

Coverless Information Hiding Method Based On Web Text

Manikanta Deva , Sri.V.Bhaskara Murthy

Mca Student, Associate Professor

Dept Of Mca

B.V.Raju College, Bhimavaram

ABSTRACT

Coverless information hiding has become a hot topic because it can hide secret information (SI) into carriers without any modification. Aiming at the problems of the low hiding capacity (HC) and mismatch in text big data, a novel method of coverless information hiding by retrieving the massive amount of web text on the Internet. First, the proposed method uses a web spider technology to capture web texts associated with SI to construct a web-text library. Second, some texts containing SI are searched and the optimal web text is selected from them. Then, the location of the SI in the selected web text is described by using a 2-D coordinate system. Finally, the URL of the web text is combined with the obtained location information and then sent to the recipient. The experimental results and analysis show that the performances are improved in terms of HC, hiding success rate, and security.

1.INTRODUCTION

Information hiding uses the insensitivities of the human sense and the redundancies of the multimedia to realize hiding in the digital carrier. It can be mainly divided into four categories: text information hiding, image information hiding, video information hiding and audio information hiding according to the different digital carriers [1]. Text is the most widely used media, so this paper focuses on text information hiding.

Most text information hiding methods are mainly divided into three types: format-based text information hiding, image-based text information hiding, and natural language information hiding [2]. The first method hides information by changing the distances between words, inserting invisible characters and modifying the formats in documents (PDF, HTML, Office) [3]_[5]. The second method treats text as binary image, thus it is possible to hide SI by combining the features of binary images with the natures of texts [6], [7]. These existing hiding methods have large HC, but the hidden carriers are modified, thus they can- not effectively resist various attacks such as re-composing, OCR and steganalysis [8]_[10]. Theoretically, as long a the carrier is modified, the embedded SI will certainly be detected. The third method uses natural language processing technology (NLP) to generate text carriers containing SI, so this method can hide information without modifying the carrier [11].

However, the generated natural texts always have some disadvantages, for example, incoherence and poor readability of semantic context, difficulties to follow linguistics and grammar rules, and deviations on statistics. A novel method called text coverless information hiding was proposed for fundamentally solve the issues mentioned above [12]. The main idea of this method is to directly retrieve carriers containing the SI from text big data shared between the sender and the recipient, which employs existing carriers to hide information and SI could be hidden without any modification

to the carriers. The method has been widely concerned by a lot of researchers because it can resist various steganalysis methods effectively. Reference [13] proposed a hiding method based on Word Rank Map by analyzing the frequency of words in each article.

On this foundation, [14] introduced a hash algorithm by combining word rank map and frequent words. Reference [15] used character encoding as a location tag to hide information. However, these methods have low HC, only one Chinese keyword can be hidden in one text. Therefore, some researchers proposed several multi-keywords methods for the issues. For example, [16] and [17] proposed a hiding algorithm based on multi-keywords schema. The main idea of the methods is that the length of SI will be hidden in the texts where SI has been hidden. Reference [18] proposed a word-embedding-based Steganalysis Method. Reference [19] introduced a method based on compound and selection. These methods improved HC to some extent, but the SR will be gradually decrease as the length of the SI increases. Reference [20] introduced an algorithm based on web text big data, which simplifies the processing of text big data, but the location information is easy to be discovered due to the direct encoding.

II.EXISTING SYSTEM

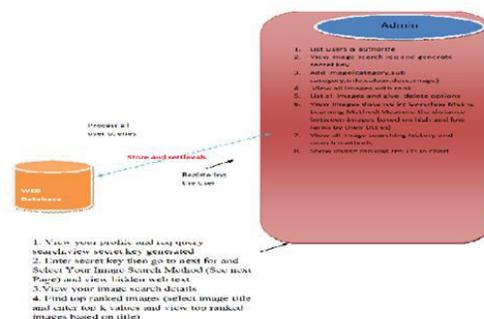
In the existing system, as a new information hiding method, coverless information hiding has become a hot issue in the field of information security. The existing coverless information hiding method can hide only one Chinese character in each natural text. However, the problem of the method is that the hiding capacity is too small. To address this problem, there is no new method named coverless multi-keywords information hiding method based on text is proposed in the previous papers. The

main idea of the method is that both the keywords and their number will be hidden in the texts. Experimental results show that the proposed method can improve the capacity of the existing coverless information hiding method based on text.

