Journal of Engineering Sciences

Vol 13, Issue 03, MARCH /2022

ISSN NO:0377-9254

COVERLESS INFORMATION HIDING METHOD BASED ON WEB TEXT

[1]MOUNIKA ASAMPELLY, [2]MURALI PONAGANTI

[1]MCA Student, [2]Assistant Professor

Department Of MCA

Sree Chaitanya College Of Engineering, Karimnagar

ABSTRACT

The coverless hiding of knowledge has become a hot topic because it can conceal confidential information (SI) without alteration from carriers. Targeted at the problems of poor hiding ability (HC) and text large data mismatch, a modern way of coverless knowledge that hides by scanning the vast volume of Internet text. Second, the approach suggested uses a browser-based spider technology in order to collect SI web texts to create a web-text library. Secondly, certain SI texts are scanned and the best web text is chosen from them. Subsequently, a 2-D coordinate scheme defines the role of the SI in the chosen web document. The site text URL is eventually paired with the position information gathered and then sent to the receiver. The experimental findings and review indicate that HC, the secret performance rate and protection are increased

I.INTRODUCTION

The concealed knowledge makes use of the senselessness of human meaning and multimedia redundancies to conceal in the optical carrier. It can primarily be separated into four categories: document hiding, picture hiding, visual hidting and audio hiding by the numerous digital carriers [1]. Text is the most commonly used media, so this paper concentrates on hiding text content.

Most strategies for hiding text information are categorised into three types: format-based text information hiding, picture-based text information hiding and the hiding of natural language information [2]. The first solution masks details by modifying terms lengths, adding intangible characters and formats in documents (PDF, Javascript, Office)[3] [5]. The second approach deals with text as a binary image, so SI can be obscured by integrating the versatility of binary images with the existence of texts [6], [7]. These current hiding techniques are vast

HC, however the secret carriers are changed, and then numerous attacks such as re-composing, OCR, and steg analysis cannot be resisted effectively[8] [10]. As long as the carrier is modified, the embedded SI would definitely be Detected. The third approach uses NLP to create SI-containing text carriers, such that details can be concealed without changed carrier[11].

However, the natural texts produced are often disadvantageous, such as incoherence and poor interpretability for semanthropic contexts, difficulties in observing language and grammar laws, and statistical differences. A new approach named text-free hiding has been introduced to radically overcome the above problems[12]. The key concept of this approach is to extract carriers carrying the SI directly from Text Big Data exchanged by the sender and the receiver, which conceal knowledge from current carriers and could be kept secret from SI without changing the carriers. Many researchers have been highly worried about the process because it can easily withstand different steganalysis approaches. Reference [13] suggested the Word Rank Map hiding approach by evaluating word frequency in each post.

On this basis[14] a hash algorithm has been implemented that incorporates word grade map and common terms. Reference [15] used character encoding to conceal details as a place tag. These methods have low HC, however, only one Chinese keyword in one document can be hidden. Consequently, some scholars advocated several keyword approaches for the problems. For instance [16] and [17] have suggested a multi-keyword scheme hiding algorithm. The core principle of the approaches is that the duration of the SI is concealed in the texts that comprise the SI. Reference [18] suggested a Steganalysis System focused on word-embedding. Reference [19] implemented a compound- and selection-based system. These techniques have increased HC to some degree, but with the duration of SI growing, the SR would decrease steadily. Reference[20] implemented a web-based text big data algorithm which simplifies text big data processing but allows for the fast discovery of position information due to direct encoding.

II.LITERATURE SURVEY

I. J. Cox and M. L. Miller, ``The _rst 50 years of electronic watermark-ing,''

It can be dated back to 1954 through electronic watermarking. Over the last 10 years, the automated watermarking has displayed significant interest, mainly because of fears regarding un authorised

copying of copyright material. The following issues are addressed in this paper: is the interest warranted? What are the technology's industrial applications? In the last 10 years, what empirical advance has been made? What are the most exciting test areas? And where should we take the next 10 years? We assume that the need in watermarking is sufficient. However, copyright applications, such as broadcast management, verification and monitoring material transmitted within organisations, are likely to become overshadowed. We often see numerous technologies that add value to media, such as annotations and site links. The most convincing of these above implementations may be. There have been tremendous advances in making these implementations possible—conceptual design, security risks and countermeasures, and the creation of a bag of tricks for effective deployment. Further development is expected in geometrical and time distortion control approaches. We foresee more exciting discoveries from knowledgeable watermarking studies.

