International Science and Technology Journal
المجلة الدولية للعلوم والتقنية
Volume 31 العدد
November 2022 نوفمبر

A New Algorithm to Hide a Secret Text in another Text

Eng . Saad bin Nasser Al Azzam

Researcher in Business Administration - Private Law - Cyber Security - Smart Cities
College of Computing - Department of Cyber Security - University of Bisha
College of Law and Judicial Studies - Private Law - University of Jeddah
College of Business - Department of Business Administration - King Khalid University
Kingdom Saudi Arabia
Snazzam.199 @gmail.com

الملخص :

هدفت الدراسة الى محاولة التعرف علي متطلبات هذه التقنية تبادلًا سريًا للبيانات ، ومن ثم حاول البشر عددًا من الطرق لتوفير الوصول إلى البيانات بسرية تامة. عززت الوسائط الرقمية الوصول إلى البيانات والنقل والكفاءة والدقة. أصبح من السهل بشكل معقول اعتراض البيانات المرسلة عبر الشبكات أو الوصول إلى مجموعة من الأجهزة ، سواء كانت مرتبطة أم لا ، لعرض المحتوى أو سرقة البيانات الحساسة أو التدخل فيها. يخفي الاتصال من خلال حمولات الوسائط المتعددة مثل النص والصورة والصوت والفيديو. يهدف هذا المشروع إلى تحسين إخفاء المعلومات من خلال دمج تركيب النص والصورة للتشفير غير المرئي والأمان والمتانة في الصور الرقمية. يقدم هذا العمل نوعًا جديدًا من إخفاء الصور الرقمية الذي يتفوق على الأساليب التقليدية في التشفير وعدم الإدراك والأمان وإخفاء النص. تبدأ أساسيات علم إخفاء المعلومات والأدبيات ذات الصلة فصلنا. يقترح هذا التحليل طريقة موحدة ومقاييس أداء لتقنيات إخفاء المعلومات. تحسين نظرية الاتصالات المخفية القائمة على نظرية المعلومات والأمان النظري وقيود القوة لفئة معينة من أنظمة إخفاء المعلومات تلبي هذه الاحتياجات. يدرس بحثنا المقترح إخفاء المعلومات عن طريق إخفاء النص والنص باستخدام مناهج كمية قياسية. تمت مقارنة خوارزمية إخفاء النص الجديدة الخاصة بنا مع طرق مجال النص الأخرى. اختبرنا العديد من الكلمات والجمل لأن الدلالات تخفي الرسالة السرية. استخدم نظامنا أنواعًا متعددة

من الخطوط ، وتم اختبار كل منها لمعرفة قوة الرسالة السرية. الخوارزمية المقترحة لدينا تلبي أعلى معايير الأمان والإدراك والقدرات. النص المخفي في نص آخر يستخدم الجودة في النظام المقترح يتضمن متوسط الخطأ المربع، الارتباط الطبيعي، الارتباط التبادلي الطبيعي، تحليل الرسم البياني، الانحراف المعياري، والاختبار والتحليل الإحصائي. تصف جملة الغلاف كيفية استخدام نظام نثر باستخدام أحرف التحكم. بالنظر إلى أرقام الأسطر لكل جملة غلاف وجمل ستيجو ، تتم معالجة مصطلحات جملة الغلاف الأخرى بشكل مشابه.

الكلمات المفتاحية: التقنيات اللغوية ، الطرق الدلالية ، أحرف التحكم

Abstract:

