A novel, reversible, Chinese text information hiding scheme based on lookalike traditional and simplified Chinese characters

Bin Feng[1], Zhi-Hui Wang[1], Duo Wang[1], Ching-Yun Chang[2] and Ming-Chu Li[1]
[1] School of software, Dalian University of Technology
Dalian, China 116620
[e-mail: wangzhihui1017@gmail.com]
[2] Computer Laboratory, University of Cambridge
Cambridge, UK
[e-mail:Ching-Yun.Chang@cl.cam.ac.uk]
*Corresponding author: Zhi-Hui Wang

Received May 15, 2012; revised August 3, 2012; accepted December 23, 2012; published January 29, 2014

Abstract

Compared to hiding information into digital image, hiding information into digital text file requires less storage space and smaller bandwidth for data transmission, and it has obvious universality and extensiveness. However, text files have low redundancy, so it is more difficult to hide information in text files. To overcome this difficulty, Wang et al. proposed a reversible information hiding scheme using left-right and up-down representations of Chinese characters, but, when the scheme is implemented, it does not provide good visual steganographic effectiveness, and the embedding and extracting processes are too complicated to be done with reasonable effort and cost. We observed that a lot of traditional and simplified Chinese characters look somewhat the same (also called lookalike), so we utilize this feature to propose a novel information hiding scheme for hiding secret data in lookalike Chinese characters. Comparing to Wang et al.'s scheme, the proposed scheme simplifies the embedding and extracting procedures significantly and improves the effectiveness of visual steganographic images. The experimental results demonstrated the advantages of our proposed scheme.

Keywords: Text data hiding, information hiding, Chinese characters, steganography

http://dx.doi.org/10.3837/tiis.2014.01.016

1. Introduction

Steganography[1, 14] (also called information-hiding) is the art and science of writing hidden messages in such a way that no one, apart from the sender and intended recipient, suspects the existence of the message; thus, it is a form of security through obscurity. In this kind of technique, secret information is embedded or hidden into some cover objects. There are different types of cover objects, including text [1, 2, 3, 4], audio, images [5, 6, 7, 8], and video.

With the rapid development of Internet, information-hiding techniques have become a topic of intense interest to those in the domain of information security. Massive amounts of text data are transferred or exchanged on the Internet every day, which makes it a very important to be able to use text as cover objects. Therefore, our focus in this paper is on the use of text as cover objects. Methods for hiding information in to text files usually modify the "cover text" to hide information, which makes the modified text file into what is known as "stego-text." For example, text size, spacing, font, or other characteristics of the text can be modified to contain the hidden information. However, compared to image or audio, text based information-hiding is considered to be the most difficult type due to the lack of redundancy in the text. To overcome this difficulty, several methods for hiding information into text have been proposed in recent years[3, 10, 16, 18, 19].

In 1998, Chapman proposed a method to disguise sensitive data as normal communications to thwart the censorship of sensitive data. He developed a software system called NICETEXT for transforming sensitive data into “harmless looking” natural language text consisting of sentences in English [2]. It also invented a system called SCRAMBLE for recovering the input, which is the original sensitive data, to NICETEXT. The success of NICETEXT system relies on large code dictionaries consisting of words categorized by type. Maher [9] implemented another text data-hiding system called TEXTO, which was designed to transform ASCII data into English sentences. In 2003, Wu and Huang [16] proposed a Markov-chain-based hiding strategy that regards the generation of text as a signal transmission from a Markov signal source. The secret messages can be embedded into the text according to a state transfer chart, which can be build in the Markov signal source from sample text. In 2004, Sun et al. [13] proposed a scheme using the left and right components of Chinese characters. Their scheme chooses those characters that have left and right components as candidates in which to hide secrets. Wang et al.[14] introduced a reversible, information-hiding scheme using left-right and up-down Chinese character set (L-R and U-D scheme) in 2009. The L-R and U-D scheme [14] includes a reversibility function and improves the hiding capacity of previous data-hiding methods [2, 3, 10, 16, 17]. However, the L-R and U-D scheme has some drawbacks that can be improved upon. The first one is that its visual steganographic effectiveness is not very good. The reassembling of the separate components of the Chinese characters causes deformation of the font, which makes it easy for attackers to visually distinguish the difference between the stego-text and the original text. The second drawback is that the scheme’s embedding and extracting procedure requires that the distances of the components of the characters be adjusted in Microsoft Word@, which is a complex procedure.