J. T. Brassil, S. Low, and N. F. Maxemchuk, ``Copyright protection for the electronic distribution of text documents

Each copy of a text document may be almost invisibly distinguished by repositioning or changing the appearance of various text items, such as lines, phrases, or characters. A particular copy should be recorded with the receiver such that the initial owner can track future unauthorised copies. In this paper we identify and evaluate several document marking mechanisms and several other label decoding mechanisms following typical forms of distortion in documentation. The marks are structured to preserve records of little meaning which are kept by persons who, whether they can be differentiated, will rather have a legitimate rather than an unauthorised copy. We will identify attacks that erase the markers and countermeasures. An architecture is defined to distribute several copies, without putting any pressure on the publisher to produce and transfer unique materials. The design helps the publisher to recognise a receiver who unlawfully redistributed the document without jeopardising the identity of people who do not work illegally. It defines two experimental processes.

T.-Y. Liu and W.-H. Tsai, Robust Watermarking in Slides of Presentations by Blank Space Coloring

A modern, robust approach is suggested to imperceptibly incorporate a watermark picture into a presentation's slides. The Watermark is separated into blocks and inserted into a repetitive pseudo-random series of the space characters in the slides. The integration is accomplished by adjusting the colours of the spatial characters to new ones that are the product of coding block contents and indexes. The embedded watermark avoids a variety of typical slide modifications, including copying and pasting of slides; adding, removing and rearranging slides; modifying slide design; and migration into file formats. A safety key is used when integrating and removing a watermark to ensure that each watermark can be removed accurately with the relevant key using a weighted vote system, where an offending presentation includes slides taken by means of presentations labelled with separate safety keys. Microsoft PowerPoint tests confirm the viability of the suggested system. A identifiable watermark of 64 diameter may be derived on average from a presentation that includes five watermarked slides. The method proposed is useful for numerous applications of slide details, including copyright security, slide authentication, slide coverage, etc.

X. G. H. Luo, ``A steganalysis method based on the distribution of space characters

Steganalysis was a significant cybersecurity research topic that helps detect hidden threats on the public network. In the last two years, the exponential growth of natural language processing technologies has culminated in a great development of coverless steganography. Previous techniques of steganalysis have yielded unsatisfactory outcomes for this modern technique of steganography and appear to be a problem. Unlike all previous methods for steganalysis in this article, the text steganalysis(TS-CNN) approach is proposed based on semanticized study, which uses the CNN to extract high-level semantitic characteristics of text and considers the subtle variations in semanticized space before and after the embedding of secret knowledge. We have compiled and published a broad text steganalysis (CT-Steg) dataset comprising a total number of 216,000 texts of different lengths and different embedding frequencies to train and validate the proposed model. Experimental findings indicate that the proposed model reaches almost 100% accuracy and is reminiscent of all previous methods. In addition, the proposed model can also estimate the potential of secret knowledge inside. These findings clearly support the fact that it is feasible and efficient to use minor shifts in the semantical space before and after insertions of hidden knowledge to perform text steganalysis.

H. Kwon, Y. Kim, and S. Lee, ``A tool for the detection of hidden data in microsoft compound document _le format

Files which use Microsoft Compound Document File Format (MCDFF) for digital forensic researchers present a problem: it is easy to hide infonnations

from MCDFF, but it is challenging to detect the hidden data in them. Through an application downloaded from the website and the Win32 API, a suspect may conceal details that might be necessary to investigate in MCDFF. There was no method before our study for the discovery of data concealed in MCDFF, which rendered MCDFF analysis challenging for science. This paper presents an overview of the features of MCDFF which are used to mask details, and a method (the "DOCdetector"). The techniques of analysing data in unused areas and inserted streams have culminated in the creation of a DOC detection application to help find and evaluate secret data.

L. Goyal, M. Raman, P. Diwan, and M. K. Vijay, ``A robust method for integrity protection of digital data in text document watermarking

Digital watermarking offers verification, confirmation and security of copyright for online multimedia content. In addition to pictures, audio and video, text is the most commonly utilised medium of communication. It must also be secured. Document watermarking methods established in the past secure the text against unauthorised copying, replication and copyright infringements. We also suggested an algorithm in this paper which ensures that the document is integral and confidential. In this procedure, watermark is created on the basis of the document's content and inserted without altering the contents of the document and encrypted in order to guarantee secrecy of the text. To authenticate and show the document's credibility, the watermark can easily be removed and tampered.

X. Chen, H. Sun, Y. Tobe, Z. Zhou, and X. Sun, ``Coverless information hiding method based on the Chinese mathematical expression

In the field of computer protection, coverless information hiding has become a hot topic as a modern information hiding tool. In each natural document, the current coverless hiding system may conceal only one Chinese character. The issue with the system, however, is that the hiding potential is too limited. To deal with this issue, this paper proposes a new approach called a coverless multi keyword knowledge hiding method based on text. The crucial point is that the document includes all the keywords and their numbers. Experimental findings demonstrate that the proposed approach will boost the ability of the current coverless text-based knowledge hiding system.