This technology requires confidential data interchange, hence humans attempted a number of methods to provide data access with complete confidentiality. Digital media has enhanced data access, transport, efficiency, and accuracy. Intercepting data sent over networks or gaining access to a range of machines, whether linked or not, to view content, steal sensitive data, or interfere with it has become reasonably easy. Steganography hides communication through multimedia payloads like text, image, audio, and video. This project aims to improve steganography by integrating text and image synthesis for invisible encryption, security, and robustness in digital images. This work introduces a new type of digital image steganography that outperforms conventional approaches in encryption, imperceptibility, security, and text concealing. Steganography basics and relevant literature start our chapter. This analysis suggests an uniform method and performance measures for steganographic techniques. Improved information theory-based steganographic communication theory and theoretical security and robustness constraints for a certain class of steganographic systems meet these needs. Our suggested research studies steganography by hiding text and text using standard quantitative approaches. Our novel text steganography algorithm was compared to other text domain methods. We tested many words and sentences because semantics hide the secret message. Our system used multiple line types, and each was tested for secret message strength. Our proposed algorithm meets the highest security, perception, and capability standards. Text hidden into another text employing quality in

the suggested system includes Mean Square Error (MSE), Normalized Correlation (NC), Normalized Cross–correlation, Histogram Analysis, Standard Deviation, and Statistical Test and Analysis. The cover sentence describes how to use the prossed system using ASCII Control characters. Given the line numbers of each cover sentence and Stego sentence, other cover sentence terms are processed similarly.

Keywords: Linguistic Techniques, semantic methods, ASCII Control characters

Introduction

This technique terms of linguistic steganography use to conceal sensitive information within text files. It focuses on neural linguistic steganography approach was proposed as a linguistic steganography methodology. It can create a steganographic paragraph based on knowledge graphs that generates coherent multi-sentence texts for concealment. It was suggested to use a subject neural linguistic steganography technology for linguistic steganography. It is focuses on the knowledge graphs, it can create a steganographic sentence with a specified theme to produce coherent multi sentence uses for writings data.

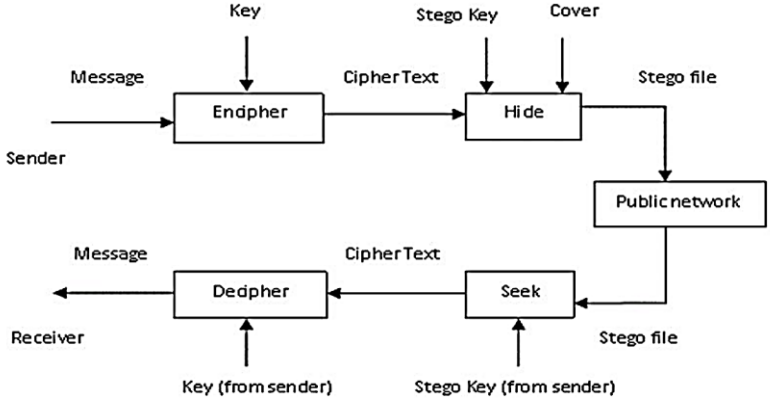

Figure 1 The Proposed Model of linguistic Text Steganograph

This method uses linguistic steganography to conceal data in plain sight inside text documents. Topic-aware neural-linguistic steganography is presented as a linguistic steganography methodology in [1].

Using knowledge graphs, it may produce steganographic paragraphs on a given subject (KGs). If you want to construct cohesive multi-sentence writings for improved concealment, a KG can provide you information about pertinent themes and material. The suggested approach reveals the precision and accuracy of the created steganographic text.
In [2], the author presented a solution to the problem of unmanageable semantic expression in steganographic texts produced by neural networks. A new obstacle that steganography models will have to overcome in the future is control cognitive imperceptibility, which the author discussed.
The Gated Recurrent Unit (GRU), Transformer, and Topic-Aware encoder models were evaluated to determine which is best for semantic extraction.
In order to produce the candidate pool from which words are randomly drawn according to the overall conditional probability distribution, categorical sampling generates steganographic phrases. Evidence from experiments demonstrates that the suggested technique may also place restrictions on the semantic expression of the steganographic text that is created.
In [3], an autonomous steganographic text generator based on an adaptive probability distribution and a generative adversarial network for languages other than English was developed. The suggested approach effectively dealt with the exposure bias that resulted from the gap between the training and inference phases.
To further minimize the ensuing discrepancy, the suggested technique calculates the candidate word space and embedding capacity based on the likelihood of word similarity. According to the results of the trials, the suggested technique provides a potential way to improve steganographic security, especially in comparison to the prior models' anti-stainless capacity.
Based on the capabilities of a commercially available language tool, Ref. [4] proposes a safe generative linguistic steganographic approach

that recursively embeds secret information using Adaptive Dynamic Grouping (ADG).
The suggested technique dynamically embeds hidden information in the created stego text by grouping tokens according to their probability at each stage, making the embedded data less detectable.
The suggested approach was shown to generate readable stego texts and provides a good level of security in experimental settings.

