JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

TEXT STEGANOGRAPHY NEW APPROACH TO INFORMATION HIDING

[1]Gaurav Khatri, [2]V. Haripriya
[1]Masters Student, [2]Professor
[1,2]School of CS & IT,
Jain (Deemed-to-be University), Bengaluru, India

Abstract: Sending encrypted messages on a regular basis will attract the attention of third parties, such as crackers and hackers, perhaps leading to attempts to break and divulge the original messages. Steganography is used in the digital world to conceal the existence of communication by disguising a secret message within another suspicious message. Steganography is frequently used in conjunction with encryption and provides an appropriate level of anonymity and security over the communication channel. This paper provides an overview of text steganography as well as a brief history of steganography, as well as some extant text-based steganography approaches. Some of the fundamental limitations with text steganography, as well as issues with existing solutions, are highlighted. A hybrid strategy of inter-word spacing and inter-paragraph spacing is presented for information concealment. Our method generates dynamic stego-text with six maximum capacity possibilities based on the length of the secret message. This report also examined the major shortcomings of each existing strategy and how our new approach might be proposed as a solution.

***IndexTerms*:-**

I. INTRODUCTION

Steganography is a technique of hiding a message or information inside another seemingly harmless message or medium like an image, audio, or video file. The main objective of steganography is to keep the hidden message undetected by everyone except the intended recipient. These techniques can be broadly classified into two categories: *spatial domain and transform domain*. In the spatial domain, the message is concealed by modifying the color or pixel values of the cover image. For instance, the LSB (least significant bit) substitution can be used to hide a message by modifying the pixel values of the cover image. In contrast, in the transform domain, the message is hidden by modifying the frequency domain representation of the cover image, such as the discrete cosine transform (DCT) coefficients of an image or the Fourier transform coefficients of an audio signal. And, can be employed for several purposes, such as secret communication, digital watermarking, and copyright protection. However, it can also be exploited for malicious purposes like hiding malware or other malicious software within seemingly harmless files. Therefore, steganography is an area of interest for both security researchers and law enforcement agencies.

Text Steganography is a technique of hiding a message or information within a text document without altering its original appearance. This method has become increasingly popular with the widespread use of electronic text communication. The ultimate aim of text steganography is to hide the presence of the concealed message, making it challenging for anyone to detect.

Steganography techniques such as hiding messages within a text using inter-word and inter-paragraph spacing hybrid methods can be useful in situations where secure communication is required. This can be helpful for individuals or organizations that want to communicate sensitive information without attracting attention or risking interception by unauthorized parties. Steganography can also be used for digital watermarking and copyright protection, where a message or identifier is hidden within a file to prove ownership or authenticity. Furthermore, it can be used in the field of forensics to uncover hidden messages in digital files. However, it can also be used for malicious purposes, such as hiding malware or other harmful software within seemingly innocent files. Therefore, it is crucial for security researchers and law enforcement agencies to be aware of and study steganography techniques to detect and prevent any malicious use.

Information hiding is important because it allows individuals and organizations to protect sensitive or valuable information from unauthorized access, theft, or interception. By using techniques such as steganography, watermarking, or encryption, information can be hidden in plain sight or made unreadable without proper authentication. This can be particularly important in fields such as finance, healthcare, or national security, where confidentiality and integrity are essential. Information hiding with steganography is a technique used to conceal a message within an innocuous carrier medium, such as an image, video, or audio file, in such a way that it is difficult to detect by anyone who is not the intended recipient. Steganography can be used to hide sensitive or confidential information, digital watermarks, and copyright information.

Steganography is different from cryptography in that cryptography involves encrypting a message to make it unreadable, while steganography hides the existence of the message. In other words, cryptography obscures the content of the message, while steganography hides the fact that a message exists. One of the advantages of steganography is that it can be used to hide information within a larger amount of data, making it less suspicious. Additionally, steganography can be used in conjunction with cryptography to provide an extra layer of security.