Since about 20% of the people in the world use Chinese characters and most of these people use Microsoft Word@ as the file tool, it is worthwhile to focus on the use of Chinese text files for information hiding via Microsoft Word@. In order to solve the above problems associated with Wang et al.’s scheme and to further improve its performance, we propose a novel scheme that uses look-alike Chinese characters in the simplified pattern and in the traditional pattern to hide information. As we know, each Chinese character has a simplified

pattern and a traditional pattern, and, we found statistically that about 50% of the commonly-used Chinese characters have simplified patterns and traditional patterns that lookalike. In Microsoft Word@, there is a method for encoding Chinese characters, known as BIG 5, that is used for the traditional Chinese characters in Taiwan, Hong Kong, and Macau. Mainland China, which uses simplified Chinese characters, uses GB2312 instead. In our proposed scheme, the information is embedded by changing the method of encoding the look-alike Chinese characters in the cover text. Since the encoding transformation operation between the traditional and simplified pattern in Microsoft Word@ is very simple, the proposed scheme achieves the goal of creating a simple process for embedding information. Our experimental results provided sufficient positive data to support the claim that the proposed scheme offers better visual steganographic effectiveness than Wang et al.'s L-R and U-D scheme, and our scheme is easier to use. The rest of this paper is organized as follows. Section2 briefly describes Wang et al.'s L-R and U-D scheme [14]. Then, the data embedding and extracting procedures of our proposed scheme are demonstrated in Section 3. Section 4 presents our experimental results, and our concluding remarks are given in Section 5.

2. Previous Work

In this section, we briefly review Wang et al.'s L-R and U-D scheme. They proposed a reversible, information-hiding scheme using left–right and up–down representations of Chinese characters.

The L-R and U-D scheme uses the mathematical expression of Chinese characters to hide the information into a text file. The mathematical expression of Chinese characters [12] was introduced into linguistic steganography by Sun, Lou, and Huang. The mathematical expression in their L-R scheme helps their scheme [13] achieves better hiding capacity and robustness than some other methods for hiding data in text [2, 17].

Wang et al.'s scheme used the mathematical expression from Sun et al.'s L-R scheme, and they added the up and down components of the Chinese characters to the candidate set. To overcome the two drawbacks of Sun et al.'s scheme, they added a reversibility function and designed a simple strategy for data extraction. During the embedding phase, their scheme chooses those characters with left and right components and up and down components as candidates in which to hide the secrets, as shown in **Fig. 1** [14]. First, the encoder scans the text to obtain the Chinese character, and then it judges whether the current character that is being processed is a candidate character or not. If it is, then, if the secret is '0,' the L-R and U-D scheme keeps the candidate character's original appearance; otherwise, the encoder uses the hiding function to hide the secret data '1' into the current embeddable character.

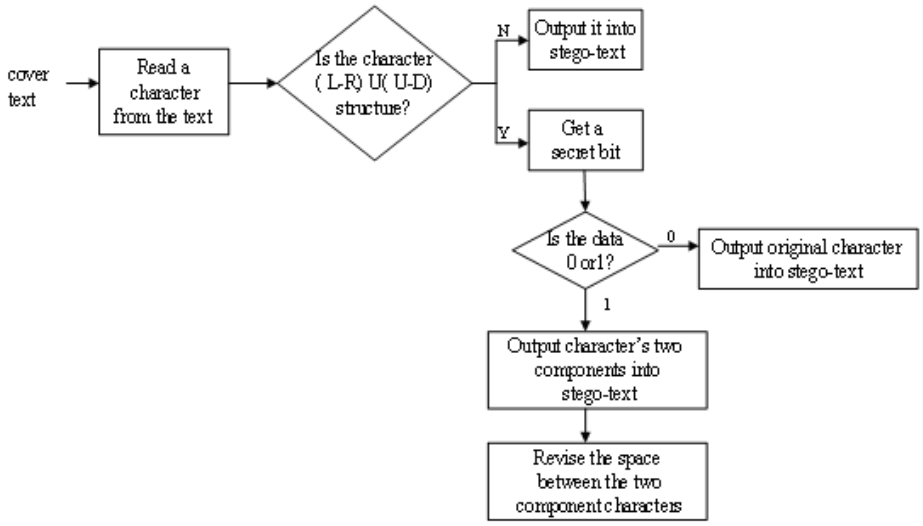

Fig. 1. Flowchart of Wang et al.'s data embedding scheme

The hiding function consists of two steps. In the first step, the encoder outputs the character's appearance by its two basic components' appearances, which can be either L-R or U-D structure. In the second step, the encoder uses Microsoft Word@'s character modification tool to change the spacing between the two glyphs, so that the two separate glyphs look like a single character rather than two separate characters. During the extracting phase, the proposed extracting procedure has two parts. The first is to extract the hidden secret, as shown in **Fig. 2**, and the second is to restore the cover text from the stego-text, as shown in **Fig. 3**.