In [5], a different linguistic steganographic approach based on the Variational Auto-Encoder was suggested, which can create the stego text automatically (VAE-Stega). The aim of this strategy was to increase the stealth and safety of the steganographic texts it created.
The VAE-Stega encoder is primarily used for two tasks: generating steganographic phrases and learning the statistical distribution properties of large-size regular texts. The suggested approach increases the stealthies of the produced steganographic phrases, as shown by the experiments.

In [6], the author developed a generative text steganographic approach using a long short-term memory (LSTM) network to build a language model from a large-scale regular text database. The produced word is verified using the LSTM network's specified conditional possibility distribution (of words) and the hidden value on the receiving end. On the receiving end, the same paradigm is utilized to decipher the code. The findings demonstrated its superiority above other efforts in its field.
In [7], the use of Recurrent Neural Networks (RNN-Stega) to automatically generate text covers from a discrete bitstream was suggested. The conditional distribution of encoding words is analyzed using fixed- and variable-length coding (FLC and VLC), respectively. The trials demonstrated a high capacity for embedding and a high degree of protection against malicious efforts.

In [8], a new approach using synonym substitution to improve linguistic steganography's embedding capability is proposed. Messages may now be concealed inside MS Word documents thanks to a modified change-tracking system that makes use of HC.

Commonly used synonyms may be used to conceal information and avoid suspicion from an outside observer.

In a similar vein, [9] proposes sampling to generate language with improved concealing capability. Mathematics is used to encode information in the cover text (AC). These findings are compared to those obtained by employing FLC and VLC, both of which are based on the use of deterministic methods to create text in order to conceal information.

Additionally, the Kullback-Leibler divergence (KL) measures the dissimilarity between two probability distributions of a given variable x that are used to control FLC embedding. In addition, VLC and AC are controlled in their own unique ways using divergence and temperature-based methods, respectively. Experimental comparisons of the AC technique's embedding performance to those of VLC and FLC revealed that the AC method performed better.

In [10], where the focus is on semantic text-based steganography, the idea of linguistic text steganography is proposed. To counteract potential security risks, banks will often use synonyms to obscure information in credit letters. These precautions are especially crucial now that LCs may be utilized digitally.

In [11], a further method of text steganography was developed using Arabic symbols to conceal information. Datasets are drawn from collections of adages and idioms. For this purpose, we use the Adhocratic algorithm (AC*) and a particular typeface (Naksh). This method's safety and breadth of coverage were evaluated. It was discovered that various Arabic geometric forms meet the criteria for steganography.

A novel linguistic text-based steganography method was proposed in ref. [12], whereby the normalized sentences are searched for the existence of commonly used letter sets or double letter pairings in the words of the paragraph. The system picks out the phrases to include in its summary, or cover text. This procedure extracts characteristics from the source language's letters and utilizes them to classify the letters. There are three distinct types of visual representations for the English alphabet. Sets of letters that include slanted lines, curved lines, or standing and sleeping lines are denoted by the letters LC, LS, and LSS,

respectively. A secret bit and its corresponding representation are identified by counting the number of individuals in each set.

