# Stego Keys Performance on Feature Based Coding Method in Text Domain

*Roshidi* Din[1,*], *Azman* Samsudin[2] , and *Sunariya* Utama[1]

[1]School of Computing, CAS, Universiti Utara Malaysia, Sintok, Kedah, Malaysia
[2]School of Computer Sciences, SoCS, Universiti Sains Malaysia, Pulau Pinang

**Abstract.** A main critical factor on embedding process in any text steganography method is a key used known as stego key. This factor will be influenced the success of the embedding process of text steganography method to hide a message from third party or any adversary. One of the important aspects on embedding process in text steganography method is the fitness performance of the stego key. Three parameters of the fitness performance of the stego key have been identified such as capacity ratio, embedded fitness ratio and saving space ratio. It is because a better as capacity ratio, embedded fitness ratio and saving space ratio offers of any stego key; a more message can be hidden. Therefore, main objective of this paper is to analyze three features coding based namely CALP, VERT and QUAD of stego keys in text steganography on their capacity ratio, embedded fitness ratio and saving space ratio. It is found that CALP method give a good effort performance compared to VERT and QUAD methods.

## 1 Introduction

With the rapid growth of Internet communication, information protection needs have become a critical issue in order to secure information [1]. One of the advanced promising researches on information protections for the next generation through untrusted communication channels is a steganology field. Unlike cryptology, which utilizes the encrypted and decrypted information of secret message which rendering the cover messages completely meaningless, steganology keeps the cover messages perceptually unchanged after concealing and detecting of the covered message. In speaking of steganology, there are two main branches [2]. Steganography is concerned with avoiding the suspicion of hidden messages in manipulated information [3]. Meanwhile, steganalysis concerned is discovering and rendering covert messages in given information [4].

Steganography is usually mislabeled with cryptography, because both are aimed at protecting valuable information. Whereas cryptography does not conceal the communication but only scrambles the data to prevent eavesdroppers understanding the content. Meanwhile steganography is study of hidden information [5] which is concealing

[*] Corresponding author: roshidi@uum.edu.my

the existing messages. There are two aspects of steganography, namely technical steganography and text steganography. Technical steganography concentrates on channel capacity which is concerned hiding information via a medium such as image, audio, video or other digitally invisible code, while text steganography concentrates on using written text to conceal secret messages [6].

## 2 Related Works on Text Steganography

Text steganography is the art of using natural language to conceal secret messages based on manipulating structure in the text [7]. There are two types of text steganography methods which are word-shift coding and feature coding.

Word-shift coding alters the text document by horizontally shifting the locations of words within the text lines to encode the document in a unique manner. Huang and Yan [8] developed a word-shift method based on character coding that modifies the inter-word spaces. Yang and Kot [9] have proposed an integrated character and word spaces which are based on facts that inter character spaces .Then, Liu et al. [10] proposed a method of word shift coding that can use in online to cover hidden message in plain text.

Next method, feature coding that alters a unique feature characteristic in text based on code word [11]. Rabah [12] has suggested some of the text features that can be altered in a large volume without making the reader aware of the existence information in the text. In Chinese letter, Sun et al. [13] has proposed a component-based method which is a novel text watermarking algorithm based on the mathematical expression. Besides, In Arabic letter, Shahreza and Shahreza [14] suggested that the information can be hidden in a text by changing the place of point in the Arabic alphabets. In Microsoft Word document, Stutsman et al. [15] has suggested a hidden message method by using generated sentences in multiple translations in term of Microsoft Word. Moreover, several methods of feature coding that manipulated pattern letter is English grammar such Changing in Alphabet Letter Patterns method known as CALP is trying to utilize English letters by mapping the binary sequence of the hidden for embedding process of the cover text [16]. Then, based multiple character-based [17] such as curve-based, vertical-based (VERT) and quadruple-based (QUAD).

Therefore, the objective of this study is to measure the fitness performance of stego keys using feature coding method in order to hide hidden text in the cover text.