Inter-word and inter-paragraph spacing can be used in steganography to hide information by adjusting the spacing between words and paragraphs in a way that is imperceptible to the human eye. This technique is known as whitespace steganography. To encode information using whitespace steganography, the first step is to determine the maximum amount of whitespace that can be modified without altering the appearance of the text. This will depend on various factors, such as the font size, type, and color, as well as the background color of the document.

II. LITERATURE SURVEY

The author begins by providing an overview of steganography and its various techniques, including image steganography and audio steganography. The focus of this paper, however, is on text steganography, which the author argues is a more challenging task due to the limited capacity of text documents. The proposed approach involves dividing the text message into several blocks and then embedding the secret message in the first letter of each block. The author suggests that this method is more secure than other text steganography techniques because it does not change the length or formatting of the original message. The author also discusses various attacks that can be used to detect the hidden message, including statistical analysis, frequency analysis, and linguistic analysis. The paper concludes by presenting experimental results that demonstrate the effectiveness of the proposed approach**.** [1]

The authors begin by providing an overview of steganography and its various applications, including data confidentiality, authentication, and integrity. They then discuss the limitations of existing text steganography techniques, which often involve modifying the text structure or content in ways that can be easily detected or compromised. To overcome these limitations, the authors propose a new steganography scheme that embeds hidden information in the white spaces of a text document. The scheme works by first dividing the text into blocks of equal length, and then identifying the white spaces in each block. The hidden information is then embedded in the white spaces using a binary encoding technique. To enhance the security and robustness of the hidden information, the authors introduce several key features of the Whitesteg scheme. These include the use of error correction codes, which can detect and correct errors that may occur during transmission, and the use of key-based encryption, which ensures that only authorized parties can access the hidden information. The authors also discuss several potential attacks on the Whitesteg scheme, including statistical analysis, linguistic analysis, and syntactic analysis. They propose several countermeasures to these attacks, including the use of randomization techniques and the manipulation of white spaces to reduce the detectability of hidden information. To evaluate the performance of the Whitesteg scheme, the authors conducted several experiments using different text documents and hidden messages. The results of these experiments showed that the scheme was able to embed and retrieve hidden information with high accuracy and reliability, while also maintaining the integrity and authenticity of the original text document. [2]

The paper begins by providing an overview of steganography and its various applications, including information hiding and watermarking. The authors then discuss the challenges of text steganography in Persian and Arabic languages, including the need to maintain the integrity and readability of the text while also hiding the hidden message. To address these challenges, the authors propose a new approach to Persian/Arabic text steganography that is based on the use of diacritical marks. Diacritical marks are small symbols that are used in Persian and Arabic languages to indicate changes in pronunciation or meaning. The proposed approach works by first dividing the text into blocks of equal length, and then embedding the hidden message in the diacritical marks of each block. The authors argue that this method is more secure and less detectable than other text steganography techniques, which often involve modifying the text structure or content in ways that can be easily detected or compromised. To evaluate the effectiveness of the proposed approach, the authors conducted several experiments using different Persian and Arabic texts and hidden messages. The results of these experiments showed that the approach was able to embed and retrieve hidden messages with high accuracy and reliability, while also maintaining the readability and integrity of the original text. The authors also discuss several potential attacks on their approach, including statistical analysis, frequency analysis, and linguistic analysis. They propose several countermeasures to these attacks, including the use of randomization techniques and the manipulation of diacritical marks to reduce the detectability of hidden messages. [3]

The paper begins by providing an overview of steganography and its various applications, including information hiding and watermarking. The authors then discuss the challenges of text steganography in Hindi language, including the need to maintain the integrity and readability of the text while also hiding the hidden message. To address these challenges, the authors propose a new approach to Hindi text steganography that is based on the use of shifting matras. Matras are vowel signs that are used in Hindi language to modify the pronunciation of a consonant. The proposed approach works by first dividing the text into blocks of equal length, and then embedding the hidden message in the matras of each block. The authors argue that this method is more secure and less detectable than other text steganography techniques, which often involve modifying the text structure or content in ways that can be easily detected or compromised. To evaluate the effectiveness of the proposed approach, the authors conducted several experiments using different Hindi texts and hidden messages. The results of these experiments showed that the approach was able to embed and retrieve hidden messages with high accuracy and reliability, while also maintaining the readability and integrity of the original text. The authors also discuss several potential attacks on their approach, including statistical analysis, frequency analysis, and linguistic analysis. They propose several countermeasures to these attacks, including the use of randomization techniques and the manipulation of matras to reduce the detectability of hidden messages. [4]