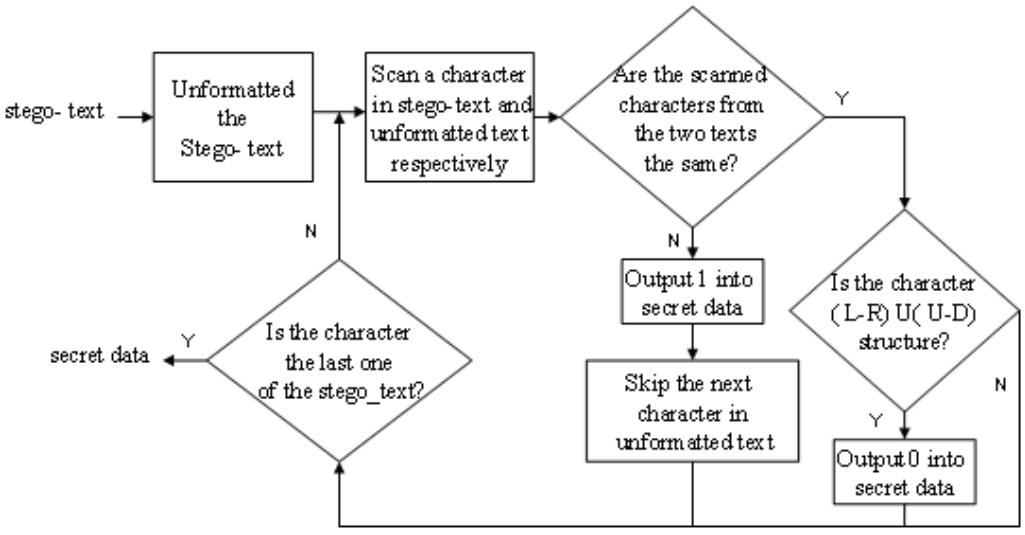

Fig. 2. Flowchart of Wang et al.'s secret data extracting procedure

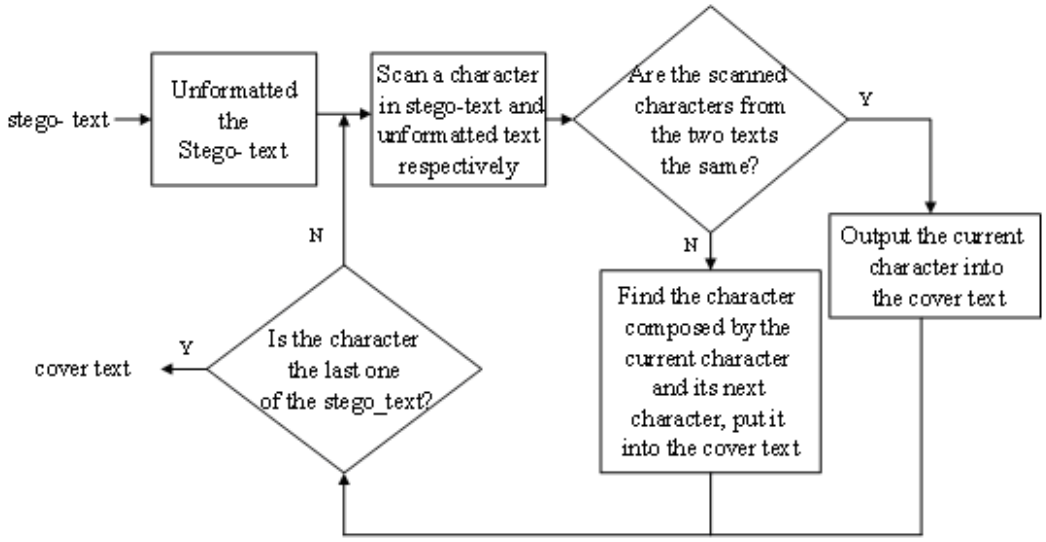

Fig. 3. Flowchart showing the procedure for restoring the cover text in Wang et al.'s embedding scheme

In the extracting procedure, the decoder builds a new, unformatted, text file and copies the content of the stego-text into it. Then, the decoder compares the contents of the unformatted text file and the stego-text file. The decoder scans the two different types of text simultaneously and builds a new text file, called the cover text file, to record the original cover text. If the two scanned characters are the same in the unformatted-text file and the stego-text file, the decoder copies the characters to the cover text file. If the character is an L-R or U-D structure character, the receiver extracts the bit ''0" as the secret datum. Otherwise, if the characters are different in both texts, the decoder extracts bit ''1" as the secret datum. Then, the decoder records the character that is currently being processed and its next character using the mathematical expression function, and the decoder determines which character can be divided into these two recorded characters. Finally, the decoder outputs the character that it found to the cover-text file. At the end of the extracting procedure, the receiver obtains two useful texts, i.e., the secret data and the original cover text.

