ISSN (Print) 2794-**7629** **Received** **07/10/2022**
ISSN (Online) 2794- **4549** Accepted 25/12/2022

FULL PAPER

The AI algorithm for Text Encryption using Steganography

Prepared by

Eng. Saad bin Nasser Al AZZam
College of Computing
Department of Cyber Security
University of Bisha
Snazzam.199@gmail.com

Abstract:

The human being sought to find a human variety of techniques to assure data access with complete confidentiality. The transition from the use of. regular text and audio data to digital media has improved data access and transport. It has become relatively simple to intercept data sent across networks or get access to a variety of machines. Steganography is the study of concealing the existence of communication by embedding hidden data in multimedia payloads such as text, image, audio, and video. This project investigates how to improve steganography by merging text and image production to provide invisible encryption, security, and robustness in digital images. Text hide into another text using quality in the proposed system include Mean Square Error (MSE), Normalized Correlation (NC), and Normalized Cross–correlation Mean Squared Error, Histogram Analysis, Standard Deviation, and Statistical Test and Analysis. The algorithm of our suggested system seems to meet the greatest standards of security, perceptions, and capability. Using the standard of ASCII Control characters for a cover word, the procedure for using the processed system is supplied from the cover sentence. The same method is used for other words from different cover sentences, given the line numbers of each cover sentence and Stego sentence.

.

Keywords: Algorithm, Secret Text, Steganography,

The Scholar Journal for Sciences & Technology sjst

1. INTRODUCTION

Text Steganography is the most advanced method of securely concealing secret messages in an innocent carrier with a dynamic tool and the capacity to adapt to new cutting-edge technology. Using cyber security for multimedia communication and providing better safety for distributed systems of digital data have attracted a lot of interest.

In the modern information hiding age, as digital media "text, "images, " "audio" "video and network provide sufficient redundancy, security, invisibility and robustness for information hiding. Digital signatures, Biometrics, Bioinformatics and spread spectrum communications to fulfill the desire of secret hidden communication.(P. Rajba, W. Mazurczyk,2016)

Watermarking is a sub-discipline of information hiding, in contrast to Steganography, which attempts to hide copyright messages into a medium with a high level of robustness to against possible attacks. Watermarking enables the identification of copyright violation or an accumulation of evidence to be used in Public and government law. Steganography's objective is to ensure that third parties do not suspect or determine whether an object is a hidden message or object. (Majeed, M.A.et.al ,2021)

The steganography in next few years, the renowned scientific field needs to reinvent itself with new technology and applications. This work used to describe and implement the new algorithm by hiding secret text in another text. The main performance and testing of the proposed method, we have focused on the security, invisibility and capacity.(Neha Rani, Jyoti Chaudhary ,2013)

