A Reliable Data Hiding Scheme Using Jigsaw Sudoku

Wanli Lyu [a)], Dongshuo Yin [b)], Yu Zhang and Jiahao Deng

Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, School of Computer Science and Technology, Anhui University, Key Lab of Industrial Image Processing and Analysis of Anhui Province, China.

[a)] Corresponding author: wanly_lv@163.com
[b)] yindongshuo@qq.com

Abstract. A distinctive data hiding method based on Jigsaw Sudoku is proposed in this paper. Different from other methods, the proposed method is based on Jigsaw Sudoku which is a kind of irregularly shaped Sudoku. Due to the novelty of Jigsaw Sudoku, the proposed method can provide a safer scheme for data hiding than that with pure Sudoku. The proposed method can hide a certain number of 9-ary numbers in a steganographic image which has a corresponding specification. With a 256*256 key matrix stretched from a size 9*9 Jigsaw Sudoku, secret message can be hidden in a vector image byte by byte in the form of pictures. Experimental results show that, the proposed method has a better peak-signal-to-noise-ratio (PSNR) compared with some other data hiding methods.

Key words: Sudoku; PSNR; hiding methods; irregularly shaped; steganographic image; corresponding specification.

INTRODUCTION

In the field of data hiding and digital watermarking, steganography is a technique that hides information in a digital media carrier. Various digital media can be used to conceal secret data, such as digital images, video streaming, electronic documents, etc. By using steganography, we can propagate the image embedded with the secret message in a public signal path avoiding being detected by possible attackers. In order to judge and contrast various hidden methods, three criteria including hiding capacity, visual quality of stego-images and security analysis are widely used. However, embedded capacity and visual quality are often inversely related. In other words, if the embedded capacity is increased, the visual quality will be reduced. Therefore, a trade-off between embedded capacity and visual quality is often made by users.

The most famous data hiding method is the least-significant-bit (LSB) replacement method proposed in 1989 [7], which is also an early simple information hiding method. LSB-based data hiding methods can be divided into two categories: LSB replacement and LSB matching [1]. The former is to simply replace the last few bits of the color value with secret messages. The latter is to increase or decrease in a certain range to maintain the original color value. Some LSB-based data hiding methods are very easy to use. However, it is hard to resist some simple attacks [8] [9]. For example, the secrets hidden by the LSB replacement method can easily be found by attackers using chi-square analysis, so some LSB-based methods are insecure.

In order to further improve the security of data hiding technique, Zhang and Wang [2] proposed exploiting modification direction (EMD) method in 2006 to improve Mielikainen's [1] LSB matching method. With the EMD method, n 5-ary numbers can be embedded into a digital image containing at least $(2n + 1)$ pixels. In the worst case, only one pixel’s value will be increased or decreased by one in each pair. In 2008, Chang et al. [3] proposed a data hiding method using traditional Sudoku inspired by EMD. Afterwards, Zhang and Zheng [4] proposed a watermarking scheme based on Fourier descriptor and Sudoku in 2015, which also shows that Sudoku becomes more and more important in information hiding system. Moreover, Nguyen and Chang [11] and Kumar et al. [10] proposed some reversible data hiding scheme using traditional Sudoku in 2015 and 2017, respectively.

The goal of the paper is to propose a novel method of information hiding to enhance the embedding capacity and the security of secret messages using Jigsaw Sudoku. Jigsaw Sudoku is a variant of classical 9×9 Sudoku. According to Jigsaw Sudoku properties, the proposed method first converts a binary secret message to secret digits in a base-9 number system and then modifies the value of the cover pixel pairs to conceal the secret digits. Due to the diversity and irregularity of each sub-block, Jigsaw Sudoku is harder for attackers to access. Therefore, the security of information hiding is improved through the proposed method.

RELATED WORK

In this section, Chang et al.'s data hiding method based on Sudoku will be described briefly. To easily convey their embedding algorithm, a flow chart is shown in Fig. 1.