3. The Proposed Scheme for Hiding Data In The Text

In this section, we present a detailed description of our proposed scheme that makes the embedding and extracting procedures for Chinese characters very simple and that provides significant improvement in the visual steganographic effectiveness.

By analyzing Chinese characters' traditional and simplified forms one by one in Microsoft Word@, we obtained two sets, i.e., a set of candidate characters to be used to embed data that have simplified patterns and traditional patterns that lookalike and a set of characters for which the two patterns are distinctly different. In **Table 1**, we listed some characters encoded in BIG 5 and in GB 2312 as examples to explain which candidate characters are defined as lookalike characters. In the list, "candidate characters" stands for the characters for which the simplified patterns encoded in GB 2312 have the same appearance with their corresponding traditional patterns encoded in BIG 5, while "non-candidate characters" identifies the characters that had distinctly different appearances in the two patterns.

Table 1. Examples of candidate characters and non-candidate characters

	Appearance of GB 2312 encoded characters	Codes for GB 2312 encoded characters	Appearance of BIG 5 encoded characters	Codes of BIG 5 encoded characters
candidate characters	大	2083	大	A46A
	日	4053	日	A4E9
	止	5425	止	A4EE
non-candidate characters	并	1802	并	A8C3
	证	5404	证	C3D2
	说	4321	说	BBA1

In this paper, we used the characters in the candidate set to hide information into the cover text. Before the data embedding and extracting phases, we had to determine which encoded standard, BIG 5 or GB 2312, is used in the Chinese cover text. We assumed that the input cover text was encoded in GB 2312.The following parts of the paper are subject to this assumption unless other information is provided. If the cover text is encoded by BIG 5 at first, then the process is almost the same as that of GB 2312 encoded cover text; the only difference is that the way in which the corresponding encoding of the lookalike character was adjusted according to the information to be embedded. After determining which encoded standard to use in the Chinese cover text, we used that standard to design efficient data embedding and extracting procedures, which are described in Sections 3.1 and 3.2, respectively.

3.1 Embedding Procedure

First, the encoder scans the text to obtain a Chinese character, then judges whether the character that is currently being processed is embeddable or not. Here, "embeddable" means that the Chinese character that is currently being processed is encoded by GB 2312 and has the lookalike appearance with its traditional pattern, which is encoded by BIG 5. The judgment is implemented by using Word's tool for making the transformation between the traditional Chinese character and the simplified Chinese character. If the result is positive, the encoder hides the secret datum into the current embeddable character by using the data embedding procedure; otherwise, the encoder outputs the character to the stego-text directly and then moves to the next Chinese character. In the secret data hiding procedure, first, the encoder extracts one bit from the information to be embedded. Second, if the extracted bit is '0', the encoder outputs the current Chinese character into the stego-text without modifying it; otherwise, the encoder transforms the current Chinese character into its traditional version by using Microsoft Word[@]'s transforming tool to convert the the simplified Chinese character to its traditional version.

The following example illustrates how this data-embedding procedure works. Let us assume that the content of the cover text is "天气阴暗", all of which are presented in the simplified pattern. First, the encoder scans the text to obtain the first character, which it reads as "天", and then, the encoder determines whether the character has the lookalike appearance as its traditional pattern. The answer is positive, so the encoder hides a secret bit in the

character. Assume that the secret data to be embedded is "10." The first secret bit "1" is embedded into "天" by outputting its traditional pattern "天" into the stego-text. Then, the encoder reads the second character in the text, which is"气", and its traditional pattern is "氣", which has an appearance that is quite different from "气". The encoder program judges this character as a non-candidate character and outputs the character directly to the stego-text without embedding any secret data. The next character is "阴", which has the traditional pattern "陰", the pattern of which also is obviously different from "阴", so the encoder outputs the original character to the stego-text. When the encoder reads "暗", which belongs to the candidate character set and the second secret bit to be embedded is "0," the encoder embeds "0" into "暗" by outputting it in the form of the simplified pattern. In the end, it is difficult for the human eye to recognize the difference between the stego-text "天气阴暗" and the cover text "天气阴暗." On the other hand, if the text is "天氣陰暗", which all presented in traditional pattern, is used as the cover text and the secret data '10' is to be embedded into it, then the stego text "天氣陰暗" can be obtained by adopting our proposed secret embedding procedure.

