 20, 9, 28, 55, 122, 36, 9, 28, 6, 3, 6, 9, 28, 11, 71, 21, 6)

Results : Encoding and Decoding Times (in Seconds)

Plaintext Size (bytes)	Encoding Time (sec)	Decoding Time (sec)
100	0.102	2.4E-05
200	0.098	3.9E-05
300	0.103	5.5E-05
400	0.114	6.8E-05
500	0.1	8.7E-05
600	0.12	9.1E-05

Table 1: Encoding and Decoding times of DBTS

Comparison of Plaintext size vs Covertext size

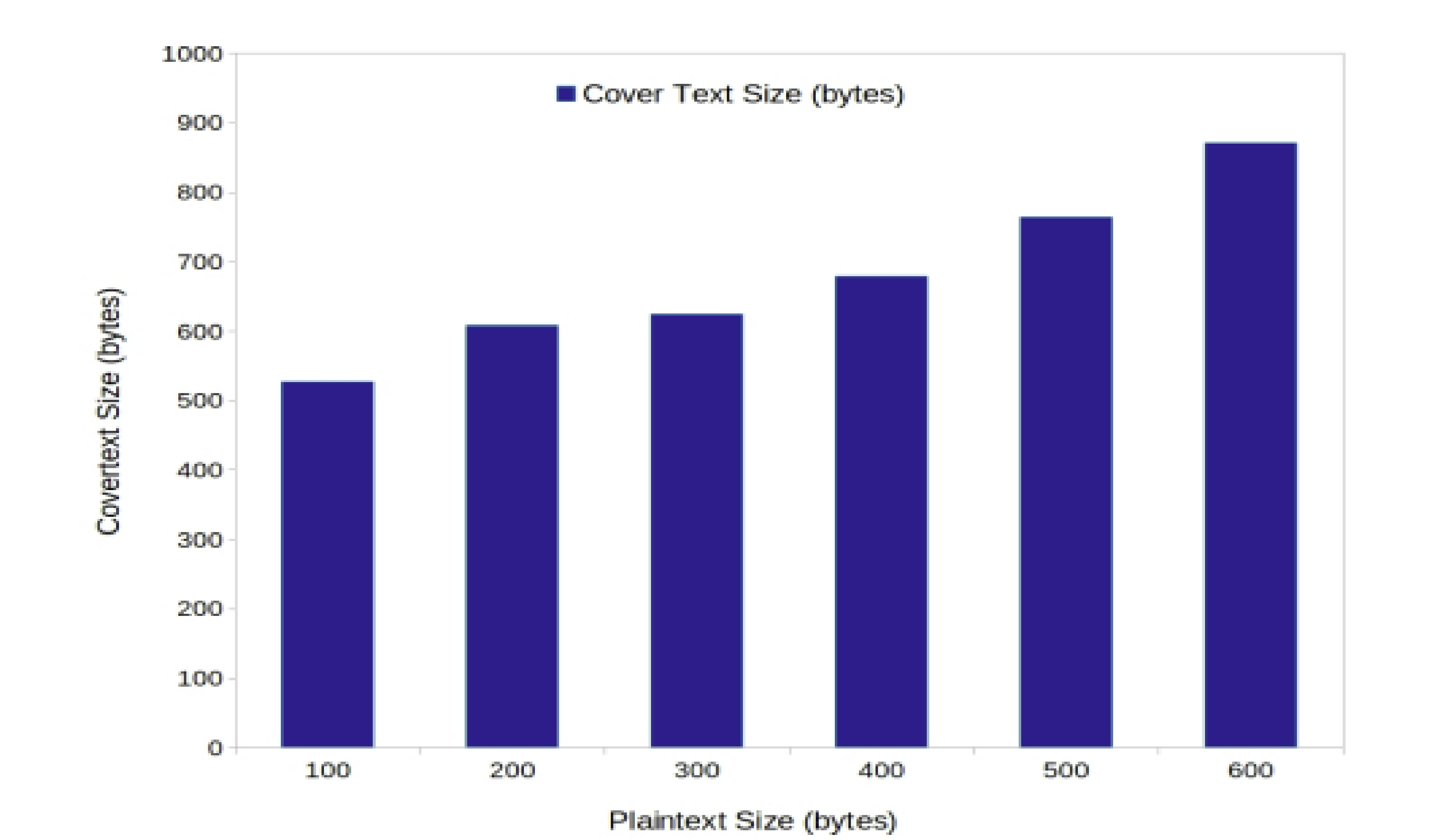

Fig 1: Comparison of Message Text and Cover Text Size of the Proposed DBTS technique.