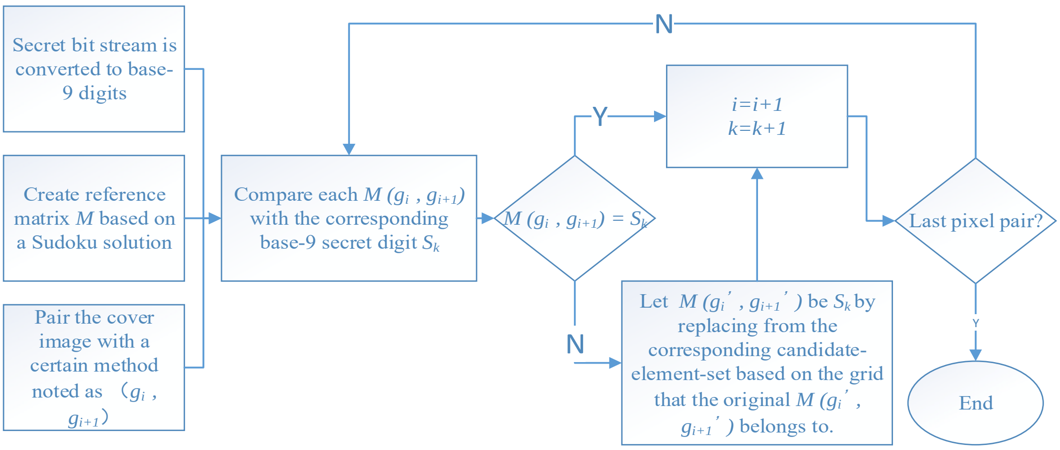


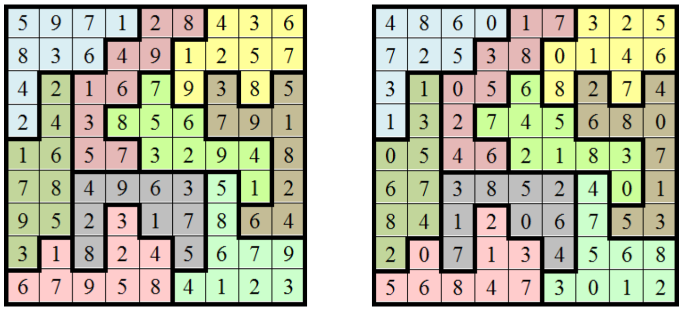
FIGURE 1. Flow chart of Chang et al.'s embedding algorithm.

PROPOSED METHOD

Jigsaw Sudoku

Sudoku is a logical puzzle with the goal of filling a 9×9 grid with numbers from 1 to 9 [5] [6]. In 1979, Garns invented the Sudoku Puzzle and Dell Magazine released it under the name "Digital Places." Sudoku was popular with publisher Nikoli in Japan in 1986 and was internationally known in 2005. In the same year, Felgenhauer and Jarvis analyze the classic 9×9 Sudoku solutions to show that total number of possible solutions is nearly 6.671×1021. In 2007, Russell and Jarvis' work [6] prove that if various possible symmetries are allowed, then the number of fundamental solutions of 9×9 Sudoku grid is 5,472,730,538.

Jigsaw Sudoku is a variant of Sudoku with the same 9×9 sized matrix, but its sub-blocks are irregular shapes rather than nine 3×3 matrices. In general, the shapes between almost every sub-block are different. The rule to solve a Jigsaw Sudoku puzzle, similar as traditional Sudoku, is to fill in the puzzle so that the numbers 1 through 9 occur exactly once in each row, column, and each sub-block. Fig.2 (a) is an example of Jigsaw Sudoku solution.



(a) (b)

FIGURE 2. (a) A Jigsaw Sudoku solution. (b) The corresponding matrix.

The Embedding Phase

First, to suit the proposed method, all the digits in Sudoku need to be decreased by one so that the Jigsaw Sudoku only contains digits from 0 to 8 like Fig.2 (b) shows. Secondly, the secret bitstream is also converted into base-9 numbers as above. The secret bit stream is converted into several α -bit segments. Then, each α -bit segment is converted into λ base-9 numeral system digits. The value of α is calculated through $\alpha = \lfloor \lambda \times \log_2 9 \rfloor$ where $\lambda = 3$. Then α is equal to 9. For example, a 9-bit segment 110010011_2 is converted into two secret digits 487_9. Let us denote the converted secret digits $S = \{s_1 s_2 s_3 ... s_n\}$, where n is the total number of the converted secret digits.