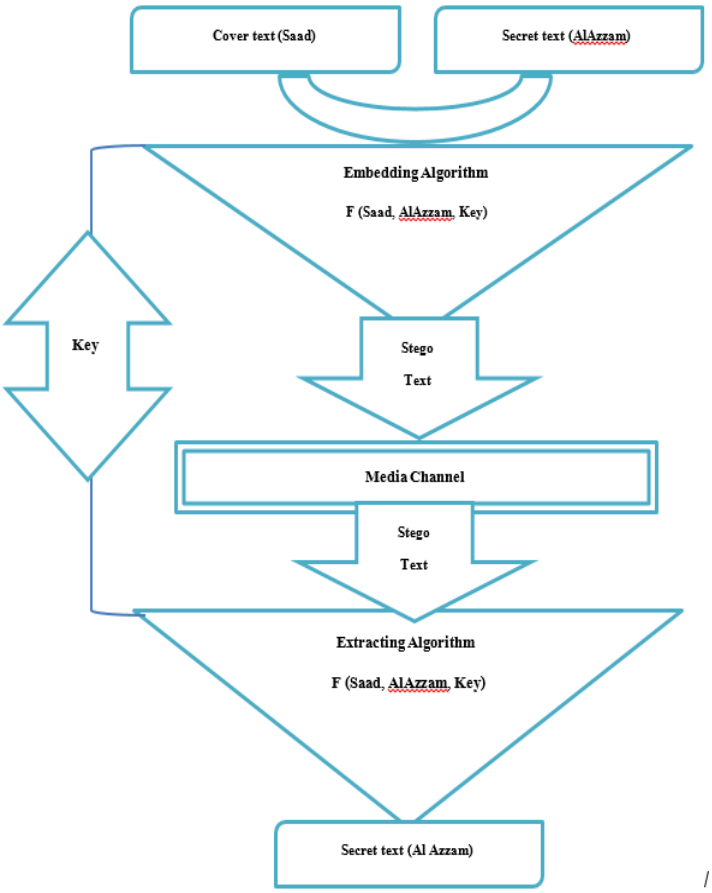


Figure 1. Hiding secret text in another text scheme.

A proposed system that focuses on how to hide the security of communication and capacity of secret data by combining text and text of hiding techniques approaches such as text, secret text, embedding algorithm, extracting algorithm, stego text and key. The unifying models. Fig.1 Illustrates the overall view of the proposed system.(K. V. Kale.et.al,2012)

2. BACKGROUND AND RELATED WORK

In this section, Author lay the groundwork for this task by looking to the past for examples before turning to the present. In this examination, we focus on data concealment by looking at the many features, needs, and schemes of steganography. To better comprehend this work, we will examine the foundational concepts of information concealing in this chapter, with a special emphasis on (Word embedding). We then

compare it to seminal works that have come before.(S. Katzenbeisser .et.al,2000)

Encoding and decoding data for use in communication is known as "information hiding." Hidden information can be made apparent or the data itself can be created. They (the Trithemius) came up with the term "steganography." It derives from the Greek words steganos, meaning "covered," and graphia, meaning "writing."
Stenography is the art and science of secretly transmitting data. Topics include how numerous parties in a communication chain might work together to obfuscate a message in a cover medium before sending it out into the public sphere. In the earliest established steganography model described by Simmons, criminal suspects Alice and Bob are each assigned their own cell. Wendy worked as a warden; therefore, they need to figure out how to get out of there. She's not going to let them have any kind of interaction with one another. To avoid Wendy's suspicion, they must have clandestine conversations.(P. Rajba, W. Mazurczyk,2019)
The information-hiding process known as steganography Numerous programmes regularly made use of the stated techniques for concealing data, which included text hiding, picture hiding, network packet hiding, software and hardware hiding, audio hiding, and video hiding. This study highlights the need for a standardised method and performance metrics to evaluate steganographic algorithms. For decades, scientists have tried to come up with new ways of conducting covert communication.(F. A. P. Petitcolas, R. J. Anderson,1999)
Exhibit the many information-hiding disciplines in Fig.2, the italicised text denotes the research-in-progress methods of blending text into text.

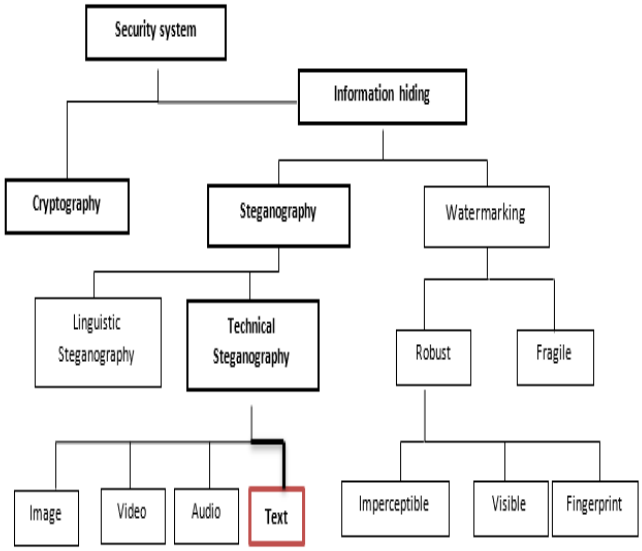


Figure 2. The bold face denotes the focus study, which is a different area of information hiding