Word-indexing compression method (WIC), suggested in [13], is another kind of linguistic steganography that may shorten the length of the experimental embedded payload. Conversely, a stego text selection approach allows for the ideal undetectable stego content to be picked among candidates. The candidate cover text is used in conjunction with a minimum-maximum weight algorithm and HC to compress the hidden message using the WIC method.
This method additionally tries to strengthen anti-stainless by using synonym replacements to incorporate 10 cover texts with low compression ratios within the corresponding compressed secret message.
Given a principle derived from the distance between a cover text and its stego text, only one stego text is selected. The suggested compression method outperforms HC and LZW coding approaches in anti-steganalysis when combined with the stego text selection criterion, and experimental findings demonstrate that this leads to a larger embedding capacity.
In [14], an alphabetic transformation approach for linguistic text steganography was presented to convert the multi-variable concealed message into alphabets. The problem of picking appropriate cover messages is therefore solved using this method.
The suggested method is a blind embedding approach that makes the hidden message completely undetectable by replacing a character with one from the cover message. According to the findings, the suggested method performs better than the existing ones in terms of its ability to conceal sensitive data.

In [15], they describe a technique for carrier-free multi-keyword text steganography by using the POS as a hidden key. Each component of the Chinese characters that make up the words is used to choose the hidden tags. You may conceal the total amount of keywords by using the POS. Also, a POS tag that is appropriate for each term in the phrase has been assigned.

After words have been segmented, their POS may be counted to gain the mapping set between POS and numbers. Since hidden tags are picked from all the Chinese character parts of the word, the POS is used to disguise the number of keywords buried in each stego text, increasing the concealing capacity.

Suggested System To New Algorithm

The suggested system is based on the semantic linguistic transform, and it focuses on how to improve secure communication and imperceptibility of secret data by concealing text and text of hiding techniques such as text, embedding, and extraction.

1. Initialize the secret key and pseudo-random number.
2. Create the Cover text's detail coefficients.
3. Create a random key of embedding process used to produce a traversing order for visiting the ascii code and line no of sentences.
4. If the secret message length ≤ the key Then Embed the secret message into the details of the n
5. Use the substitution on different lines and with different numbers of words. The embedded

Algorithm Sender : Step 1: Select a text, key: =ø // Initially, key is empty

Step 2: Read the text word-by-word

for each read word in the text do

if the word is selected then

// Processing of the selected word.

// The selection of a word will be done using a random function, i.e.

// if the outcome of the random function is 1 then the word will be

// selected else it will not be selected,

ask the admin for the new target meaning to the selected word

if the selected word does not appear in the stego_table then

insert the selected word in the stego_table

assign the target meaning to the selected word in the stego_table

else // the selected word appears in the stego_table

if (old target meaning of the selected word)
(new target meaning of the word) then,
replace the (old target meaning of the word) by the,
(the new target use a meaning of the word) in the stego_table

Endif
endif

append (line_number(selected word), word_number (selected word)) at the end of
the key
// line_number(selected word) will be represented by the corresponding
// character in the ASCII code table. Exp: the number 82 will be represented // by the
character ‘R’

endif
end for
Ask admin for the phrase to insert at the bottom of the text,
insert given phrase at the bottom of the text,
append the key at the end of the inserted phrase.

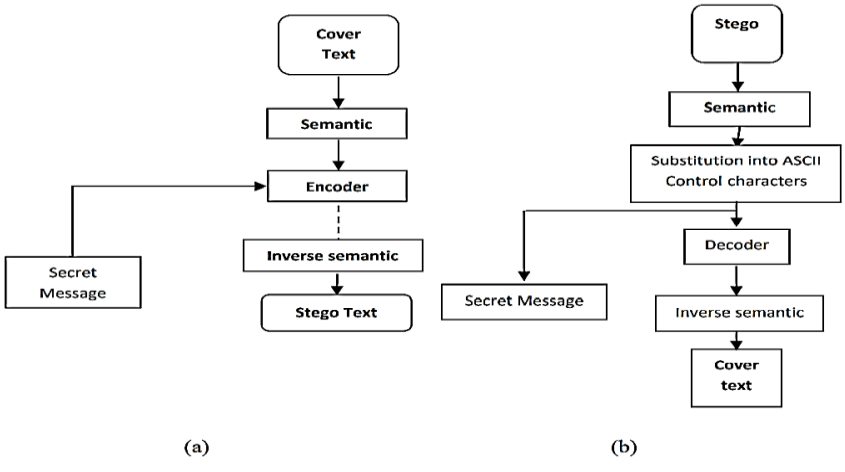

Figure 2 Embedding and Extracting Stego text scheme (a) Embedding, (b) Extracting.