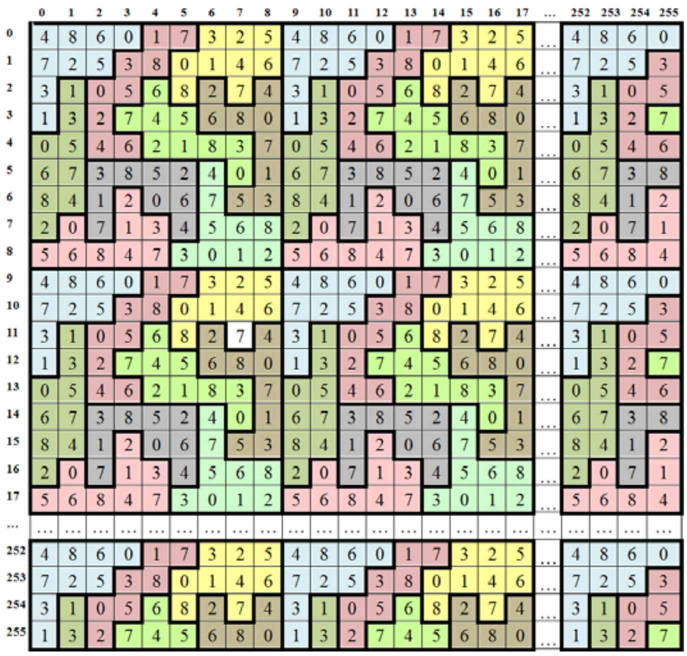


FIGURE 3. An example of key matrix $K*$ stretched from Fig.2 (b).

$K*$, as a key reference matrix stretched from a decided 9×9 Jigsaw Sudoku solution, is a 256×256 matrix. The main idea of the proposed method is to modify the original pixel pairs of the cover image based on $K*$ to embed secret bits. Fig.3 presents an example of the key matrix $K*$.

After the key reference matrix K^* has been prepared, cover pixels are divided into pixel pairs. Let I be the cover image, thus, $I=\{g_1, g_2, ..., g_{H\times M}\}$ where H and M are the height and length of I. For each pair of pixels (g_i, g_{i+1}), it can locate an element on key matrix K^* whose row and column are g_i and g_{i+1}. Considering Jigsaw Sudoku properties, three different candidate sets noted as CEH, CEV, and CEB respectively, are created based on three rules as follows.

RULE 1: (i.e. chose the element set for CEH)

If $g_{i+1} \in (3, 252)$, then

$$CE_H = \{K(g_i, g_{i+1}-4), K(g_i, g_{i+1}-3), K(g_i, g_{i+1}-2), K(g_i, g_{i+1}-1), K(g_i, g_{i+1}), K(g_i, g_{i+1}+1), K(g_i, g_{i+1}+2), K(g_i, g_{i+1}+3), K(g_i, g_{i+1}+4)\};$$

Else If $g_{i+1} \leq 3$, then

$$CE_H = \{K(g_i,0), K(g_i,1), K(g_i,2), K(g_i,3), K(g_i,4), K(g_i,5), K(g_i,6), K(g_i,7), K(g_i,8)\};$$

Else If $g_{i+1} \geq 252$, then

$$CE_H = \{K(g_i,247), K(g_i,248), K(g_i,249), K(g_i,250), K(g_i,251), K(g_i,252), K(g_i,253), K(g_i,254), K(g_i,255)\};$$

RULE 2: (i.e. chose the element set for CEV)

If $g_{i+1} \in (3, 252)$, then

$$CE_H = \{K(g_i-4, g_{i+1}), K(g_i-3, g_{i+1}), K(g_i-2, g_{i+1}), K(g_i-1, g_{i+1}), K(g_i, g_{i+1}), K(g_i+1, g_{i+1}), K(g_i+2, g_{i+1}), K(g_i, +3g_{i+1}), K(g_i+4, g_{i+1})\};$$

Else If $g_{i+1} \leq 3$, then

$$CE_H = \{K(0, g_{i+1}), K(1, g_{i+1}), K(2, g_{i+1}), K(3, g_{i+1}), K(4, g_{i+1}), K(5, g_{i+1}), K(6, g_{i+1}), K(7, g_{i+1}), K(8, g_{i+1})\};$$

Else If $g_{i+1} \geq 252$, then

$$CE_H = \{K(247, g_{i+1}), K(248, g_{i+1}), K(249, g_{i+1}), K(250, g_{i+1}), K(251, g_{i+1}), K(252, g_{i+1}), K(253, g_{i+1}), K(254, g_{i+1}), K(255, g_{i+1})\};$$

RULE 3: (i.e. chose the element set for CEB)

If g_i or $g_{i+1} \in \{252, 253, 254, 255\}$, then

g_i or $g_{i+1} = 251$,

as pretreatment considering the feasibility of the method and the quality affected to the cover image due to the fact that 256 equals to 9×28 plus 4.