3.2. Extracting Procedure

In the extracting procedure, the decoder scans the characters in the stego-text. If the simplified pattern of the scanned character and the traditional pattern of the scanned character do not lookalike, the decoder copies the characters to the reconstructed cover text file. If the two patterns of the character have the same appearance and if the character is presented in the simplified pattern form, which means it is encoded by GB 2312, the decoder extracts "0" as the secret bit and outputs the character directly into the reconstructed cover text file. Otherwise, if the character is a traditional Chinese character encoded by BIG 5, the decoder extracts "1" as the secret data and outputs the character's simplified version into the reconstructed cover text file. The next paragraph gives an example to illustrate our data extracting procedure.

Let us consider the example given earlier in Section 3. The stego-text is "天气阴暗". First, the decoder scans the character"天", which is a candidate character encoded by BIG 5 since it has the lookalike appearance as its simplified pattern. The judgment is made by using Microsoft Word@'s transforming tool to convert the traditional Chinese character to the simplified Chinese character. Then, the decoder extracts "1" as the hidden secret data and outputs the simplified pattern of "天" into the reconstructed cover text file. The second scanned character, "气", is not a candidate character, so the decoder outputs "气" without any changes to the reconstructed cover text. The next scanned character is "阴", which is not a candidate character either, so the decoder outputs it directly to the reconstructed cover text. The last character "暗" is a candidate character, and it is presented in simplified pattern, so the decoder extracts "0" as the hidden secret bit and outputs "暗" directly to the reconstructed cover text. Finally, the decoder obtains the lossless, reconstructed cover text "天气阴暗" and the secret data "10."

4. Experimental Results

In this paper, we used Microsoft Word 2007@ to implement both the proposed scheme and Wang et al.'s scheme. The first test cover text is shown in **Fig. 4**.

超新星，一颗巨星生命的终结。一颗具有几十至上百倍太阳质量的恒星身躯四散溅落到太空之中，但是，大灾之后，星系再生。从新的废墟中，新的恒星又会诞生。它们会在这种气体尘埃云中点燃生命之火，而超新星则使星云物质更加丰富。爆炸所产生的冲击带来了许多金属元素，比如铁合金，重元素仅仅产生于最巨大的恒星之中。于是，一颗超新星养育了他的子孙后代。恒星的产生过程是这样的，首先在自引力作用下物质塌缩，而后形成一个盘，在盘的中心，一颗恒星开始闪耀，在它周围产生环状物，环分解而形成行星，太阳系就是这样演化的。地球在这兰色的“可居住”区域的轨道上安然旋转着，在这里，生命可能存活。