Results:

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.

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

Line NO	Cover Sentences	Stego Sentences	Word No	Cover word	ASCII Control chara	Line no + ASCII Contr	Stego Word	line and word	Line no	ASCII Control	ASCII Control chara	Line no + ASCII Contr
1.	On Thursday, Muha	Shoot the target at the oil	1	sings	4ACK	46	weapons	3BS	3	8	BS	38
2.	After that, eat dinner at 9	Go home until use a new	2	bird	4ENQ	45	use	2EOT	2	4	EOT	24
3.	In the city of Dammam is located the	Oil is guarded by the military	3	voice	4LF	410	hit	4EOT	4	4	eot	44
			4	house	3VT	311	depot	1BE	1	7	bel	17
4.	In the morning the bird sings	In the evening hit oil tank	5	Thursda	1STX	12	the	1STX	1	2	stx	12
			6	on	1SOH	11	shoot	1SOH	1	1	soh	11

The improved results shows in Tables (1) and (Figures (3) in the form of cover texts and Stego text of Mean, Standard Deviation, and Standard Error mean, where the result of linguistic (semantic) technique obtained the best result.

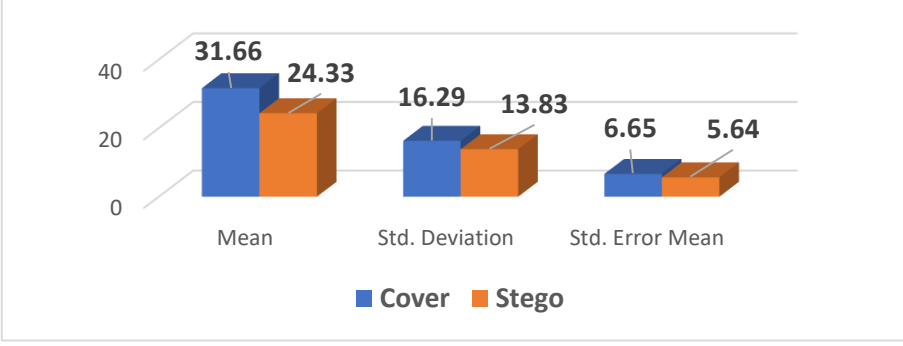

Figure 3show the cover text and Stego text comparison. The extracted results are to the cover texts and the Stego text of Mean, Standard Deviation, and Standard Error mean.

Table 2 show the various test results of the measurements that applied to the cover text and stego text. The extracted results are to the cover texts and Stego text of Mean, Standard Deviation, Error mean and correlation.

	Mean	Std. Deviation	Std. Error Mean	Correlation	Sig.	T	Sig.
Cover	31.6667	16.29315	6.65165	.819	.046	1.916	.113
Stego	24.3333	13.83715	5.64899				

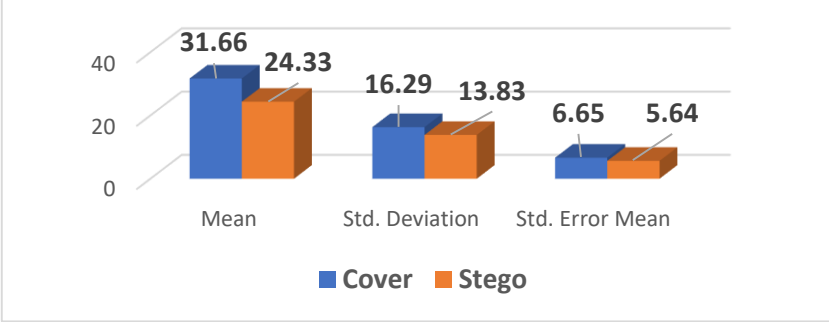

Figure 4 show the cover text and Stego text comparison. The extracted results are to the cover texts and the Stego text of Mean, Standard Deviation, and Standard Error mean.