Then

$$CE_B = \{K(g_a, g_b) | (g_a, g_b) \in Grid_{(g_i, g_{i+1})}\},$$

where $Grid_{(g_i, g_{i+1})}$ is the corresponding Jigsaw Sudoku grid that (g_i, g_{i+1}) belongs to.

Fig. 4 shows an example of CEH, CEV, and CEB of (4,3) using different colors. One of the three candidate elements sets are selected to reposition (g_i, g_{i+1}) to (g_i', g_{i+1}') so that $K(g_i', g_{i+1}')$ is equal to the corresponding secret base-9 digit.

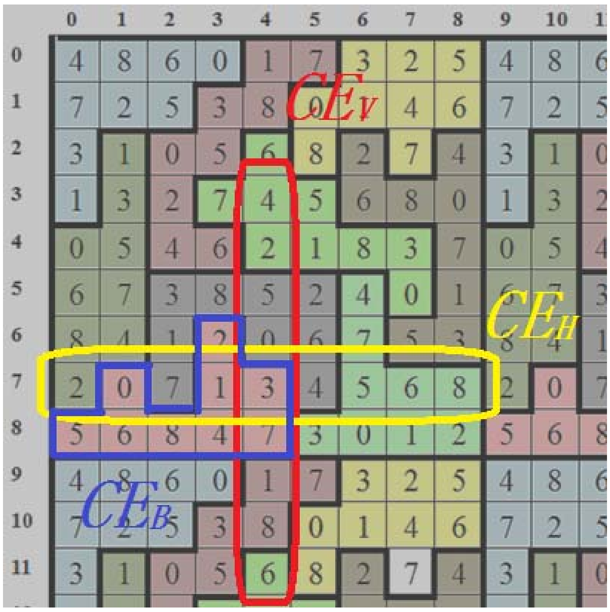


FIGURE 4. Illustration of CEH (yellow area), CEV (red area) and CEB (blue area).

The Extraction Phase

Same as above, the stego-image is divided into pixel pairs (g_i', g_{i+1}'), the secret base-9 digits can be located accurately with key matrix K^*. Each $K(g_i', g_{i+1}')$ refers to a secret base-9 digit, and finally all the secret digits are extracted form the whole secret message.

EXPERIMENTAL RESULTS

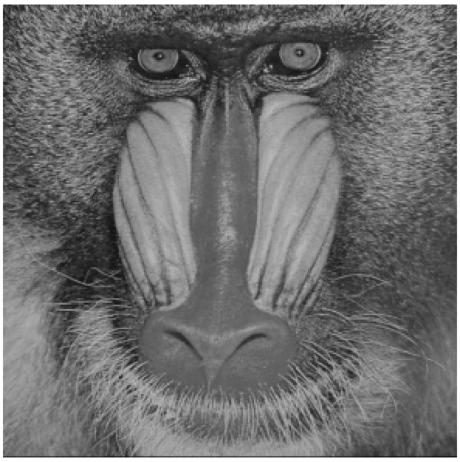
(a) Tiffany (b) Lena

(c) Gold hill

(d) Bridge

(e) Boats

(f) Barbara



(g) Baboon

(h) Zelda

FIGURE 5. The eight test images.

In this section, a 512×256 sized secret image is embadded in nine classical grayscale images sized 512×512 shown in Fig. 4 through the proposed method using MATLAB R2016a. At the same time, EMD method, Chang et al.’s method and Zhang and Zheng’s method are also implemented in our simulation to form a comparison among these three methods.

Two criteria are used for evaluating when comparing the three methods’ performance, which are also widely used in other imformation hiding system: Embedding Capacity and Visual Quality of the stego-image.

The peak-signal-to-noise-ratio (PSNR), which is commonly used to compare visual quality in imformation hiding system to evaluate the similarity between original image I and the corresponding stego-image I’, is calculated in our simulation. It is esay to calculate PSNR through its definition as follows.

$$PSNR = 10 \times \log_{10}(\frac{255^2}{MSE})dB \tag{1}$$

where 255 is the maximum value of each pixel and the Mean-Square-Error (MSE) for an image is defined in Equation (2).

$$MSE = \frac{1}{H \times W}\sum_{i}^{H}\sum_{j}^{W}(I_{ij} - I_{ij}')^2 \text{ a)} \tag{2}$$

where H and W are the height and wideth of the cover image and stego-image, Iij and Iij’ are the row i, column j pixel of the original image and the stego-image. A larger PSNR means that stego-image has a good visual qulity, on the other hand, the less the worse.