## 3 Experimental Designs of Research

There are several stages that involved in the experimental design which is shown in **Fig. 1.**

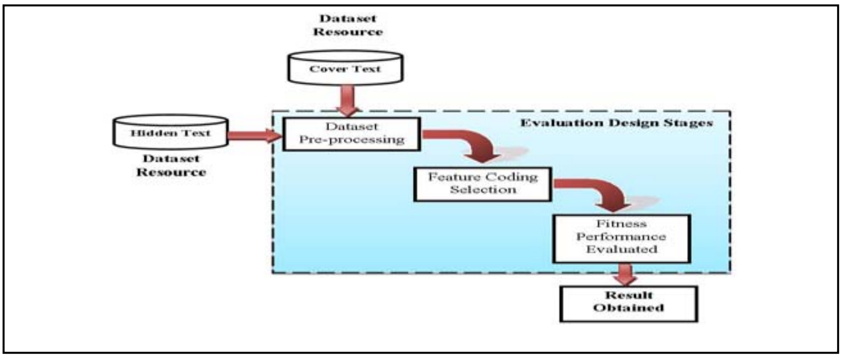

**Fig. 1.** Visualization of experimental design

DOI: 10.1051/shsconf/20173300036

This study consist of dataset such as cover text and hidden text that shown in **Fig. 2**. Then, three techniques are considered as feature coding methods used as the stego keys namely Changing Alphabet Letter Patterns called CALP [16], Vertical Straight Line called VERT and Quadruple Categorization called QUAD [17].

## 4 Fitness Performances Evaluated

The objective of this stage is to measure the fitness performance of the stego keys in text steganography domain from several researchers' works [18-21] such as capacity ratio, embedded fitness ratio and saving space ratio. Capacity ratio (CR) is used to determine the total of a hidden text that can be embedded in the cover text that can formulate as;

$$CR = \left[ \frac{Total\ Bits\ of\ Hidden\ Text}{Total\ Bits\ of\ Cover\ Text} \right] \text{x } 100\% \tag{1}$$

$$= \frac{\sum_{i=1}^{1<m<n} a_i}{\sum_{i=1}^{1<m<1n} b_i} \times 100\% \tag{2}$$

where

*a* = *Total bits of hidden text*
*b* = *Total bits of cover text*

Meanwhile, embedded fitness ratio (ER) is used to determine the total fitness of a hidden text can be embedded in a cover text. The ER of the stego keys can be calculated as;

$$ER = \left[ \frac{Total\ Bits\ of\ Stego\ Text - Total\ Bits\ of\ Cover\ Text}{Total\ Bits\ of\ Cover\ Text + Total\ Bits\ of\ Hidden\ Text} \right] \text{x } 100\% \tag{3}$$

$$= \frac{\sum_{i=1}^{1<m<n} a_i}{\sum_{i=1}^{1<m<n} b_i} \times 100\% \tag{4}$$

where

*a* = *Total number of embedded bits*
*b* = *Total bits of expected stego text*

Finally, saving space ratio (SSR) is used to determine the total space of text that can be saved during embedding process in a cover text that can be calculated as;

$$SSR = \left[ \frac{Total\ Bits\ of\ Expected\ Stego\ Text - Total\ Bits\ of\ Stego\ Text}{Total\ Bits\ of\ Expected\ Stego\ text} \right] x\ 100\% \tag{5}$$

$$= \frac{\sum_{i=1}^{1<m<n} a_i}{\sum_{i=1}^{1<m<n} b_i} \times 100\% \tag{6}$$

where

*a* = *Total number of saving space bits*
*b* = *Total bits of expected stego text*

## 5 Fitness Performance Analysis

The fitness performance analysis of stego keys are investigated using 20 different types of cover text. The result of evaluated the stego keys are elaborated in the following discussion.

### 5.1. Fitness performance of CALP

This section explains the measurement used to measure the fitness performance of CALP on an analyzed text. Thus, the score values of capacity ratio (CR), embedded fitness ratio (ER) and saving space ratio (SSR) for CALP technique is shown in **Fig. 2**.