The performance of our suggested technique is compared to the cover text and a hidden secret message in figure 5.2, where the mean, standard devotion, and standard error mean are shown for six words in four lines of sentences for comparison with our proposed method.
The placement of a hidden string of Cover text in front of Stego text that is independent of cover text. This method embeds secret bits by selecting a line in Stego text for the structure of our algorithm, which consists of six words drawn from four lines of sentence using ASCII Control characters.

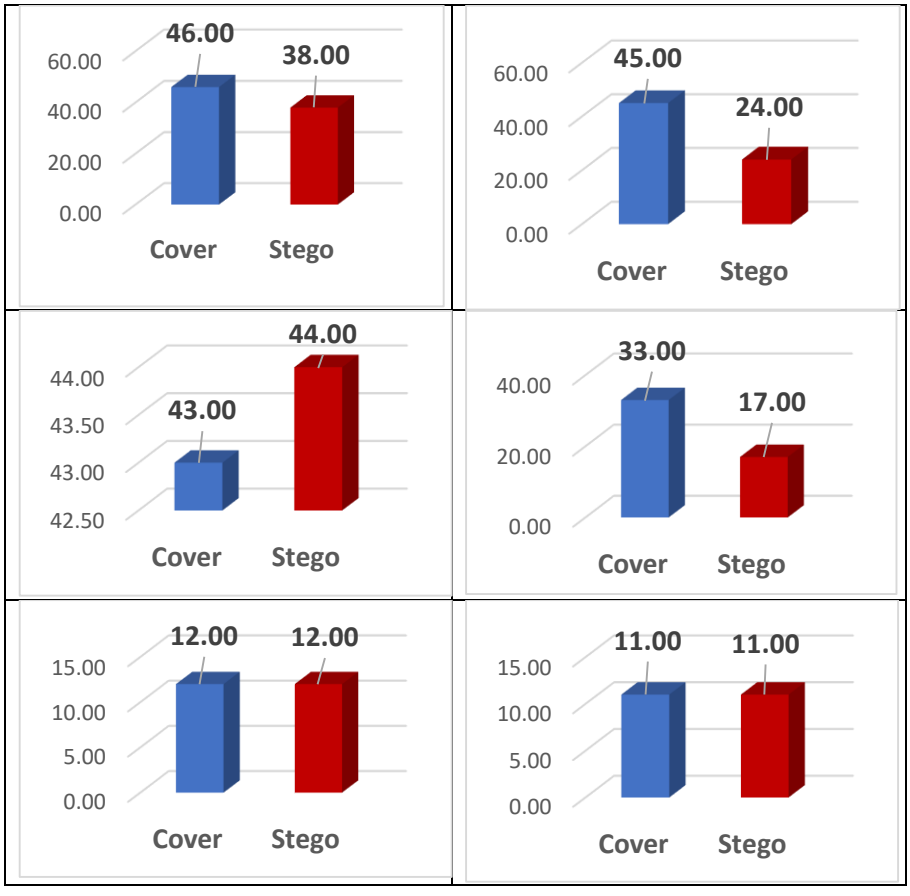

Figure 5 show the histogram of six Cover text and (Stego text) comparisons

the given the Histogram of cover and Stego text for six words choosing from a four lines of sentences given the New Algorithm to Hide a Secret

Text in Another. No significant visual difference is present after the embedding procedure.

Comparisons with various text domain algorithms were made to provide the ultimate judgment of our novel text steganography algorithm. We tested on many words and sentences because the key uniqueness of our strategy is the way semantics are used to hide the secret message. Several types of lines were used in our implementation, and the strength with which the secret message could survive was measured for each of them. Our proposed system's algorithm appeared to meet the highest levels of security, perceptions, and capacity.

DISCUSSION AND CONCLUSIONS

DISCUSSION

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 with malevolent intent. In recent years, steganography and cryptographic approaches have become well-known sub-disciplines of information concealment.

