Data Hiding on Text Using Big-5 Code

Jun-Chou Chuang[1] and Yu-Chen Hu[2]

[1]*Department of Computer Science and Communication Engineering*
Providence University
200 Chung-Chi Rd., Shalu, Taichung 43301, Republic of China (Taiwan)
[2] *Department of Computer Science and Information Management*
Providence University
200 Chung-Chi Rd., Shalu, Taichung 43301, Republic of China (Taiwan)
lzchung@pu.edu.tw, ychu@pu.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

With the advance of computer and Internet technologies, protecting personal information becomes an important issue. Traditional Cryptography for symmetric encryption schemes and non-symmetric encryption schemes can provide high-level data security, but they are not flexibility for all kinds of media. Steganography also called data-hiding [1-7] is the covert communication which embeds secret information into a meaningful media with imperceptibly and only the authorized users can extract the hidden data. The meaningful media used to hidden secret message is called the cover media, and the encoding result is called the stego media. Generally speaking, a data hiding scheme should satisfy the followings two issues. Firstly, the hiding capacity should be as large as possible. Secondly, the visual quality of the embedding result should not distinguishable from the cover media.

Digital image is the most popular camouflage media and the main reason is image pixels can be distortion. Images are very easy to make imperceptible modification. When an image pixel is tiny modified, it is not easily awarded something difference by the human eyes. Recently, some researches concerning the data hiding scheme on the text documents. The embedding payload of a text document is less than a digital image because it is not easy to find the redundancy information in a text document. In generally, text hiding schemes [8-15] 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. The rest of this paper is organized as follows. We give a briefly review of the Liu's text hiding method and Wang's text hiding method in Section 2. The proposed scheme is presented in Section 3. The experimental results are described in Section 4. Finally, conclusions will be given in Section 5.

2. Related Works

In 2004, Liu et al. [10] proposed an L-R scheme for Chinese text hiding scheme. The scheme divides a Chinese character into left and right parts. The embedding procedure contains two steps. The first step is to embed secret bits into the selected Chinese characters. A Chinese character can be divided into left and right parts would be selected to hide data. If they want to embed bit 0, a Chinese character must be divided into left and right parts. If they want to embed bit 1, a Chinese character does not need to divide into left and right parts. The second step is to adjust the width of the inter-word space in order to keep acceptable visual quality. In Figure 1, a bit string, 01 is embedded into two Chinese words "很" and "好". The secret bit 0 means we do not need to divide a Chinese character into left and right parts. In the data extracting procedure, the secret can be extracted from the stego text by the human eyes directly.

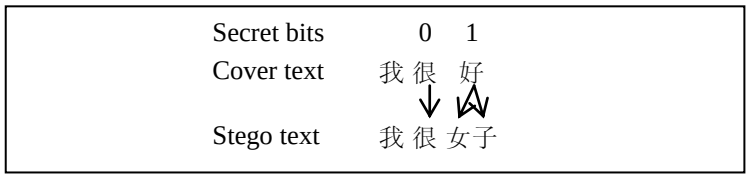

Figure 1. An Example of Liu's Text Hiding Scheme

In 2009, Wang et al. [14] applied the concepts of Liu et al.'s scheme to propose an improvement version. The method combines the up and down components (U-D scheme) of Chinese characters rather than L-R scheme only in order to increase the hiding capacity. An example of Wang's scheme is listed in Figure 2. In Figure 2, a bit string, 011 is embedded into three Chinese words "位", "明" and "星".

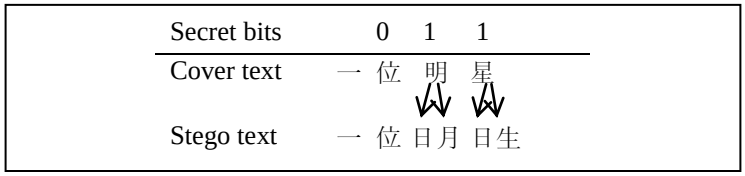

Figure 2. An Example of Wang's Text Hiding Scheme

3. The Proposed Scheme

Wang et al. applies the L-R scheme and U-D scheme partition a Chinese word. Their scheme has two drawbacks. First, an embeddable Chinese word should satisfy L-R or U-D partition. Second, the visual quality of the embedding result still leaves a lot of places can be improved. For that, we proposed an effective 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 advantage of our scheme gets good visual quality of the stego document and it does not cause any visual distortion. A portion of the Big-5 table is listed in Table 1. 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. The block diagram of the proposed scheme is shown in Figure 3.

