

Text Hiding Scheme Using Big-5 Code

Jun-Chou Chuang¹, Chun-Chi Lo², Yu-Chen Hu³, Piyu-Yu Tsai⁴

¹ Department of Computer Science and Communication Engineering

² Department of Computer Science and Information Engineering

³ Department of Computer Science and Information Management

Providence University

lzchung@pu.edu.tw

⁴ Department of Computer Science and Information Engineering

National United University

pytsai@nuu.edu.tw

Abstract. This paper presents a text hiding scheme using Big-5 code. Some text-hiding schemes embed secret information at between-word and between-character by adding tabs or spaces. Liu et al. proposed a Chinese text data hiding scheme to divide the Chinese character into left and right parts for data embedding. However, the adjusted spaces or divided characters of between-word may look like strange and it may expose the risk of security. Therefore, we intend to design a text hiding scheme using Big-5 code. The secret is hidden into spaces between-word and between-character of a cover text by placing a Big-5 code either 20 or 7F. The visual quality of the stego-document is the same as the original document and therefore reducing the suspicious of attention by hackers. Experimental results show that the visual quality of the proposed scheme achieves good results and feasibility.

Keywords: Text hiding, data hiding, image hiding, Big-5.

1 Introduction

Text hiding schemes [1-6] can be classified into two types, content format and language semantic. The content format methods adjust the width of tracking, the height of leading, number of white spaces, font sizes, and etc. The language semantic methods change the meaning of a phrase or a sentence in a text document. The traditional text hiding schemes embed secret information at between-word and between-character by adding tabs or white spaces. However, the adjusted white spaces of between-word may look like strange. Therefore, we intend to design a text hiding scheme using Big-5 code. The secret is first converted into binary and then embedded into white spaces between-word and between-character of a cover text by placing a Big-5 code either 20 or 7F.

2 The Proposed Scheme

In our Scheme, the secret is hidden into spaces between-word and between-character of a cover text by placing a Big-5 code either 20 or 7F. The advantage of our scheme gets good visual quality of the stego document and it does not cause any visual distortion. The hexadecimal number of a white-space in the Big-5 table is 20. We can see that the Big-5 code 7F is a blank character and it can simulate as the white-space function. That is, we use the two Big-5 codes 20 and 7F to embed secret bit 0 or 1, respectively.

To embed secret into a cover text, we shall adjust the content of a cover text. We need to add a white-space in each between-word and between-character. Secret messages are sequentially converted into 0's and 1's binary stream. One white-space of between-word and between-character in a cover text is used to hide one secret bit. If we want to embed secret bit 0, the Big-5 code of white-space 20 is adopt. If we want to embed secret bit 1, the Big-5 code of blank character 7F is adopt. After finishing the secret embedding, we add an end-of-code 7F to indicate no secret of input. The hiding capacity of a cover text can be determined before data embedding. We can calculate the total number of white-spaces in a cover text. Assume a cover text contains w characters; the embedding payload of a cover text is $(w-1)$ bits.

In the extraction phase, the decoder sequentially scans the stego-text to find white-space of between-character and between-word. The extracted secret bit is "0" or "1" depends on the extracted Big-5 code is 20 or 7F. The length of the secret can be determined by scanning the location of the last 7F code in the stego text.

3 Experimental Results and Discussions

We have made two experiments on Chinese cover text and English cover text. The secret bits are randomly generated by the pseudo random number generator with a known seed. The embedding payload of the two cover texts in Fig. 1 are 243 bits and 141 bits, respectively. The evaluated criterion on the proposed method concerns on the embedding payload of the cover text. To demonstrate the hiding capacity, we compare the hiding capacity of our scheme, Liu's scheme and Wang's scheme in Table 1. The Liu's method and the Wang's method cause visual distortion of Chinese words and do not have enough embedding payload. From the experiment results in Table 1, the embedding payload of our scheme is better than Liu's method and Wang's method. Bessie, our scheme can apply to any kind of text format. On the other hand, our scheme does not adjust the width of white-space. Therefore, the visual quality of the proposed scheme achieves good results.

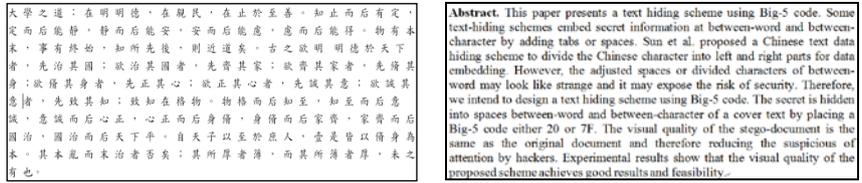


Fig. 1. Two cover texts.

Table 1. Compare the embedding payload of Chinese cover text to our scheme, Liu’s method, and Wang’s method.

Method	Payload (bits)
Liu et al. [5]	~63
Wang et al. [6]	~73
Our method	243

4 Conclusions

The paper presents a text hiding technique using Big-5 code. The proposed method applies two Big-5 codes 20 and 7F to encode the secret bit. The Big-5 code 20 is to represent secret bit “0” and the Big-5 code 7F is used to represent the secret bit “1”. The secret bit is hidden into the white-space of between-character and between-word in a cover text. Experimental results show that the visual quality of the proposed scheme achieves good results and feasibility.

References

1. Chang, C. C., Wu, C. C. and Lin, I. C.: A Data Hiding Method for Text Documents Using Multiple-Base Encoding. High Performance Networking, Computing, Communication Systems, and Mathematical Foundations, vol. 66, pp. 101--109 (2010)
2. Liu, Y., Sun, X., Gan, C. and Wang, H.: An Efficient Linguistic Steganography for Chinese Text. IEEE International Conference on Multimedia and Expo, pp. 2094--2097, Beijing, China (2007)
3. Liu, T.Y. and Tsai, W.H.: A New Steganographic Method for Data Hiding in Microsoft Word Documents by a Change Tracking Technique. IEEE Transactions on Information Forensics and Security, vol. 2, no. 1, pp. 24--30 (2007)
4. Lee, I.S. and Tsai, W.H.: A New Approach to Covert Communication via PDF Files. Signal Processing, vol. 90, no. 2, pp. 557--565 (2010)
5. Liu, Y., Sun, X., Gan, C. and Wang, H.: An Efficient Linguistic Steganography for Chinese Text. IEEE International Conference on Multimedia and Expo, pp. 2094--2097, Beijing, China (2007)
6. Wang, Z.H., Chang, C.C., Lin, C.C. and Li, M.C.: A Reversible Information Hiding Scheme Using Left-Right and Up-Down Chinese Character Representation. The Journal of Systems and Software, vol. 82, no. 8, pp. 1362--1369 (2009)