J. Zhang, L. Wang, J. Shen, and H. Lin, "Text hiding system dependent on word rank map text

The hiding of textual knowledge has drawn the interest of many scientists and accomplished a great deal. However, there is a big weakness in text knowledge hiding methods like text type, process creation, text picture and so on that cannot endure steganography detection. A new approach for covering coverless text details is introduced based on the rating map of terms. First of all, stego-vectors are generated by the rank map directly from the secret message. Then, some standard texts, including the sgeto-vectors produced, will be searched for text big data. Finally, the confidential knowledge may be transmitted without modifying the stego-texts to the recipient. The algorithm proposed has a greater theoretical and functional relevance, as it is robust for virtually any existing form of steganalysis.

III.SYSTEM ANALYSIS AND DESIGN

EXISTING SYSTEM:

Under the current framework, coverless information hiding has become a hot problem in the field of information protection as a modern information hiding tool. In each natural document, the current coverless hiding system may conceal only one Chinese character. The issue with the system, however, is that the hiding potential is too limited. There is no new method called multi-keyword coverless knowledge hiding method, based on text, suggested in previous articles, to resolve this problem. The crucial point is that the document includes all the keywords and their numbers. Experimental findings demonstrate that the proposed approach will boost the ability of the current coverless text-based knowledge hiding system.

PROPOSED SYSTEM:

· The scheme proposes a coverless web-based knowledge hiding method which uses the large number of web pages on the Internet to disguise the SI. Firstly the keywords and hidden keywords sought and, with mature search engines, the web pages affiliated with SI are retrieved. Web texts are then fetched on the Internet by utilising Web spider technologies to create a web text library.

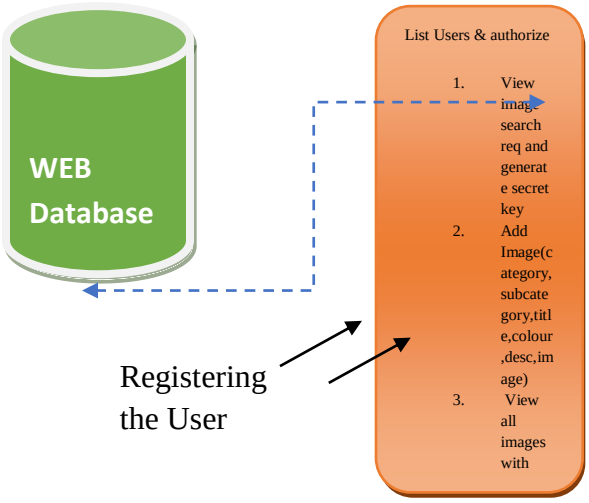
A sequence of web text sets of hidden keywords are contained in the library. The best web text is chosen as stegotext from web text collections. The place of a hidden keyword is defined as a co-ordinate type in the site document. After the position information is coupled with the key, the URL parameters are compressed and packed. The URL is sent as a transmission tool to the receiver. There is no change trace of the proposed system in the site text and it has strong HC and a secure SR.

Journal of Engineering Sciences

Vol 13, Issue 03, MARCH /2022

ISSN NO:0377-9254

IV.Architecture Diagram



V.MODULES IMPLEMENTATION

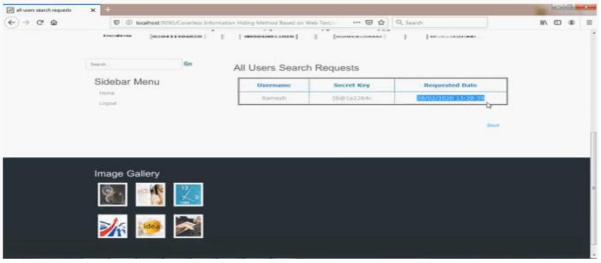
Admin

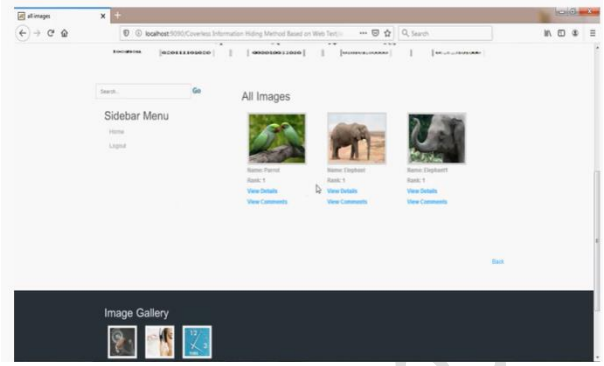
Admin requires a correct username and password in this module. After logging successfully, he or she is able to do multiple operations such as display all users, permit and their information, view the user select request and create a hidden key, add Images and its details such as(category, sub-category, picture name, colour, c and image) and view all images in rank.