Fig. 4. Cover text

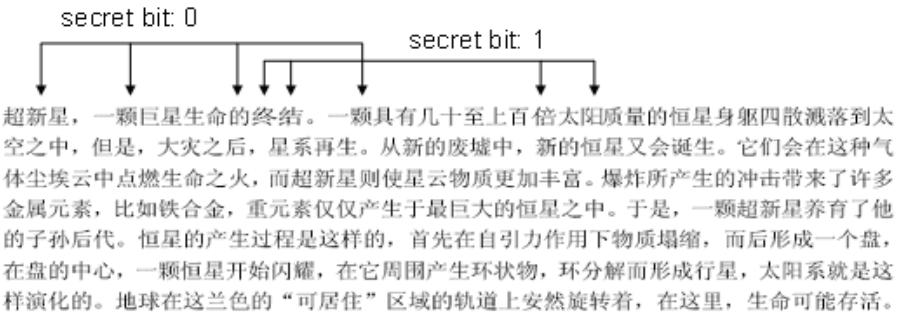

Fig. 5. Stego-text of Wang et al.’s scheme

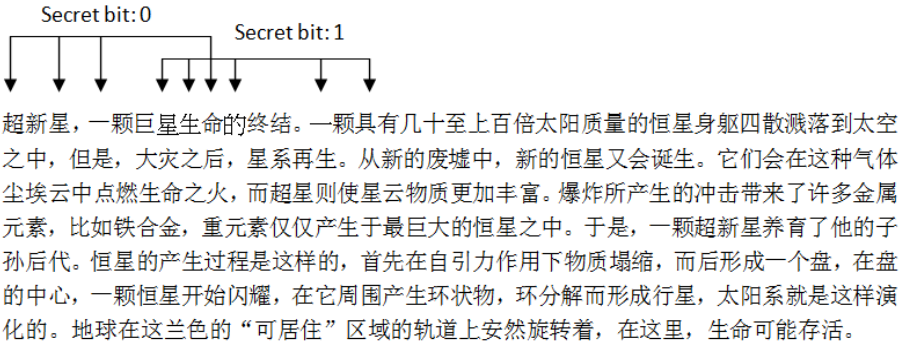

Fig. 6. Stego-text of our proposed scheme

In this cover text, there are 278 Chinese characters, 147 of which are candidate characters in the proposed scheme. So, there are 147 secret bits that can be embedded into the cover text by utilizing the proposed scheme. Since there are 115 L-R and U-D Chinese characters in the cover text, Wang et al.’s scheme can embed 115 secret bits. Obviously, in this example, our proposed scheme has a higher hiding capacity than Wang et al.’s scheme. In statistic, our proposed scheme’s look-alike, traditional, simplified Chinese characters make up 50% of the total characters, which means the amount of the secret data can be embedded is about a half size of the cover text without making any expansion to it. We assume that the secret data to be embedded into **Fig. 4** are “00011011.” **Fig. 5** shows the stego-text for Wang et al.’s scheme after the secret bits have been embedded, and **Fig. 6** shows the stego-text for our proposed scheme after the secret bits have been embedded. By observing the two sets of stego-text, it is apparent that the characters, which were used to embedded secret bit ‘1’, looks unnormal due

to the components' space adjustment in Wang et al.'s scheme, while those characters appear no different from their shapes in the cover text by using the proposed scheme. Therefore, the visual steganographic effectiveness of our proposed scheme is much better than that of Wang et al.'s scheme. In addition, during the implemetation of Wang et al.'s scheme, we found that the modification tool used to adjust the appearance of the characters for Wang et al.'s scheme is more difficult to implement than the transformation tool used in the proposed scheme. Since adjusting the distances between the components of a character is a very complex procedure, especially for characters with the up-down structure, it is more convenient and effective that this can be accomplished in our proposed scheme just by pressing the transformation tool button in Microsoft Word@.

To further demonstrate our scheme's advantage, we use other three different types of cover texts to test the visual steganographic effectiveness and the hiding capacity between our proposed scheme and Wang et al.'s scheme. **Fig. 7** shows the cover texts, the cover text 1 is a paragraph obtained from a novel, the cover text 2 is a paragraph cited from the news paper, and the cover text 3 is a paragraph extracted from a science journal. **Fig. 8** shows the corresponding stego-texts generated by our proposed scheme. From **Table 2**, we can see that when the same cover text is adopted, comparing with Wang et al's scheme, the occupied memory of the stego-text generated by our proposed scheme is smaller than 13 KB in average, while the required memory of the stego-text generated by Wang et al.'s scheme is about 35 KB in average. In terms of hiding capacity, Wang et al.'s scheme only hides about 37 bits in average, but ours hides about 84 bits in average which means the hiding capacity of ours is over about two times of Wang et al.'s scheme when the test cover texts are used.

Cover text 1:

病房里，一个生命垂危的病人从房间里看见窗外的一棵树，在秋风中一片片地掉落下来。病人望着眼前的萧萧落叶，身体也随之每况愈下，一天不如一天。她说："当树叶全部掉光时，我也就要死了。"一位老画家得知后，用彩笔划了一片叶脉青翠的树叶挂在树枝上。最后一片叶子始终没掉下来。只因为生命中的这片绿，病人竟奇迹般地活了下来。

Cover text 2:

中国证券报发改委：采取八措施巩固经济企稳回升势头

国家发展改革委主任张平 25 日说，当前我国经济发展正处在企稳回升的关键时期，一旦松懈就有可能出现反复，因此下半年要采取八大措施巩固这种回升势头。另外，张平表示，将从五方面加快发展方式转变和经济结构调整。

Cover text 3:

我国推行的重工业优先发展战略是一种典型的赶超战略。在经济上赶超先进国家，是所有落后国家和地区共同的热切愿望。然而，几乎所有实行赶超战略的经济，大都陷入诸如日益加深的城乡贫困化、旷日持久的高通货膨胀，以及积重难返的经济结构失衡的困境之中。

Fig. 7. Three cover texts.