2.1 Text steganography

All of these strategies for concealing text within another text fall into one of three categories: linguistic techniques, format-based randomness, or random and statistical generation. (N. F. Johnson and S. Jajodia,1998)

2.1.1. Format Based Methods

In this technique, sensitive information is hidden by the use of deliberate typos in text design. The stego file may be deciphered by opening it in a word processor, which will reveal any typos or extra white space. Font size changes may raise suspicions in the mind of a human reader. If the original plaintext can be accessed, a comparison with the allegedly steganographic text will reveal the areas of transformation.(Ross J. Anderson, Fabien A. P. Petitcolas,1998)

2.1.2. Random and Statistical Generation

steganography is aware of the plaintext, he or she may decide to write a new cover text instead. For instance, one can conceal

information in a seemingly meaningless sequence of symbols. Using statistical factors like vocab size and letter frequencies, another method generates words that appear to have the same statistical qualities as genuine words in the target language. (Ari Moesriami Barmawi,2016)

2.4.1 Linguistic Steganography

The term "linguistic steganography" refers to a subset of steganography that uses language as a "hiding place" in messages and takes into consideration the attributes of created and updated linguistic text technique and uses the varied methods as scenarios. Linguistic steganography is a method that employs context-free grammars to build a tree of data, which may then be used to conceal information by assigning the left branch the value '0' and the right branch the value '1'. A different technique uses a Greibach-Normal-Form grammar, where the first choice represents bit 0 and the second option represents bit 1 in a production. However, this approach is not without its flaws. as it makes use of the diminutive grammar, there will be a great deal of verbatim recitation. Second, there is no discernible semantic structure to the text despite its impeccable grammatical order. The end result is a series of unrelated phrases.(Deepali Bhat, Krithi V, Manjunath KN,2017)

3. METHODOLOGY

Steganographic techniques have been proposed for a range of applications in recent years. Majority of linguistic steganography methods is utilizes changes to the syntax or the lexicon. Even if the data is vulnerable to cover modification, the goal of these tactics is to conceal it.

The main goals of this study are to provide new, useful, and computationally effective methods for text steganography that are based on linguistics, format-based methodologies, statistical and random generation, or a combination of both. For security, concealment, and capacity methods, a text multimedia system combining text and text Steganography is used.

A text steganography technique involves in English text a mechanism for generating an embedded message, strategies for embedding the message, and techniques for extracting the message. For security, the encrypted communication usually used a secret key that was shared between the sender

and the receiver. The assumption that the encrypted message is a binary random sequence has been used in many statistical evaluations of text Steganography.

Each use of text steganography has its own set of requirements, which are determined by the application's goal. This kind of embedding will be accomplished through the use of shared secret keys. The confirmed message upon extraction should resemble the first message, demonstrating that the stego object's content has not modified since it was marked.(S.Changder, D. Ghosh,2010)

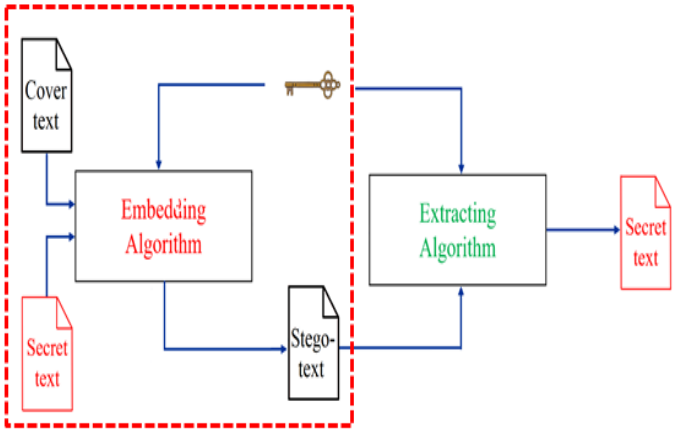


Fig.3. The work process approach to be achieved in text steganographic

In this article, we examine how text steganography techniques are used to leverage imperceptibility to hide hidden data (text) within other media (text). The most important necessity is security, even when statistical tools are used, it is impossible for the human eye to comprehend. In the most basic scenario,

the embedder works by adjusting synonym replacement by looking up synonyms in a word dictionary. Secret text is then compressed using the Huffman Compression Algorithm.(Aleksandr A. Maliavko,2018)