Our proposed work investigates that research of text steganography looked at the evolution of text steganography and how new technologies are being used in text steganography as linguistics or semantic process. To help researchers by compiling current methods, these technologies were linked with existing steganography categories. The Mean Square Error (MSE), Normalized Correlation (NC), and Normalized Cross–correlation Mean Squared Error, Histogram Analysis, Standard Deviation, Statistical Test and Analysis are all standard ways to analyze text and text quality quantitatively in proposed system.

The 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, and the same method is used for additional words from the cover sentences, given the line numbers of every cover sentence and Stego sentence. While the Stego word extraction process employs Stego Sentences processes, given the line numbers of each Stego word, and the Stego word is extracted from the Stego sentence using the standard

of ASCII Control characters of a Stego word, the same procedure is utilized for other words from Stego sentences.

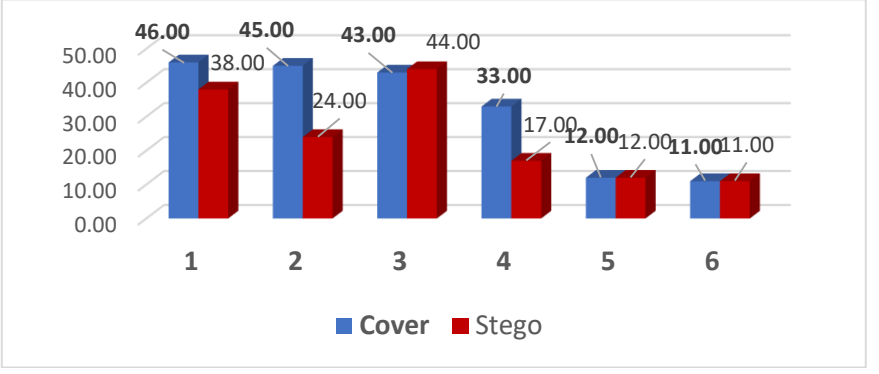

Figure 6 show all six histogram of Cover text and (Stego text) comparisons

In the existing metrics above shown in fig 6.2 focused on text hiding for measurement like "MSE", "Standard Divisions", "Standard error mean", are having major goals, such as that the embedded data must be invisible to the observer, including the observer's tools like computer analysis, and that it should have the capacity and security options, where this measurement is nearer to the original image. In our new algorithm of hiding text in another text uses above metrics and text quality measurement used to judge the existence of secret message between the cover text and Stego text. The new algorithm of hiding text in another text gives a good method of hiding for the purpose of security and imperceptible that improves the method respectively.

The results of existing related hiding text in other text steganography systems are classified using the cover text and Stego text methods' extended categorization of ASCII standard characters with line of sentences.

These issues of information hiding relate to the A New Algorithm to Hide a Secret Text in Another Text and Stego the following list specifies the main contributions of this experimental work:

- Scientific literature in the scope of digital text Steganography (hiding text into another text) surveyed and systematically summarized.

- An extended classification of text Steganography is presented to clarify advantages of text steganographic method over other currently existing methods.
- The existing related hiding text and text steganography approaches are classified according to the extended classification.
- A novel of hiding text and application-friendly approaches to digital steganography is proposed.
- Several methods for the text steganography extensions providing applicability are embedded and extracted with the basic algorithm such as Deep learning , Neural network for linguistic (lexical, semantic).
- The process of embedding and extracting algorithms work well in linguistic technique, which widely suggested and used in the field of information hiding.
- The security, capacity, impeccability and performance aspects of the proposed approaches are analyzed to survive against various hidden data.
- Experimental results are generated for a number of text and text sequences to determine the performance of the linguistic methods.

CONCLUSION

In this work we should be understood within the parameters of the specified scope in this thesis, research never ends, so it is anticipated that future study in cyber security will examine areas outside the parameters of this master's thesis. It is hoped that the limitations of this work may serve as a springboard for additional investigation.

The proposed system's efficacy and efficiency can be improved and enhanced in terms of capacity, security, impeccability, and robustness. It is difficult to find a method of text steganography required for both the security and robustness, as a result, the potential to develop a novel method to meet our needs is worth exploring. One crucial component of data concealing is to take advantage of the cover object's redundancy in order to save space for data embedding. It is worthwhile to research how to take advantage of the cover object's redundancy as much as feasible, and the cover object's properties should be further exploited for performance optimization.