The paper begins with an overview of steganography and its various applications, including data confidentiality, copyright protection, and digital watermarking. The authors then discuss the challenges of text steganography, including the need to maintain the readability and naturalness of the text while hiding the secret message. To address these challenges, the authors propose a new synonym text steganography approach that works by replacing certain words in the text with their synonyms. The selection of words and synonyms is done in a way that does not change the overall meaning or context of the text, but provides a way to encode the hidden message. The authors argue that this method is more secure and less detectable than other text steganography techniques, which often involve modifying the text structure or content in ways that can be easily detected or compromised. To evaluate the effectiveness of the proposed approach, the authors conducted several experiments using different text documents and hidden messages. The results of these experiments showed that the approach was able to embed and retrieve hidden messages with high accuracy and reliability, while also maintaining the readability and naturalness of the original text. The authors also discuss several potential attacks on their approach, including statistical analysis, frequency analysis, and semantic analysis. They propose several countermeasures to these attacks, including the use of multiple synonym replacements, the use of uncommon synonyms, and the manipulation of the order and position of the replaced words. [5]

The paper begins with an overview of steganography and its various applications, including data confidentiality, copyright protection, and digital watermarking. The authors then discuss the challenges of text steganography in the context of e-mail communication, including the need to maintain the formatting and naturalness of the e-mail while hiding the secret message. To address these challenges, the authors propose a new approach that works by inserting hidden messages into the HTML formatting tags of an e-mail. The approach uses a codebook to map each character in the hidden message to a corresponding formatting tag, which is then inserted into the e-mail body. The authors argue that this method is more secure and less detectable than other e-mail based text steganography techniques, which often involve modifying the text structure or content in ways that can be easily detected or compromised. To evaluate the effectiveness of the proposed approach, the authors conducted several experiments using different e-mail messages and hidden messages. The results of these experiments showed that the approach was able to embed and retrieve hidden messages with high accuracy and reliability, while also maintaining the formatting and naturalness of the original e-mail. The authors also discuss several potential attacks on their approach, including statistical analysis, frequency analysis, and content analysis. They propose several countermeasures to these attacks, including the use of random formatting tags, the use of multiple hidden messages, and the manipulation of the order and position of the inserted tags. [6]

Steganography is the art of hiding secret information within a cover medium such as an image, audio, or text. In the case of SMS steganography, the cover medium is the text message itself. The goal of SMS steganography is to embed secret messages within SMS text messages in such a way that the embedded messages are invisible to the human eye and can only be detected using a steganalysis tool. The proposed technique in this paper uses a set of emoticons to represent binary digits (bits) of the secret message. The emoticons act as the cover medium for the secret message. The technique consists of three main steps: encoding, embedding, and decoding. In the encoding step, the secret message is first converted into binary digits using ASCII code. Each binary digit is then mapped to a specific emoticon from a predefined set of emoticons. The emoticon set consists of 40 different emoticons, with each emoticon representing a unique 5-bit binary code. In the embedding step, the binary-coded secret message is inserted into the cover text message using a predefined set of rules. The rules specify the placement of the emoticons within the text message and the order in which they appear. In the decoding step, the recipient extracts the secret message from the cover text message by decoding the emoticons back into binary digits using the predefined mapping between emoticons and binary codes. The binary digits are then converted back into ASCII code to retrieve the original message. The paper presents the experimental results of the proposed technique using two metrics: message distortion and capacity. Message distortion measures the amount of distortion introduced to the cover message due to the embedding of the secret message. Capacity measures the maximum length of the secret message that can be embedded within a given cover message length. The experimental results show that the proposed technique is able to embed secret messages with minimal distortion to the cover message. The average message distortion for the proposed technique is 0.36, which is significantly lower than the distortion introduced by other SMS steganography techniques. The proposed technique also has a high capacity, with a maximum secret message length of 170 bits for a cover message length of 160 characters. [7]