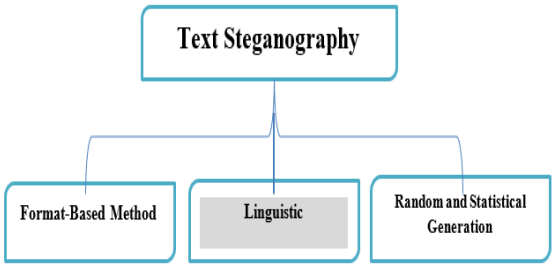


Fig 4. Random and statistical generation, as well as "linguistic, format-based, random and statistical generation".

In a Linguistic domain are the most well-known text steganography methods that focused on changing the embedding and use of physical properties of text symbols. Here we go over some of the ways for altering features so that they are undetectable to the naked eye, uses to hide the secret material.(Ei Nyein Chan Wai, May Aye Khine,2011)
Hiding Data in Wordlist is a technique without the use of any special characters, concealing a message within a list of words. Another method that smoothly combines Arabic and Unicode Standard characters, pseudo-spaces, and Kashida (extension characters) to enable safe use at the individual

level. A proposed method for embedding secret information by using the key to match to the necessary characters.(Keshav Joshi,2018)
Using A text steganography method based on text format was used to improve concealed data storage by using the justified written text included inside PDF documents. The secret information was hidden using a format-specific steganography technique that involved embedding it in cover objects. (Alfin Naharuddin, Adhi Dharma,2018)
Method makes use of the justified formatted text found in PDF documents to improve hidden data storage. In order to convey the text of the messages, a format operation must be carried out after choosing a subset of cover-elements. The stego-text is chosen to extract and align during the extraction process in order to obtain the hidden message.(N.Subramanian,et al,2021)

Algorithm 3.1 Embedding process: (formatting techniques)
for $i = 1,...,1(c)$ **do**
$S_i \leftarrow C_i$
end for
for $i = 1 ...,l(m)$ **do**
Compute index j_i where to store i th message bit
$s_{ji} \leftarrow c_{ji} \leftrightarrows m_i$
end for
Algorithm 3.2 Extraction process: least significant bit (LSB)
for $i = 1,...,1\ (M)$ **do**
Compute index j_i where the i th message bit is stored
$m_i \leftarrow$ LSB (C_{ij})
end for

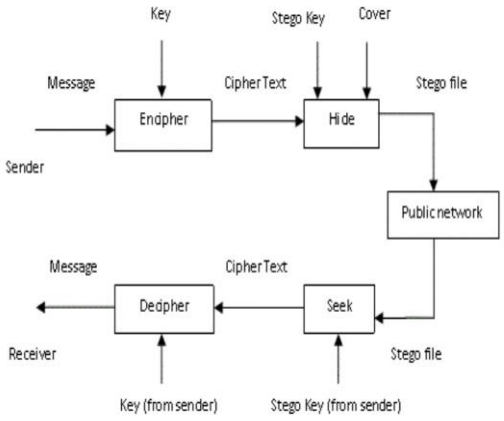
3.1. RANDOM TEXT STEGANOGRAPHY AND STATISTICAL GENERATION

The work developed a mathematical text steganography method based on the Markov Chain model that focuses on transition probability. This technique additionally encodes the state transition-binary sequence diagram necessary by the receiver to decode the information. The results revealed that this model has a higher potential for concealment than earlier strategies.
Established a coverless plain text steganography approach based on the parity of binary search tree (BST) based on Internet texts. To boost security and hiding capacity, apply a randomized indexed word dictionary to a set of emails. Huffman developed an email-based text Steganography system with a reversible feature.(Shivani Sharma, Dr. Avadhesh,2016)