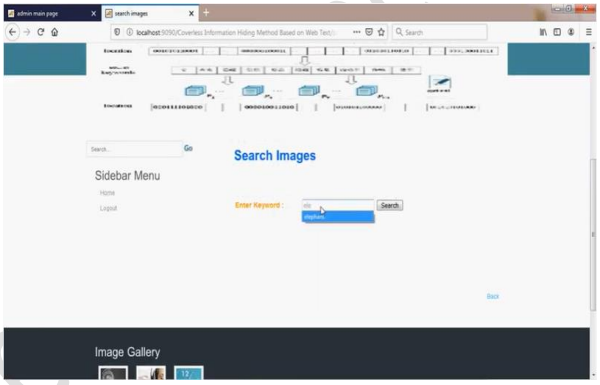
- **User**

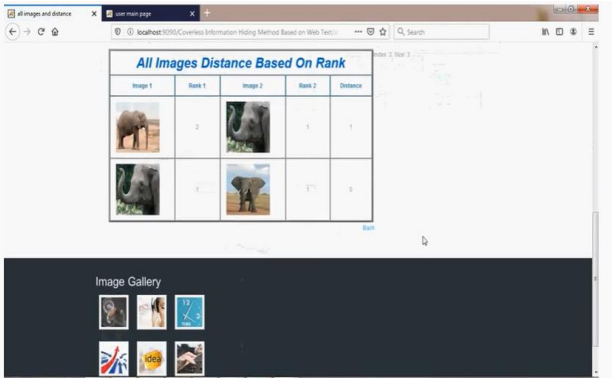
Numbers of users are present in this module. Until executing such procedures, the consumer may log. After active registration, you can log in with correct username and password. Submit a secret key search request to search pictures and view a secret key answer, search pictures by entering a secret key if the search page otherwise opens shows the error message, see all the image search details (keyword, search process and date on a search), and view hidden site text, and view top-ranking photos by supplying.

VI.SCREEN SHOTS









VII.CONCLUSIONS

This thesis proposed a new method of web text hiding coverless information. The approach considers the current vast Internet text as large data and uses spider technologies to create a web text library. And hidden web texts are obtained to mask knowledge. The system not only ensured the dissimulation of knowledge but also increased the dissimulation ability. However, confidentiality is not sufficient because a substantial portion of the website includes confidential details. The next research is how the retrieval mechanism is managed to overcome this issue.

REFERENCES

[1] I. J. Cox and M. L. Miller, ``The _rst 50 years of electronic watermark-ing," *EURASIP J. Adv. Signal Process.*, vol. 2002, no. 2, Dec. 2002,Art. no. 820936.

[2] K. Bennett, ``Linguistic steganography: Survey, analysis, and robustnessconcerns for hiding information in text," Dept. Linguistic, Purdue Univ.,West Lafayette, IN, USA, CERIAS Tech. Rep. 13, 2004.

[3] J. T. Brassil, S. Low, and N. F. Maxemchuk, ``Copyright protection forthe electronic distribution of text documents," *Proc. IEEE*, vol. 87, no. 7,pp. 1181_1196, Jul. 1999.

[4] T.-Y. Liu and W.-H. Tsai, *Robust Watermarking in Slides of Presentationsby Blank Space Coloring: A New Approach* (Lecture Notes in ComputerScience), vol. 5510. Berlin, Germany: Springer-Verlag, 2009, pp. 49_64.

[5] A. Koluguri, S. Gouse, and P. B. Reddy, ``Text steganography methods andits tools," *Int. J. Adv. Sci. Tech. Res.*, vol. 2, no. 4, pp. 888_902, 2014.

[6] M. Wu and B. Liu, ``Data hiding in binary image for authenticationand annotation," *IEEE Trans. Multimedia*, vol. 6, no. 4, pp. 528_538,Aug. 2004.

[7] Z. H. Xia, S. H. Wang, X. M. Sun, and J. Wang, ``Print-scan resilientwatermarking for the chinese text image," *Int. J. Grid Distrib. Comput.*,vol. 6, no. 6, pp. 51_62, 2013.

[8] X. G. H. Luo, ``A steganalysis method based on the distribution of spacecharacters," in *Proc. IEEE Int. Conf. Commun., Circuits Syst.*, vol. 1,Jun. 2006, pp. 54_56.

[9] H. Kwon, Y. Kim, and S. Lee, ``A tool for the detection of hidden data inmicrosoft compound document _le format," in *Proc. Int. Conf. Inf. Sci.Secur. (ICISS)*, Jan. 2008, pp. 141_146.

[10] L. Goyal, M. Raman, P. Diwan, and M. K. Vijay, ``A robust method forintegrity protection of digital data in text document watermarking," *Int.J.Innov. Res. Sci. Technol.*, vol. 1, no. 6, pp. 14_18, Nov. 2014.