III.PROPOSED SYSTEM

The system proposes a coverless information hiding method based on web text, which takes advantage of the large number of web pages on the Internet to hide SI. The searched keywords and secret keywords are extracted firstly, and the web pages associated with SI are obtained by using mature search engines. Then, web texts on the Internet are fetched by using web spider technology to construct a web text library. A series of web text sets containing secret keywords are retrieved in the library. The optimal web text is selected from web text sets as stegotext. The location of each secret keyword in the web text is described as the form of the coordinates. After the location information is combined with the key, it is compressed and packaged into the parameters of the URL. The URL is sent to the recipient as a transmission medium. The proposed method does not have any modification trace in the web text, and it has high HC and stable SR.

IV.SYSTEM ARCHITECTURE



V. IMPLEMENTATION

Admin

In this module, admin has login by valid user name and password. After login successful he can do some operations such as view all user, authorize and their details, view users search request and generate secret key, Add Images and its details like(category, sub category, image name,color,desc and image), view all images with rank ,List all images perform operations like (edit or delete) , view all images distance based on rank high to low , view all images search history and search method ,view all images ranking results in chart.

User

In this module, there are n numbers of users are present. User should register before doing some operations. After registration successful he can login by using valid user name and password. Login successful he will do some operations like view profile details, Send secret key request for searching images and view secret key response, Search images by entering secret key if it match search page will be opened otherwise show error message, view all images search details like (keyword, search method and date on searched) and view hidden web text and view top ranked images by providing top 'k' value.

VI.CONCLUSIONS

This paper proposed a novel method of coverless information hiding based on Web Text. The method regards existing massive internet text as big data, and constructs web text library by using spider technology. And web texts that contain secret information are retrieved to hide information. The method not only ensured the concealment of information, but also improved the hiding capacity. However, the security is not satisfying when secret information contains a large part of a web page. The next research is

how to control the retrieval process to solve this problem.

REFERENCES

- [1] I. J. Cox and M. L. Miller, "The first 50 years of electronic watermarking," *EURASIP J. Adv. Signal Process.*, vol. 2002, no. 2, Dec. 2002, Art. no. 820936.
- [2] K. Bennett, "Linguistic steganography: Survey, analysis, and robustness concerns 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 for the 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 Presentations by Blank Space Coloring: A New Approach (Lecture Notes in Computer Science)*, vol. 5510. Berlin, Germany: Springer-Verlag, 2009, pp. 49_64.
- [5] A. Koluguri, S. Gouse, and P. B. Reddy, "Text steganography methods and its 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 authentication and 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 resilient watermarking 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 space

- characters," 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 in microsoft 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 for integrity protection of digital data in text document watermarking," Int. J. Innov. Res. Sci. Technol., vol. 1, no. 6, pp. 14_18, Nov. 2014.
- [11] P. Wayner, *Disappearing Cryptography: Information Hiding: Steganography Watermarking*, 2nd ed. San Francisco, CA, USA: Morgan Kaufmann, 2009.
- [12] X. Chen, H. Sun, Y. Tobe, Z. Zhou, and X. Sun, "Coverless information hiding method based on the Chinese mathematical expression," in Proc. 1st Int. Conf. Cloud Comput. Secur., Nanjing, China, 2015, pp. 133_143.
- [13] J. Zhang, J. Shen, L. Wang, and H. Lin, "Coverless text information hiding method based on the word rank map," in Proc. Int. Conf. Cloud Comput. Secur., Nanjing, China, Nov. 2016, pp. 145_155.
- [14] J. Zhang, Y. Xie, L. Wang, and H. Lin, "Coverless text information hiding method using the frequent words distance," in Proc. 3rd Int. Conf. Cloud Comput. Secur., Oct. 2017, pp. 121_132.
- [15] X. Chen, S. Chen, and Y. Wu, "Coverless information hiding method based on the Chinese character encoding," J. Internet Technol., vol. 18, no. 2, pp. 313_320, Mar. 2017.