3.2. LINGUISTIC TEXT STEGANOGRAPHY

This technique terms of linguistic steganography use to conceal sensitive information within text files. It can create a steganographic paragraph based on knowledge

graphs that generates coherent multi-sentence texts for concealment. It is focuses on the knowledge graphs with a specified theme to produce coherent multi sentence uses for writings data. (N.R. Zaynalov,et al,2020)



D

The process of hiding of text into another text. Steganography aims to utilize the application of concealing information. It is implemented in real time in order to satisfy the design goals of security and imperceptibility. Given the line numbers of each cover sentence and Stego sentence, the researcher detached the message embedding and extraction techniques from them. The steganography architectural design is the process of combining text and data for the text-mining purpose.(Ala'a M. Alhusban, Jehad Q,2017)

Author have decoupled the message embedding and extraction processes from the Cover Sentences, Stego Sentences. Given the line Numbers of every cover sentences and stego sentences, the cover word is given from the cover sentence using the standard of ASCII Control characters of a cover word. The same process is used for other cover sentences.(K. Wang, Q. Gao,2019)

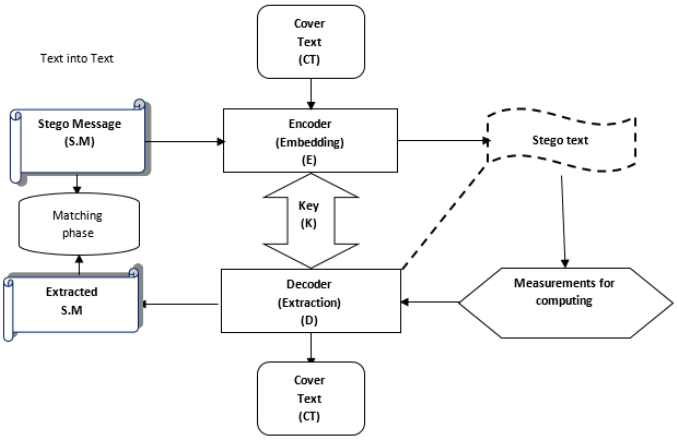


Fig 6. The process of hiding Stego message scheme.

Encryption, decryption, data, and concealment are all techniques used to encode, decrypt, encode, and decode secret communications. It is believed that attempting to satisfy several communications at the same time will considerably improve the effectiveness and efficiency of ciphertext encryption. There are five essential aspects of a data hiding technique that can be utilized to guide the construction of a good security of search effectiveness and efficiency.

Newly introduced hiding secret data strategy involves blending text into text as secret information, composed of the following processes: process of embedding secret messages such as (text, ciphertext, cover text), detected and extract secret message processes. To show the efficiency, security and imperceptibility performance of our scheme, executive tests have taken place for the steps of the proposed system.(R. Gurunath et al,2021)

4.1. EMBEDDING FUNCTION PROCESS

Figure 7 shows the process of hiding the secret message such as text/ cipher text / text into text in linguistic domain transformation text

steganography as the semantic or lexical transform. It described as embedding and extraction processes from the Cover Sentences, Stego Sentences. The cover text of a different size of sentences is transformed into an ASCII Control characters and number

of lines and sentences.(M. Taleby Ahvanooey et al,2018)

Embedding algorithm

Input: *f* indicates as a Cover text of size M_1 *x* M_2, $W_I \epsilon$ {-1, 1} be the text, text and shared a secret key.

Output: Stego Text

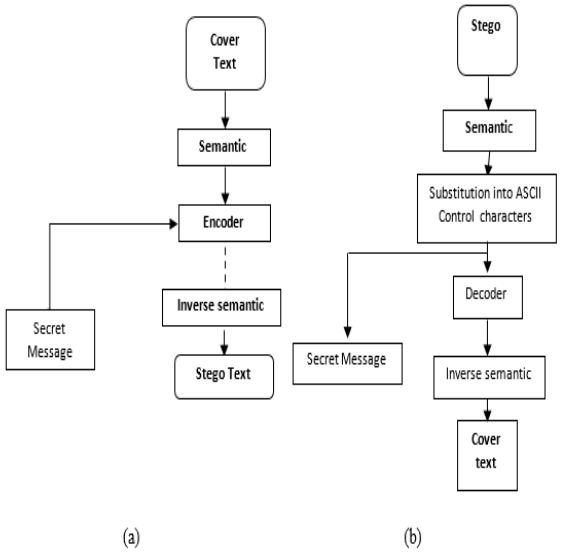


Fig 7. Embedding and Extracting Stego text scheme (a) Embedding, (b) Extracting.