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 a secret bit 0, the Big-5 code of white-space 20 is applied. If we want to embed a secret bit 1, the Big-5 code of blank character 7F is applied. 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.

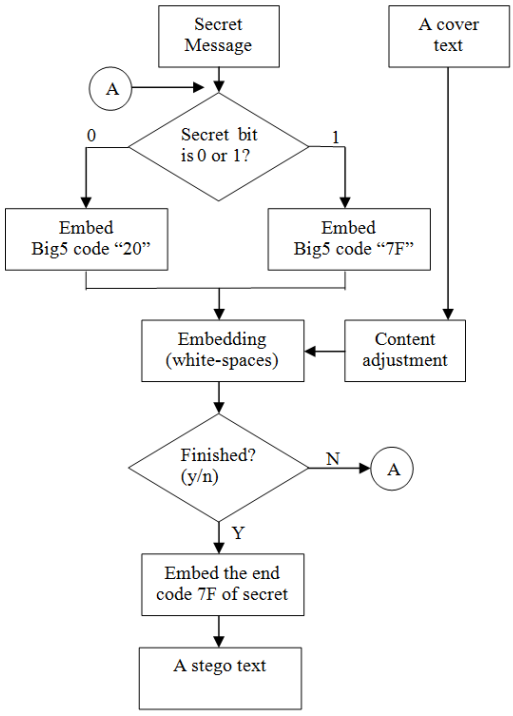

Figure 3. The Block Diagram of the Proposed Scheme

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.

Table 1. Two Big-5 Codes 20 and 7F in the Big-5 Table

1	space	20	17	0	30	33	@	40	49	P	50	65	`	60	81	p	70
2	!	21	18	1	31	34	A	41	50	Q	51	66	a	61	82	q	71
3	"	22	19	2	32	35	B	42	51	R	52	67	b	62	83	r	72
4	#	23	20	3	33	36	C	43	52	S	53	68	c	63	84	s	73
5	$	24	21	4	34	37	D	44	53	T	54	69	d	64	85	t	74
6	%	25	22	5	35	38	E	45	54	U	55	70	e	65	86	u	75
7	&	26	23	6	36	39	F	46	55	V	56	71	f	66	87	v	76
8	'	27	24	7	37	40	G	47	56	W	57	72	g	67	88	w	77
9	(	28	25	8	38	41	H	48	57	X	58	73	h	68	89	x	78
10	)	29	26	9	39	42	I	49	58	Y	59	74	i	69	90	y	79
11	*	2A	27	:	3A	43	J	4A	59	Z	5A	75	j	6A	91	z	7A
12	+	2B	28	;	3B	44	K	4B	60	[	5B	76	k	6B	92	{	7B
13	,	2C	29	<	3C	45	L	4C	61	\	5C	77	l	6C	93	|	7C
14	-	2D	30	=	3D	46	M	4D	62	]	5D	78	m	6D	94	}	7D
15	.	2E	31	>	3E	47	N	4E	63	^	5E	79	n	6E	95	~	7E
16	/	2F	32	?	3F	48	O	4F	64	_	5F	80	o	6F	96		7F

4. 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 cover text in Figures 4 and 5 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 2. 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 2, the embedding payload of our scheme is better than Liu's method and Wang's method. Besides, 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.

大學之道：在明明德，在親民，在止於至善。知止而后有定，
定而后能靜，靜而后能安，安而后能慮，慮而后能得。物有本
末，事有終始，知所先後，則近道矣。古之欲明 明德於天下
者，先治其國；欲治其國者，先齊其家；欲齊其家者，先脩其
身；欲脩其身者，先正其心；欲正其心者，先誠其意；欲誠其
意者，先致其知；致知在格物。物格而后知至，知至而后意
誠，意誠而后心正，心正而后身脩，身脩而后家齊，家齊而后
國治，國治而后天下平。自天子以至於庶人，壹是皆以脩身為
本。其本亂而末治者否矣；其所厚者薄，而其所薄者厚，未之
有也。

Figure 4. Chinese Cover Text

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. Sun 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.