Then, for embedding capacity, bit-per-pixel which is also commonly used in similar papers is calculated according to its definition as follows.

$$C = \frac{|S|}{H \times W}\text{(bpp)} \tag{3}$$

where S is the number of all secret bits that embedded into all cover pixels. C is also positively related to embedding capacity, that is, the larger C is, the more secret bits are embedded.

TABLE 1. The result of Embedding Capacity and PSNR comparisons.

Images	EMD Method		Chang et al.’s Method		Zhang and Zheng’s Method		Proposed Method	
	PSNR	C	PSNR	C	PSNR	C	PSNR	C
Tiffany	52.11	1	45.02	1.5	44.73	1.5	47.55	1.5
Lena	52.12	1	44.97	1.5	44.68	1.5	47.29	1.5
Gold hill	52.11	1	44.84	1.5	44.76	1.5	47.49	1.5
Bridge	52.11	1	44.07	1.5	44.71	1.5	47.44	1.5
Boats	52.11	1	44.94	1.5	44.69	1.5	47.50	1.5
Barbara	52.11	1	44.77	1.5	44.50	1.5	47.30	1.5
Baboon	52.11	1	44.68	1.5	44.62	1.5	47.77	1.5
Zelda	52.12	1	44.89	1.5	44.65	1.5	47.52	1.5
Average	52.11	1	44.65	1.5	44.67	1.5	47.48	1.5

The result of our simulation is summrized in Table 1. The result shows that the proposed method perform better than EMD method in embedding capacity and equal to Chang et al.’s method. What’s more, the proposed method performs a 47.48 dB PSNR on average which is better than Chang et al’s method based on the properties of traditional Sudoku and Zhang and Zheng’s method based on fourier descriptor and Sudoku. Last but not least, the proposed method is much simple in theory and easier to implement than Zhang and Zheng’s method.

CONCLUSION

Steganography is a science of data hiding and delivery. This paper proposed a new version of Sudoku data hiding based on Chang et al.'s method using Jigsaw Sudoku rather than traditional Sudoku. Due to the large number of possible solutions of Sudoku, it is very hard to get the key matrix $K*$. So, the security of the proposed method is guaranteed and only the target users can extract the secret information. The experimental results show that the proposed method's visual quality is higher than 47 dB in average which is 2.83dB higher than Chang et al.'s method and 2.81dB higher than Zhang and Zheng's method. Moreover, the proposed method achieves higher embedding capacity in comparison with EMD method.

REFERENCES

1. J. Mielikainen, "LSB Matching Revisited," IEEE Signal Processing Letters, Vol. 13, No. 5, May 2006, pp. 285-287.
2. X. Zhang and S. Wang, "Efficient Steganographic Embedding by Exploiting Modification Direction," IEEE Communications Letters, vol. 10, no. 11, pp. 781-783, 2006.
3. C.-C. Chang, Y.C. Chou, and T.-D. Kieu, "An information hiding scheme using Sudoku". Proc. Third Int. Conf. Innovative Computing, Information and Control, Dalian, China, June 2008, pp. 17–22.
4. L. Zhang and H. Zheng, "A High Capacity Watermarking Scheme Based on Fourier descriptor and Sudoku." International Symposium on Computational Intelligence and Design IEEE, 2016:637-642.
5. B. Felgenhauer and F. Jarvis, "Mathematics of Sudoku I," Mathematical Spectrum, 39(1):15–22, September 2006.
6. E. Russell and F. Jarvis, "Mathematics of Sudoku II," Mathematical Spectrum, 39(2):54–58, January 2007.
7. L. F. Turner. Digital Data Security System. Patent IPN, WO 89/08915, http: // en. wikipedia. org/ wiki/ Mathematics of Sudoku, 1989.
8. S. Guillermito: Chi-square Steganography Test Program, http://www.guillermito2.net/stegano/tools/index.html
9. A.D. Ker, "Steganalysis of LSB Matching in Grayscale Images," IEEE Signal Processing Letters 12(6), 441–444 (2005).
10. M. S. Kumar, et al, "Data hiding with dual based reversible image using sudoku technique," Advances in Computing, Communications and Informatics (ICACCI), 2017 International Conference on. IEEE, 2017.
    T. S. Nguyen, and C.-C. Chang, "A reversible data hiding scheme based on the Sudoku technique," Displays 39 (2015): 109-116.