4.2. TEXT QUALITY MEASUREMENT FOR COMPUTING EVALUATION

The text quality technique would be to calculate the mean values and variances of a few tiny places in the text and compare them to the original and hidden data content. In the last two decades, various objective quality measuring methods for text quality evaluation have been created. They are based on numerical text quality indicators.(Rajeev Kumar,et al,2016)

The Mean Square Error (MSE) and Normalized Correlation (NC) are the most widely used Objective

quality measurements for picture evaluation. They are valuable in the optimization context

because they are simple to calculate and have obvious physical implications. The MSE and NC are also mathematically convenient ways to analyze text and text quality.(Tapodhir Acharjee,et al,2020)

Mean Square Error (MSE)

The squared difference between the embedded and original text serves as a measure. It can be used for an image of size N x M defined as.

$$MSE = \frac{1}{MN}\sum_{j=1}^{M}\sum_{k=1}^{N}(x_{j,k} - x'_{j,k})^2$$

(5.1)

Where X (n, m) is the original text and X' (n, m) is the embedded text.

Standard Deviation (SD):

$$SD = \sqrt{\frac{\sum(Xi-\mu)^2}{N}}$$

(5.2)

The standard deviation is a statistic that indicates how much variance or dispersion there is in a group of statistics. A low standard deviation indicates that the values are close to the set's mean, whereas a high standard deviation means they are scattered over a larger range. SD denotes population standard deviation, N denotes population size, and Xi denotes each value from the population.(Q. M. Ashraf et al,2016)

Cross - correlation with a normalized coefficient (NC)

$$NC = \sum_{j=1}^{M}\sum_{k=1}^{N} x_{j,k} \cdot x'_{j,k}$$

(5.3)

In this section, a novel technique using a hybrid of text and text is offered for the purpose of hiding information. In order to provide more security and limit the distortion of the stego text, an alternative means of encrypting the data is presented.(Wei Zhang, et al,2020)

5. RESULTS AND DISCUSSIONS

In this chapter, we present and discuss the Chapter 1 Results and Discussions of our text Steganography by hiding text into another text through linguistic text steganography. The results and analyses of the experiments are available. To raise the capacity rate and security of cover text, all of the strategies discussed previously were used.(Toru Segawa,2016)

The aim of our study was to ensure the security and imperceptivity of the proposed algorithm. It is studied using linguistic (Semantic / lexical) approaches to measure the

differences between the cover text and the StegO text using MSE, DV, text histogram, and NC.(Anandaprova Majumder,2018)

Table 1: A New Algorithm to Hide a Secret Text in Another Text. | Stego

Line	Cover Sente	Stego Sente	Word	Cover wc	ASCII Co characte Cover w	ine no + A ontrol cha Numners) word	Stego W	line and v numbe	Lin	ASCII Co raracters (Stego wc	ASCII Co aracters (Stego w	ine no + A ontrol cha Numbers) word
1	On Thursd hammad wer new hous	ot the targ oil depo	1	sings	4ACK	46	weapo	3BS	3	8	BS	38
2	r that, eat din am	home unti ew military comes	2	bird	4ENQ	45	use	2EOT	2	3	eot	24
3	the city of D located the l beautiful ho	is guarded tary with v	3	voice	4LF	410	hit	4EOT	4	4	eot	44
			4	house	3VT	311	depo	1BEL	1	7	bel	17
4	the morning ngs with a be voice	the evening tank	5	Thursda	1STX	12	the	1STX	1	2	stx	12
			6	on	1SOH	11	shoo	1SOH	1	1	soh	11

6. CONCLUSIONS AND FUTURE WORK

6.1. CONCLUSIONS

The great advancement is currently a hot topic in both the corporate and public sectors because to the recent increase in interest in information hiding strategies. Utilize to protect the system's integrity and stop the exploitation of digital media by criminals. In recent years, steganography and cryptographic approaches have become well-known sub-disciplines of information concealment.