III. METHODOLOGY

The system that enables users to hide a secret message within a text using the inter-word and inter-paragraph spacing hybrid method is a steganography technique that uses the spacing between words and paragraphs to conceal the secret message. The objective is to modify the spacing between words and paragraphs in a way that the text appears identical but includes a hidden message.

To achieve this, the system generates stego-text dynamically by determining the maximum capacity of the text based on the length of the secret message. The system offers six choices for the maximum capacity, enabling the user to select how much of the text they want to use for hiding the message. Once the user chooses the maximum capacity, the system uses inter-word and inter-paragraph spacing to hide the message. Inter-word spacing refers to the space between words, while inter-paragraph spacing refers to the space between paragraphs. By adjusting the spacing between words and paragraphs, the system can insert pieces of the secret message without affecting the visible appearance of the text.

For example, if the user wants to hide the message "HELLO" in a paragraph of text, the system will calculate the maximum capacity of the text based on the length of the message and generate stego-text accordingly. The system will then change the spacing between words and paragraphs to insert the message, resulting in a text that looks identical but contains the hidden message.

Overall, the inter-word and inter-paragraph spacing hybrid method is an efficient steganography technique that enables users to hide messages within text without altering the visible appearance of the text. The system's dynamic generation of stego-text and the six options for maximum capacity provide users with flexibility and control over the amount of text used to hide the message.

Various methods can be used to encode information using whitespace steganography, but one of the most common involves using variations in the length of the spaces between words and paragraphs to represent binary data. For instance, a longer space between two words may represent a "0," while a shorter space represents a "1." Similarly, different variations in the length of spaces between paragraphs may be used to represent different characters or symbols.

Conclusion

The expected outcomes and advantage of this code is that it provides a simple and effective way to hide secret messages within a given text without raising suspicion. The use of inter-word and inter-paragraph spacing helps to hide the message in plain sight, making it difficult for anyone to detect. This could be useful for covert communication or for adding a bit of fun to a piece of text

References

[1] Delina B. Information hiding: A new approach in text steganography. InProceedings of the International Conference on Applied Computer and Applied Computational Science, World Scientific and Engineering Academy and Society (WSEAS 2008) 2008 (pp. 689-695).

[2] Por LY, Ang TF, Delina B. Whitesteg: a new scheme in information hiding using text steganography. WSEAS transactions on computers. 2008 May;7(6):735-45.

[3] Shirali-Shahreza MH, Shirali-Shahreza M. A new approach to Persian/Arabic text steganography. In5th IEEE/ACIS International Conference on Computer and Information Science and 1st IEEE/ACIS International Workshop on Component-Based Software Engineering, Software Architecture and Reuse (ICIS-COMSAR'06) 2006 Jul 10 (pp. 310-315). IEEE.

[4] Changder S, Debnath NC, Ghosh D. A new approach to Hindi text steganography by shifting matra. In2009 International Conference on Advances in Recent Technologies in Communication and Computing 2009 Oct 27 (pp. 199-202). IEEE.

[5] Shirali-Shahreza MH, Shirali-Shahreza M. A new synonym text steganography. In2008 international conference on intelligent information hiding and multimedia signal processing 2008 Aug 15 (pp. 1524-1526). IEEE.

[6] Tutuncu K, Abi Hassan A. New approach in E-mail based text steganography. International Journal of Intelligent Systems and Applications in Engineering. 2015 May 26;3(2):54-7.

[7] Nagarhalli TP. A new approach to SMS text steganography using emoticons. International journal of computer applications. 2014;975:8887.