Stego text 1:

病房里，一个生命垂危的病人从房间里看见窗外的一棵树，在秋风中一片片地掉落下来。病人望着眼前的萧萧落叶，身体也随之每况愈下，一天不如一天。她说："当树叶全部掉光时，我也就要死了。"一位老画家得知后，用彩笔划了一片叶脉青翠的树叶挂在树枝上。最后一片叶子始终没掉下来。只因为生命中的这片绿，病人竟奇迹般地活了下来。

Secret bits:1010011111000110101000111010101010110001111000111100101011000111100111000110

Stego text 2:

中国证券报 发改委：采取八措施巩固经济企稳回升势头

国家发展改革委主任张平 25 日说，当前我国经济发展正处在企稳回升的关键时期，一旦松懈就有可能出现反复，因此下半年要采取八大措施巩固这种回升势头。另外，张平表示，将从五方面加快发展方式转变和经济结构调整。

Secret bits: 1111000110110011110110010101111100010001110110101111011100001110

Stego text 3:

我国推行的重工业优先发展战略是一种典型的赶超战略。在经济上赶超先进国家，是所有落后国家和地区共同的热切愿望。然而，几乎所有实行赶超战略的经济，大都陷入诸如日益加深的城乡贫困化、旷日持久的高通货膨胀，以及积重难返的经济结构失衡的困境之中。

Secret bits: 0111011001100110010100110111011101001000110101111001001000111

Fig. 8. The stego-texts with their embedded bits.

Table 2. The performance results of the Wang et al.'s scheme and the proposed scheme

	Wang et al.'s scheme	The proposed scheme	Wang et al.'s scheme	The proposed scheme
Cover text 1	34.2	11.7	53	78
Cover text 2	36.4	12.1	39	81
Cover text 3	36.4	12.1	41	89
average	35.7	12.0	44.3	82.7

5. Conclusion

In recent years, text data hiding technology has increased at an astounding rate. The requirements for text hiding schemes are becoming more and more stringent in terms of hiding capacity, effectiveness, robustness, and usability. Therefore, in this paper, we proposed a simple and more secure data embedding scheme for text. To improve the visual

steganographic effectiveness and capacity of Wang et al.'s scheme, we proposed the transformation between traditional and simplified patterns of the same characters, since there is a set of Chinese characters that lookalike when they are encoded by GB 2312 and by BIG 5, and, all we have to do to transform the character between its simplified pattern and its traditional pattern is to press the transformation button in Microsoft Word@. The proposed scheme makes the embedding and extracting procedure reversible, and the original cover text can be easily derived after the secret data are extracted. In addition, based on the experimental results we obtained, our scheme improved the hiding capacity to some extent. Furthermore, the most significant difference between our proposed scheme and Wang et al.'s scheme is that we can use the transformation tool instead of dividing and merging the two neighboring component characters in the stego-text and revising the distance between them, which is difficult and time-consuming to accomplish.

In conclusion, by transforming traditional and simplified Chinese characters, our hiding scheme improves the visual steganographic effectiveness and hiding capacity, reduces the memory required for embedding information, and offers a simple and reversible embedding and extracting procedure. The experimental results prove that our proposed scheme has a powerful advantage over the L-R and U-D scheme.

Acknowledgments

This work was supported by the National Nature Science Foundation of China under Grant No. 61272374, No. 60133012 and No. 61272371.

References

[1] M. J. Atallah, V. Raskin, M. Crogan, C. Hempelmann, F. Kerschbaum and D. Mohamed, "Natural Language Watermarking: Design, Analysis, and a Proof-of-Concept Implementation," in *Proc. of the 4th International Workshop on Information Hiding, Pittsburgh, PA, USA*, pp. 185-200, 2001. Article (CrossRef Link).

[2] M. Chapman, "Hiding the Hidden: A Software System for Concealing Ciphertext as Innocuous Text," *Information and Communications Security,* pp. 335-345, 1997. Article (CrossRef Link).

[3] Y. L. Chiang, L. P. Chang, W. T. Hsieh and W. C. Chen, "Natural Language Watermarking Using Semantic Substitution for Chinese Text," in *Proc. of IWDW2003, Seoul, Korea*, pp. 129-140, 2003.Article (CrossRef Link).

[4] R. G. Kammer, "Data Encryption Standard (DES)", *National Institute of Standards and Technology*, 1999.Article (CrossRef Link).

[5] C. C. Chang, P. Y. Pai, C. M. Yeh, and Y. K. Chan, "A High Payload Frequency-based Reversible Image Hiding Method," *Information Sciences*, vol. 180, no. 11, pp. 2286-2298, 2010. Article (CrossRef Link).