Measure Square Error (MSE), Normalized Correlation (NC), and Normalized Cross–correlation. Mean Squared Error (NC) are all ways to analyze text and text quality quantitatively in proposed system. Process of using prosed system given as the cover word is

given from the cover sentence using the standard of ASCII Control characters of a cover word.

6.2. FUTURE WORK:

The proposed system's efficacy and efficiency can be improved and enhanced in terms of capacity, security, impeccability, and robustness. It is hoped that the limitations of this work may serve as a springboard for additional investigation. One crucial component of data concealing is to take advantage of the cover object's redundancy in order to save space for data embedding.

Deep learning-assisted text generation allows for semantic control, which is especially useful for lengthy manuscripts. The usage of steganography and encryption methods and techniques is crucial for giving the embedding procedure an additional layer of security. As a result, it is possible to examine an additional security layer for embedding methods using non-sequence or random embedding spots.

REFERENCES

1. Majeed, M.A.; Sulaiman, R.; Shukur, Z.; Hasan, M.K. "A Review on Text Steganography Techniques," Mathematics, 2021, 9, 2829. https://doi.org/10.3390/math9212829
2. Neha Rani, Jyoti Chaudhary "Text Steganography Techniques: A Review," International Journal of Engineering Trends and Technology (IJETT) – Volume 4 Issue 7- july 2013.
3. K. V. Kale, Najran N. H. Aldawla, M. M. Kazi, "Steganography Enhancement by combining text and image through Wavelet Technique ," in International Journal of computer & Applications (IJCA), Vol.51 No.21, pp. 0975 – 8887, August 2012.
4. S. Katzenbeisser and F. A. P. Petitcolas, "Information Hiding Techniques for Steganography and Digital Watermarking," Norwood, MA, USA: Artech House, Inc,2000.
5. F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding-a survey," Proceedings of the IEEE, vol. 87, pp. 1062, 1999.
6. N. F. Johnson and S. Jajodia, "Exploring Steganography: Seeing the Unseen," IEEE Computer, pp. 26-34, Feb 1998.
7. Ross J. Anderson and Fabien A. P. Petitcolas, "On the Limits of Steganography" IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 16, NO. 4, MAY 1998.
8. Ari Moesriami Barmawi, "Linguistic Based Steganography using Lexical Substitution and Syntactical Transformation," IEEE, 2016.
9. Deepali Bhat, Krithi V, Manjunath KN, Srikanth Prabhu, Renuka A. "Information Hiding through Dynamic Text Steganography and Cryptography" 2017 , IEEE
10. Aleksandr A. Maliavko, "The Lexical and Syntactic Analyzers of the Translator for the El Language" 2018 , IEEE.
11. Ei Nyein Chan Wai, May Aye Khine, "Syntactic Bank-based Linguistic Steganography Approach," 2011 International Conference on Information Communication and Management IPCSIT vol.16 (2011) © (2011) IACSIT Press, Singapore
12. Alfin Naharuddin, Adhi Dharma Wibawa, "A High Capacity and Imperceptible Text Steganography Using Binary Digit Mapping on ASCII Characters", 2018 , IEEE , (ISITIA).
13. Shivani Sharma, Dr. Avadhesh Gupta, Munesh Chandra Trivedi, Virendra Kumar Yadav," Analysis of Different Text Steganography Techniques: A Survey" , 2016 Second International Conference on Computational Intelligence & Communication Technology, IEEE , 2016
14. N.R. Zaynalov,et al:," UNICODE For Hiding Information In A Text Document", 14th International Conference on Application of Information and Communication Technologies (AICT), IEEE, 2020.
15. Ala'a M. Alhusban and Jehad Q. Odeh Alnihoud, "A MELIORATED KASHIDA BASED APPROACH FOR ARABIC TEXT STEGANOGRAPHY", International Journal of Computer Science & Information Technology (IJCSIT) Vol 9, No 2, April 2017.
16. K. Wang, Q. Gao: A Coverless Plain Text Steganography Based on Character Features", VOLUME 7, IEEE .2019.
17. R. Gurunath et al.: "Novel Approach for Linguistic Steganography Evaluation Based on Artificial Neural Networks", VOL 9, IEEE, 2021.
18. M. Taleby Ahvanooey et al.: "Innovative Text Steganography Technique for Hidden Transmission of Text Message via Social Media", VOL 6, IEEE,2018.
19. Rajeev Kumar,et al:,"A Space based reversible high capacity text steganography scheme using Font type and style", ICCCA, IEEE, 2016.
20. Tapodhir Acharjee,et al:,"XORSTEG: A New Model of Text Steganography",
21. J. R. Jayapandiyan et al.:, "Enhanced Least Significant Bit Replacement Algorithm in Spatial Domain of Steganography" Vol 8, IEEE, 2020.
22. Cihan Yakar, Suat Ozdemir," Steganography Application for UTF8 Encoded Texts".