Comparison of Encoding time with Banerjee *et al.*[3]

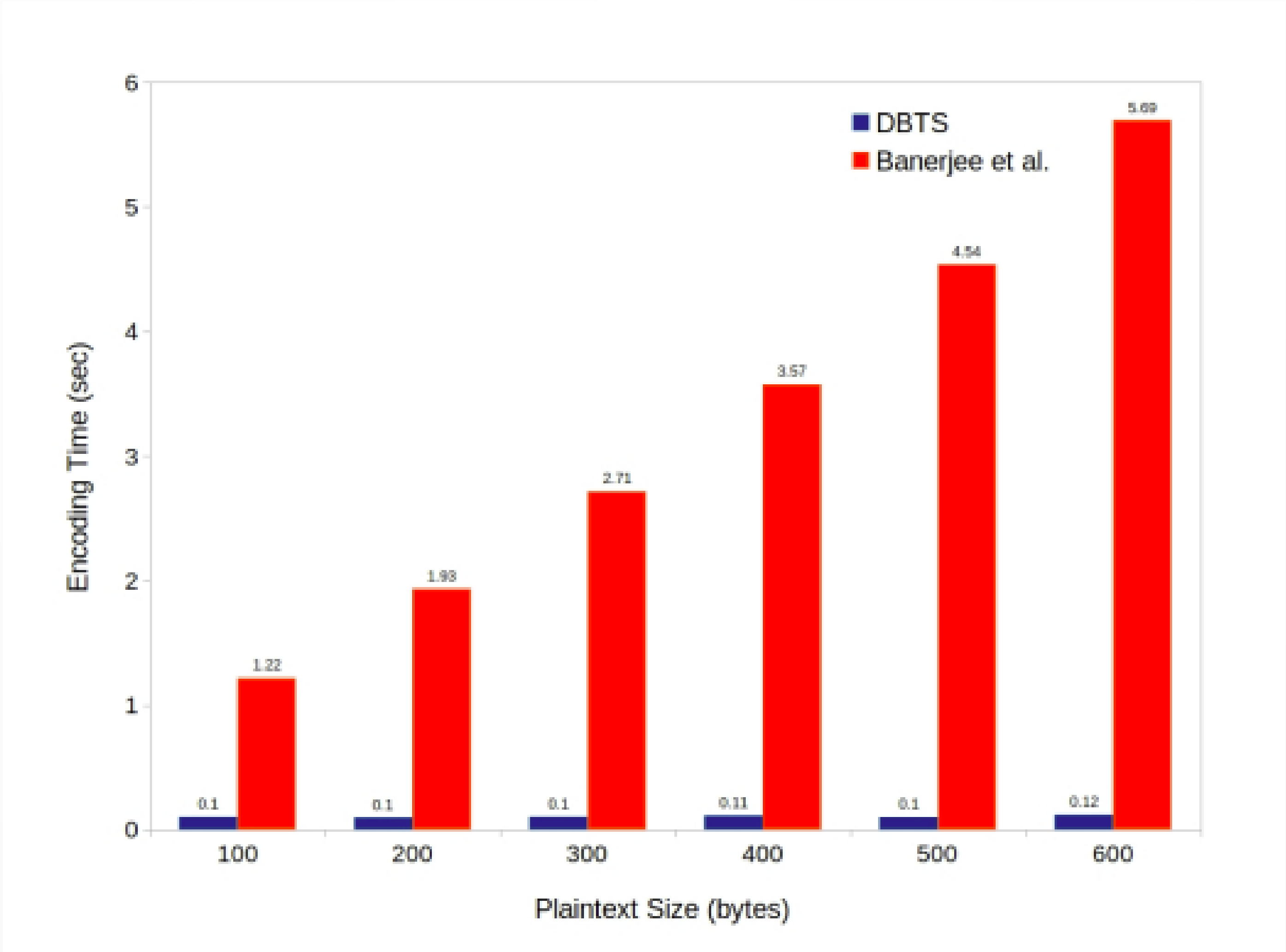

Fig. 2: Comparison of Encoding Time between Proposed DBTS and Banerjee et al.[3]

Comparison of Decoding time with Banerjee et. al.