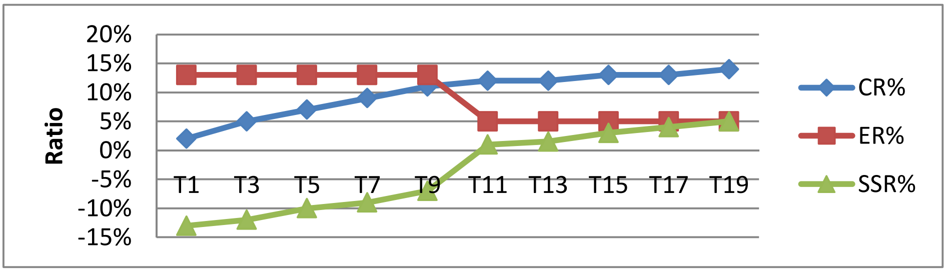

**Fig. 2**. Score values of fitness performance for CALP

**Fig. 2** shows the number value of fitness performance for CALP. The result shows that the CR has a stable increase from 2% until closely 15% throughout the analyzed text. As for ER, the percentage ratio maintains between 14% and 13% throughout T1 to T9 and suddenly declines from 13% to 5% between T9 and T10. The percentage remains unchanged at 5% between T10 to T20. As for SSR, the percentage ratio starts at almost -11% at T1. Increase is seen at T1 when the percentage moves from -6% to 1% at T10 and gradually increases from 1% to 5% between T10 and T20.

### 5.2. Fitness performance of VERT

**Fig. 3** shows the score value of fitness performance for VERT. The result shows that the CR has a steady increase from almost 1% to nearly 12% between T1 and T17 and slightly maintains at almost 12% from T17 to T20. As for ER, the percentage ratio has a gradual decrease from at nearly 13% to 12% between T1 and T9. However, there is a sudden spike closely at 12% to almost 16% of the percentage from T9 to T20. Meanwhile for SSR, the percentage ratio starts at less than -12% at T1 and sharply increases to -5% at T10. The trend remains stable at -5% from T10 to T17 before it falls slightly to -6% at T20.

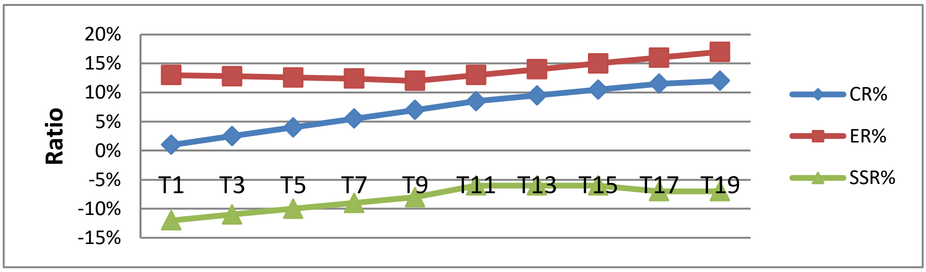

**Fig. 3**. Score values of fitness performance for VERT

### 5.3. Fitness performance of QUAD

**Fig. 4** shows the score value of fitness performance for QUAD. The result shows that the CR has a steady increase from almost 1% to closely 12% from T1 to T17 and remains stable at 12% at T18. As for ER, the percentage ratio gradually decreases from nearly 13% to 12% between T1 and T10. Then, there is a sharp increase from T10 to T20 with a percentage of 12% and 17% respectively. Meanwhile, for SSR, the percentage ratio starts with less than -12% and steadily increases to almost -5% from T1 to T10. Then, the trend shows a constant movement from T10 to T17 with a percentage of -5% before it has a slight decrease to -6% at T20.