23. Q. M. Ashraf et al.:"TOPSIS-Based Service Arbitration for Autonomic Internet of Things", VOL 4, IEEE, 2016.
24. F. X. K. Akotoye, et al:,"CHARACTER PAIR TEXT STEGANOGRAPHY BASED ON THE ENHANCED PARAGRAPH APPROACH"
25. Wei Zhang, et al:, "Coverless Text Steganography Method Based on Characteristics of Word Association", 20th International Conference on Communication Technology, IEEE, 2020.
26. Toru Segawa, "Low-Power Optical Packet Switching for 100-Gb/s Burst Optical Packets With a Label Processor and 8 × 8 Optical Switch", JOURNAL OF LIGHTWAVE TECHNOLOGY, VOL. 34, NO. 8, APRIL 2016.
27. Anandaprova Majumder, Suvamoy Changder,"A Generalized Model of Text Steganography by Summary Generation using Frequency Analysis", IEEE, 2018.
28. N.Subramanian,et al.:,"Image Steganography: A Review of the Recent Advances",VOL 9,IEEE, 2021.
29. Keshav Joshi, "A New Approach of Text Steganography Using ASCII Values", IJERT, Vol. 7 Issue, May,2018. http://www.ijert.org.
30. Mohammed Abdul Majeed, et al:, "A Review on Text Steganography Techniques", Mathematics 2021, 9, 2829. https://doi.org/10.3390/math9212829.
31. S.Changder, D. Ghosh, N. C. Debnath, "Linguistic Approach for Text Steganography through Indian Text", 2nd International Conference on Computer Technology and Development 2010, IEEE, 2010.
32. P. Rajba, W. Mazurczyk, "Information Hiding Using Minification" VOL 9,IEEE, 2021.
33. Shalaw Mshir, Asaf Varol, "A New Model for Creating Layer Planes Using Steganography for Text Hiding", IEEE, 2019.
34. Yueyao Xu, "Unsupervised Deep Learning for Text Steganalysis", 2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI),IEEE,2020.
35. Henning Titi Ciptaningtyas, et al:, "Text Steganography on Sundanese Script using Improved Line Shift Coding".
36. Zhongliang Yang, et al:, "A Fast and Efficient Text Steganalysis Method".
37. Nuzhat Naqvi, et al:, "Correction to: Multilayer Partially Homomorphic Encryption Text Steganography (MLPHE-TS): A Zero Steganography Approach", Springer Science+Business Media,2019. https ://doi.org/10.1007/s1127 7-018-5868-1.
38. Afra Ibrahim Alaqeel, Mohamed Saad Saleh, "Developing a Performance-based Tool for Arabic Text Steganography",2021 National Computing Colleges Conference (NCCC), IEEE, 2021.
39. Sunita Chaudhary, et al:, "Aggrandize text security and hiding data through text steganography", IEEE,2016.