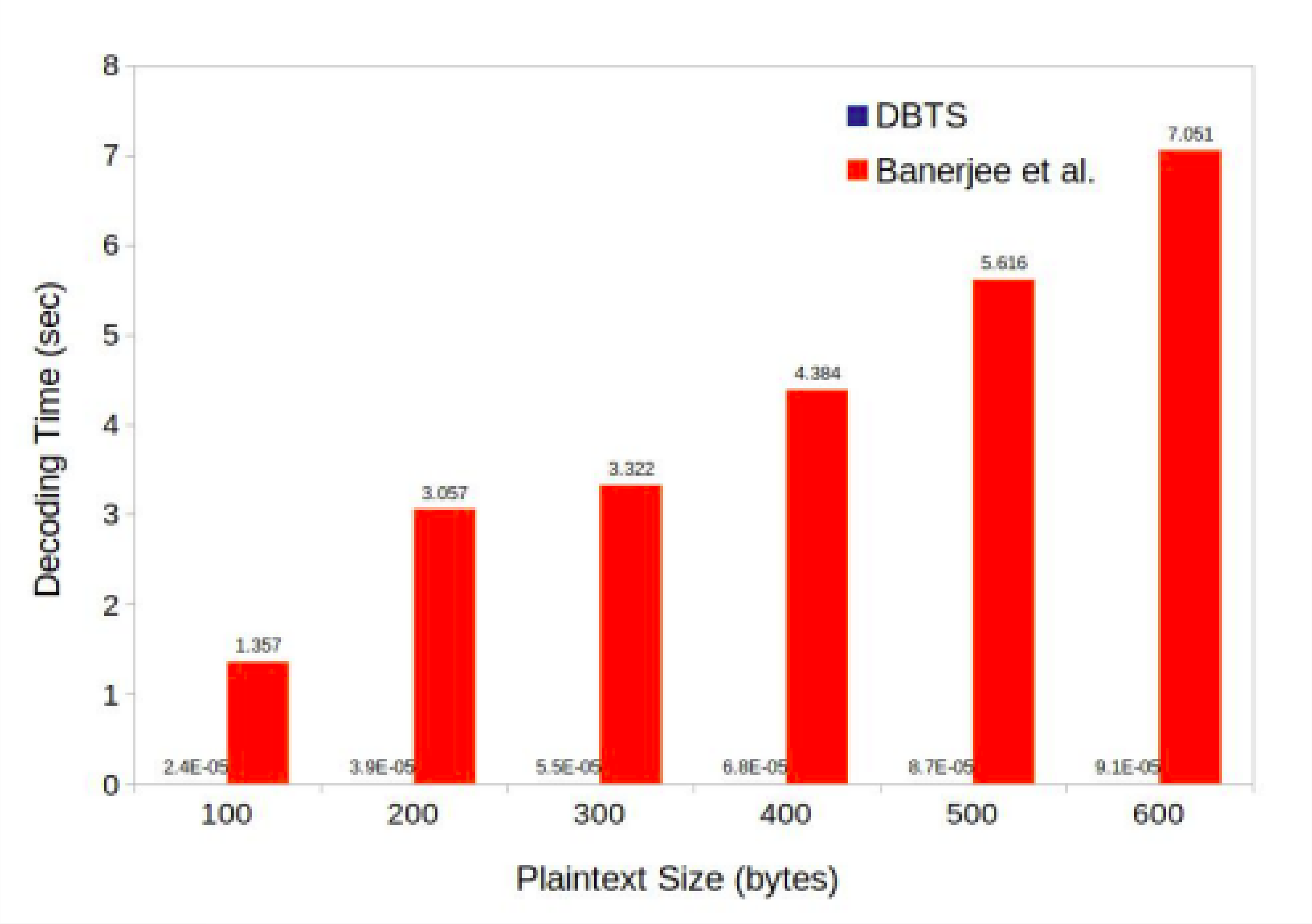

Fig. 3: Comparison of Decoding Time between Proposed DBTS and Banerjee et al..