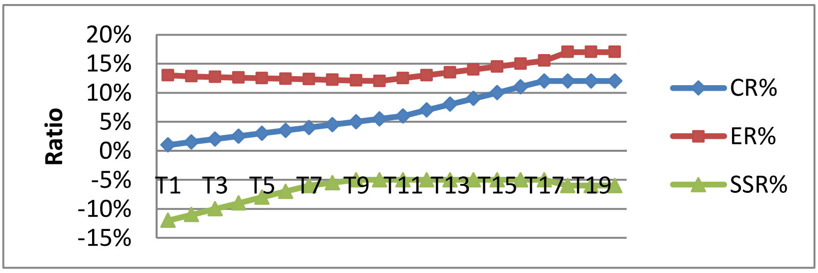

**Fig. 4**. Score values of fitness performance for QUAD

## 6 Conclusions

This study evaluated the three methods in implementation of fitness measurements of the stego keys based on capacity ratio, embedded fitness ratio, and saving space ratio. The three categories of text steganography techniques have been evaluated. It has been identified that CALP technique gives a better effort performance compared to VERT and QUAD technique. However, VERT and QUAD techniques give a quite similar result for embedding ratio and saving space performance compare to CALP. The results demonstrate that all of the methods are able to perform consistently with the utilization of hidden text and cover text. In future, this study proposes to evaluate a robustness of each technique in order to find a strength capability on text steganography from steganalysis activities.

This work was supported by 2016 Non-PI University Grant under RIMC Grant, Universiti Utara Malaysia (UUM).

## References

1. J. M. Ahmed Z. M. Ali., IJCSNS **11,** 4 (2011)
2. R. Chandramouli, N. D. Memon, Proceeding SPIE Security and Watermarking of Multimedia Contents 5020,173 – 177 (2003)
3. C. Y. Chang, S. Clark, The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 591 – 599 (2010)
4. H. A. Aboalsamh, H. Mathkour, M. Mursi, G. M. Assassa, Proceedings of the 12th WSEAS International Conference on Computers, 1011-1018 (2008)
5. R. Chandramouli, N. D. Memon, SPIE Security and Watermarking of Multimedia Contents 5020,173 – 177 (2003)
6. M. Chapman, G. I. Davida, M. Rennhard, Proceedings of the Information Security Conference (ISC '01), Malaga, Spain, 156 -165. (2001).

7. M. Chapman, G. I. Davida, M. Rennhard, Proceedings of the Information Security Conference (ISC '01), Malaga, Spain, 156 -165 (2001)
8. D. Huang, H. Yan, IEEE Trans. Circuits and Systems for Video Technology, **11**, 1237-1245 (2001)
9. H. Yang, A. C. Kot, IEEE International Conference on Multimedia and Expo, Taipei, Taiwan, **2**, 955 – 958 (2004)
10. M. Liu, Y. Guo, L. Zhou, Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 807-811 (2009)
11. M. V. Nasab, B. M. Shafiei, Aust. j. basic appl. sci, **5**, 12 (2011)
12. K. Rabah, Information Technology Journal, **3**, 3 (2004)
13. X. Sun, G. Luo, H. Huang, Proceedings of the 3rd ACM International Conference on Information Security, Shanghai, China, **85**, 76-81 (2004)
14. M. H. S. Shahreza, M. H. Shahreza, International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 1524-1526 (2008)
15. R. Stutsman, M.J. Atallah, C. Grothoff, K. Grothoff, Proceeding of the 21st Annual ACM Symposium on Applied computing (SAC 2006), Dijon, France, 338-345 (2006)
16. S. Bhattacharya, P. Indu, Duta, S. A. Biswas, G. Sanyal, Journal of Global Research in Computer Science, **2**, 3 (2011)
17. S. Dulera, D. Jinwala, A. Dasgupta, IJNSA, **3**, 6 (2011)
18. M. Agarwal, Computer Networks and Communications (NetCom). Springer, New York, 477-488 (2013)
19. L. Y. Por, D. Beh, T. F. Ang, S. Y. Ong, IAJIT **10**, 1 (2013)
20. C. F. Lee, and H. L. Chen, Springer Berlin Heidelberg, 155-179 (2013)
21. P. Akhilandeswari, J. G. George, Proceedings of International Conference on Internet Computing and Information Communications, Springer, India, 1-7 (2014)