[6] C. C. Chang and T. D. Kieu, "A Reversible Data Hiding Scheme Using Complementary Embedding Strategy," *Information Sciences*, vol. 180, no. 16, pp. 3045-3058, 2010.Article (CrossRef Link).

[7] C. C. Chang, Y. P. Hsieh, and C. Y. Lin, "Lossless Data Embedding with High Embedding Capacity Based on Declustering for VQ-compressed Codes," *IEEE Trans. Inf. Forensics Security*, vol, 2, no. 3, pp. 341-349, 2007.Article (CrossRef Link).

[8] C. C. Chang and C.Y. Lin, "Reversible Steganographic Method Using SMVQ Approach Bases on Declustering," *Information Sciences*, vol. 177, no. 8, pp. 1796-1805, 2007.Article (CrossRef Link).

[9] K. Maher, TEXTO, Article (CrossRef Link).
[10] B. Murphy and C. Vogel, "The Syntax of Concealment: Reliable Methods for Plain Text Information Hiding," *International Society for Optics and Photonics*, 2007.Article (CrossRef Link).
[11] R. Rivest, A. Shamir and L. Adleman, "Cryptographic Communications System and Method", *U.S. Patent 4405829*, 1977.Article (CrossRef Link).
[12] X. M. Sun, H. W. Chen, L. H. Yang, and Y. Y. Tang, "Mathematical Representation of a Chinese Character and its Applications," I*nternational Journal of Pattern Recognition and Artificial Intelligence*, vol. 16, no. 8, pp. 735-747, 2002.Article (CrossRef Link).
[13] X. M. Sun, G. Luo, and H. J. Huang, "Component-based Digital Watermarking of Chinese Texts," in *Proc. of the 3rd International Conference on Information Security, Shanghai, China*, pp. 76-81, 2004.Article (CrossRef Link).
[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*, vol. 82, no. 8, pp. 1362-1369, 2009.Article (CrossRef Link).
[15] O. Vybornova and B. Macq, "Natural Language Watermarking and Robust Hashing Based on Presuppositional Analysis," in *Proc. of the IEEE International Conference on Information Reuse and Integration (IRI 2007), Las Vegas, Nevada, USA*, pp. 177-182, 2007.Article (CrossRef Link).
[16] K. Winstein, "Lexical Steganography through Adaptive Modulation of the Word Choice Hash," Article (CrossRef Link)..
[17] S. F. Wu and L. S. Huang, "Research on Information Hiding," *US-China Education Review*, pp. 77-81, 2003.Article (CrossRef Link).
[18] "Spam Mimic," 2000.Article (CrossRef Link)..Spammimic.com
[19] P. Wayner, "Mimic Functions," *Cryptologia XVI*, pp. 193-214, 1992.Article (CrossRef Link).

Bin Feng received the BS degree in Computer Science and Technology in 2002 from the LiaoCheng University, Shandong, China, and the MS degree in software engineering in 2006 from the Dalian University of Technology, Dalian, China. He has been an assistant in TaiShan College among 2002-2004. He is currently a full engineer of Computer Science at DaLian University of Technology (DLUT) (Dalian, China), where he has been since September 2006.Since 2011 he is currently pursuing his PhD degree in computer software and theory from the Dalian University of Technology, Dalian, China. His research interests include data hiding, image processing, network and information security.

Zhi-Hui Wang received the BS degree in software engineering in 2004 from the North Eastern University, Shenyang, China. She received her MS degree in software engineering in 2007 and the PhD degree in software and theory of computer in 2010, both from the Dalian University of Technology, Dalian, China. Since November 2011, she has been a visiting scholar of University of Washington. Her current research interests include information hiding and image compression.

Duo Wang is currently a senior student in software school of Dalian University of Technology, Dalian, China. She majors in software engineering. Her research interests include information hiding, and image processing.

Ming-Chu Li received a Ph.D. degree from the University of Toronto, Toronto, Canada, in 1998. Prof. Li is currently a Full Professor of Computer Science at DaLian University of Technology (DLUT), Dalian, China, where he has been since September 2004. He is also the Vice Dean of School of Software of DLUT. His research interests include Hamiltonian Graph Theory, NP-Theory and Algorithms, Network and Information Security, Reputation Systems, and Grid computing and its applications.

Ching-Yun Chang received her B.S. in Electrical Engineering and Computer Science at National Tsing Hua University, Taiwan, in 2006, and her M.Sc. in Computer Science at University of Oxford, United Kingdom, in 2009. She is currently pursuing her Ph.D. in Computer Science at University of Cambridge, United Kingdom. Her current research interests include natural language processing, multimedia processing and information security.