Figure 5. English Cover Text

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. [10]	~63
Wang et al. [14]	~73
Our method	243

5. Conclusion

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.

Acknowledgements

This work is supported by the National Science Council, Taiwan, R.O.C. under contract NSC 100-2221-E-126-008. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

References

[1] J. C. Chuang and Y. C. Hu, "An Adaptive Image Authentication Scheme for Vector Quantization Compressed Image", Journal of Visual Communication and Image Representation, 22, 5 **(2011)**.
[2] D. Kundur and D. Hatzinakos, "Digital watermarking for Telltale Tamper Proofing and Authentication", Proceedings of the IEEE, 87, 7 **(1999)**.
[3] C. S. Lu and H.Y.M. Laio, "Multipurpose Watermarking for Image Authentication and Protection", IEEE Transactions on Image Processing, 10, 10 **(2001)**.
[4] J. C. Chuang, C. C. Chang and Y. C. Hu, "A Novel Color Image Hiding Scheme Using Error Diffusion Technique", Imaging Science Journal, 56, 1 **(2008)**.
[5] C. C. Chang, P. Y. Lin and J. C. Chuang, "A Grayscale Image Steganography Based upon Discrete Cosine Transformation", Journal of Digital Information Management, 8, 2 **(2010)**.
[6] J. C. Chuang, and C. C. Chang, "Using a Simple and Fast Image Compression Algorithm to Hide Secret Information", International Journal of Computers and Applications, 28 **(2006)**.
[7] C. C. Chang, J. C. Chuang and Y. C. Hu, "Spatial Domain Image Hiding Scheme Using Pixel-Values Differencing, Fundamental Informaticae, 3, 70 **(2006)**.

[8] C. Chen, S. Z. Wang and X. P. Zhang, "Information Hiding in Text Using Typesetting Tools with Stego-Encoding", International Conference on Innovative Computing, Information and Control, (**2006)**, 1, 459-462, Beijing, China.
[9] C. C. Chang, C. C. Wu and I. C. Lin, "A Data Hiding Method for Text Documents Using Multiple-Base Encoding, High Performance Networking", Computing, Communication Systems, and Mathematical Foundations, 66, **(2010)**.
[10] Y. Liu, X. Sun, C. Gan and H. Wang, "An Efficient Linguistic Steganography for Chinese Text", IEEE International Conference on Multimedia and Expo, **(2007)**, 2094-2097, Beijing, China.
[11] T. Y. Liu and W.H. Tsai, "A New Steganographic Method for Data Hiding in Microsoft Word Documents by a Change Tracking Technique", IEEE Transactions on Information Forensics and Security, 2, 1 **(2007)**.
[12] I. S. Lee and W. H. Tsai, "A New Approach to Covert Communication via PDF Files", Signal Processing, 90, 2 **(2010)**.
[13] I. S. Lee and W. H. Tsai, "Security Protection of Software Programs by Information Sharing and Authentication Techniques Using Invisible ASCII Control codes", International Journal of Network Security, 10, 1 **(2010)**.
[14] Z. H. Wang, C. C. Chang, C. C. Lin and M.C. Li, "A Reversible Information Hiding Scheme Using Left-Right and Up-Down Chinese Character Representation", The Journal of Systems and Software, 82, 8 **(2009)**.
[15] A. Ibrahim and A. Zabian, "Algorithm for Text Hiding in Digital Image for Information Security", International Journal of Computer Science and Network Security, 9, 6 (**2009)**.

Authors

Jun-Chou Chuang received his PhD. degree in computer science and information engineering from the Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi, Taiwan in 2004. Currently, Dr. Chuang is an assistant professor in the Department of Computer Science and Communication Engineering, Providence University, Sha-Lu, Taiwan. His research interests include multimedia security, data hiding, digital watermarking and signal processing.

Yu-Chen Hu received his PhD. degree in computer science and information engineering from the Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi, Taiwan in 1999. Currently, Dr. Hu is a professor in the Department of Computer Science and Information Management, Providence University, Sha-Lu, Taiwan. Dr Hu Servers as the Editor-in-Chief of International Journal of Image Processing since 2009. In addition, he is the managing editor of Journal of Information Assurance & Security. He joints the editorial boards of several other journals. His research interests include image and signal processing, data compression, information hiding, and data engineering.