Attacks on the processes

- An adversary requires to guess the length of the secret plaintext in order to recover it from the covertext.
- When an adversary is able to find the length of the plaintext, then it needs to extract all meaningful sentences possible out of given covertext message that requires exponential computational time.
- As a result, the process of uncovering the secret message from the covertext is extremely tedious given that the adversary known the algorithm and the plaintext length.

Conclusion and Future Work

- This paper contributes a new English sentence dictionary based steganographic technique called DBTS.
- The key point is the selection of English sentence where the secret plaintext can be hidden. The stegokey is generated accordingly.
- The results shows the proposed technique required smaller encoding and decoding time compared to existing one.
- The DBTS in some cases bears unusual lenghtier key, it is needed to be minimized.
- We may look for an alternative dataset to build a more robust dictionary.

References

- 1. Tatoeba : a collection of sentences and translations., https://tatoeba.org/en/
- 2. Acharjee, T., Konwar, A., Kumar Ram, R., Sharma, R., Goswami, D.: Xorsteg: A new model of text steganography. In: 2016 International Conference on Communication and Electronics Systems (ICCES). pp. 1–4 (2016). https://doi.org/10.1109/CESYS.2016.7889820
- 3. Banerjee, I., Bandyopadhyay, K., Rashid, T.A., Mohan, A., Alsadoon, A., Kumar, A.: Face geometry and handwritten characters based biometric text steganography. In: 2021 IEEE Bombay Section Signature Conference (IBSSC). pp. 1–6. IEEE (2021)
- 4. Chaudhary, S., Dave, M., Sanghi, A.: Aggrandize text security and hiding data through text steganography. In: 2016 IEEE 7th Power India International Conference (PIICON). pp. 1–5. IEEE (2016)
- 5. Gharavi, H., Rajaei, B.: A robust steganography algorithm based on curvelet transform. In: Electrical Engineering (ICEE), Iranian Conference on. pp. 1624–1628. IEEE (2018)
- 6. Johnson, N.F., Duric, Z., Jajodia, S.: Information hiding: steganography and watermarking-attacks and countermeasures: steganography and watermarking: attacks and countermeasures, vol. 1. Springer Science & Business Media (2001)
- 7. Karthikeyan, B., Deepak, A., Subalakshmi, K., MM, A.R., Vaithiyanathan, V.: A combined approach of steganography with lsb encoding technique and des algorithm. In: 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB). pp. 85–88. IEEE (2017)

References contd...

- 8. Kataria, S., Kumar, T., Singh, K., Nehra, M.S.: Ecr (encryption with cover text and reordering) based text steganography. In: 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013). pp. 612–616. IEEE (2013)
- 9.Kumar, M., Soni, A., Shekhawat, A.R.S., Rawat, A.: Enhanced digital image and text data security using hybrid model of lsb steganography and aes cryptography technique. In: 2022 Second international conference on artificial intelligence and smart energy (ICAIS). pp. 1453–1457. IEEE (2022)
- 10. Majeed, M.A., Sulaiman, R., Shukur, Z., Hasan, M.K.: A review on text steganography techniques. Mathematics 9(21) (2021). https://doi.org/10.3390/math9212829
- 11. Majumder, A., Changder, S.: A generalized model of text steganography by summary generation using frequency analysis. In: 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). pp. 599–605. IEEE (2018)
- 12. Mishra, A.K., Puthal, D., Tripathy, A.K.: Graphcrypto: Next generation data security approach towards sustainable smart city building. Sustainable Cities and Society 72, 103056 (2021)
- 13. Naharuddin, A., Wibawa, A.D., Sumpeno, S.: A high capacity and imperceptible text steganography using binary digit mapping on ascii characters. In: 2018 International Seminar on Intelligent Technology and Its Applications (ISITIA). pp. 287–292. IEEE (2018)
- 14. Wu, N., Ma, W., Liu, Z., Shang, P., Yang, Z., Fan, J.: Coverless text steganography based on half frequency crossover rule. In: 2019 4th International Conference on Mechanical, Control and Computer Engineering (ICMCCE). pp. 726–7263. IEEE (2019)

Thank You
Any Questions?