Our technology has also been implemented in the media, such as sentences, pdf, MS-word, and many text formats. As a result, after our method has been proposed, the question of how to discover the message hidden in the media becomes critical. The types of secret messages between the data in the cover object will be included to the totally secure model in future theoretical study. Analyzing the chance that the hidden message can be resistant at a given data hiding rate will yield a more precise model.

In the field of text steganography, several algorithms have been successfully developed. As a result, the suggested algorithms are solely concerned with the concealment of text and images. These currently proposed technologies to different forms of multimedia, such as video, audio, and various types of image and text. 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. Further research into these coupled approaches is possible. The approach is subject to attack due to the sequential selection of embedding positions. As a result, it is possible to examine an additional security layer for embedding methods using non-sequence or random embedding spots.

References

[1] Li, Y., Zhang, J., Yang, Z., & Zhang, R. (2021). Topic-aware neural linguistic steganography based on knowledge graphs. ACM/IMS Transactions on Data Science, 2(2), 1-13.

[2] Yang, Z., Xiang, L., Zhang, S., Sun, X., & Huang, Y. (2021). Linguistic generative steganography with enhanced cognitive-imperceptibility. IEEE Signal Processing Letters, 28, 409-413.

[3] Zhou, X., Peng, W., Yang, B., Wen, J., Xue, Y., & Zhong, P. (2021). Linguistic steganography based on adaptive probability distribution. IEEE Transactions on Dependable and Secure Computing.

[4] Zhang, S., Yang, Z., Yang, J., & Huang, Y. (2021). Provably secure generative linguistic steganography. arXiv preprint arXiv:2106.02011.

[5] Yang, Z. L., Zhang, S. Y., Hu, Y. T., Hu, Z. W., & Huang, Y. F. (2020). VAE-Stega: linguistic steganography based on variational auto-encoder. IEEE Transactions on Information Forensics and Security, 16, 880-895.

[6] Kang, H., Wu, H., & Zhang, X. (2020). Generative text steganography based on LSTM network and attention mechanism with keywords. Electronic Imaging, 2020(4), 291-1.

[7] Yang, Z. L., Guo, X. Q., Chen, Z. M., Huang, Y. F., & Zhang, Y. J. (2018). RNN-stega: Linguistic steganography based on recurrent neural networks. IEEE Transactions on Information Forensics and Security, 14(5), 1280-1295.

[8] Mahato, S., Khan, D. A., & Yadav, D. K. (2020). A modified approach to data hiding in Microsoft Word documents by change-tracking technique. Journal of King Saud University-Computer and Information Sciences, 32(2), 216-224.

[9] Yang, R., & Ling, Z. H. (2019, November). Linguistic steganography by sampling-based language generation. In 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) (pp. 1014-1019). IEEE.

[10] Chaw, A. A. (2019). Text steganography in Letter of Credit (LC) using synonym substitution based algorithm. International Journal for Advance Research and Development, 4(8), 59-63.

[11] Hamzah, A. A., Khattab, S., & Bayomi, H. (2021). A linguistic steganography framework using Arabic calligraphy. Journal of King Saud University-Computer and Information Sciences, 33(7), 865-877.

[12] Majumder, A., & Changder, S. (2018, August). 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.

[13] Xiang, L., Wu, W., Li, X., & Yang, C. (2018). A linguistic steganography based on word indexing compression and candidate selection. Multimedia Tools and Applications, 77(21), 28969-28989.

[14] Naqvi, N., Abbasi, A. T., Hussain, R., Khan, M. A., & Ahmad, B. (2018). Multilayer partially homomorphic encryption text steganography (MLPHE-TS): a zero steganography approach. Wireless Personal Communications, 103(2), 1563-1585.

[15] Liu, Y., Wu, J., & Xin, G. (2017, July). Multi-keywords carrier-free text steganography based on part of speech tagging. In 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) (pp. 2102-2107). IEEE.