Bulletin of Electrical Engineering and Informatics
Vol. 9, No. 2, April 2020, pp. 764~770
ISSN: 2302-9285, DOI: 10.11591/eei.v9i2.2069

Analysis review on feature-based and word-rule based techniques in text steganography

Farah Qasim Ahmed Alyousuf[1], Roshidi Din[2]
[1]Information Technology Department, Lebanese French University, Iraq
[2]School of Computing, UUM College of Arts and Sciences, Universiti Utara Malaysia, Malaysia

Article Info

Article history:

Received Oct 20, 2019
Revised Dec 28, 2019
Accepted Jan 24, 2020

Keywords:

Feature-based
Text steganography
Word-rule based

ABSTRACT

This paper presents several techniques used in text steganography in term of feature-based and word-rule based. Additionally, it analyses the performance and the metric evaluation of the techniques used in text steganography. This paper aims to identify the main techniques of text steganography, which are feature-based, and word-rule based, to recognize the various techniques used with them. As a result, the primary technique used in the text steganography was feature-based technique due to its simplicity and secured. Meanwhile, the common parameter metrics utilized in text steganography were security, capacity, robustness, and embedding time. Future efforts are suggested to focus on the methods used in text steganography.

Corresponding Author:

Farah Qasim Ahmed Alyousuf,
Information Technology Department,
Lebanese French University, Erbil, Iraq.
Email: frhalyousuf@gmail.com

1. INTRODUCTION

Nowadays, steganography becomes the most critical approach used to secure the data. The word steganography means, hide the secret data like the text or digital format. It aims to conceal the secret data ultimately between two parties and make it not visual to the third party and without any suspicions about the existing of any hidden information. There are some types of steganography have been divided into two mediums, which are digital steganography and natural language steganography. Digital steganography is the art that deals with the digital medium, for example, image, video, and audio, while natural language steganography deals with the text files. Even though digital steganography has the main considerations by the researchers, however, the text is the most critical data that needs to be secured because most of the documentation will be in the text such as sending critical information or assigning urgent appointments [1]. Additionally, natural language steganography is classified into two types, which are linguistic steganography and text steganography. Linguistic steganography deals with text (a secret message) that will be hidden in a text medium [2]. Meanwhile, text steganography changing the format of the text or a specific character without changing the meaning of the sentences [3, 4]. Hiding the data by using natural language steganography needs some techniques. Each type has its techniques used by the researchers in text steganography, which are, word-rule based and feature-based technique [5]. Meanwhile, there are five techniques used in linguistic steganography such as, synonym substitution, syntactic substitution, semantic substitution, PCFG, and hybrid technique.

Word-rule based known as the technique that embeds the secret message by shifting the text horizontally or vertically, and it is consist of two categories, which are line-shift coding and word-shift

coding [6]. Meanwhile, feature-based defined as the technique that changes the feature of the characters based on a code word. Sometimes it is slightly shifted up and down, or changing the length of the characters to embed the message in the text data [7]. Therefore, this paper presents a review on feature-based and word-rule based techniques by analyzing the advantages and drawbacks for each technique. Besides, the performance metric for text steganography technique will also be highlighted.

2. PERFORMANCE OF TECHNIQUES IN TEXT STEGANOGRAPHY

There are many techniques used in text steganography that contribute to increase the performance of embedding the hidden message. Thus, this section will study the performance of these techniques based on feature-based and word-rule based techniques.

2.1. Feature-based technique

The feature-based technique works with the shape or the format of the text, for example, changing the size or changing the font type. This technique can make reader assumes that there are no changes in the text, so that reader can't recognize the secret message hidden in the cover text. Based on the current studies, there are two types of feature-based techniques which are language-based (it concentrates on the language used), and letter-based (it can be implemented in any language that used A-Z letters) [8].

Tables 1 and 2 have shown two techniques utilized in feature-based technique. By using language-based, the robustness of the text steganography algorithm will be increased. However, if there was an enhancement in the capacity performance, the security strength will be decreased. Since there is a relationship performance between the quality and the security of text steganography algorithms, the increment in the capacity performance will also affect the quality performance of the cover text, and it will be easy to be attacked. Meanwhile, by using the letter-based technique, the scheme will have a high capacity and a robustness performance scheme.

Table 1. Review on feature-based technique used (language-based)

Techniques used	Proposed methods/technique	Advantages	Disadvantages	Review
English-Based	Encrypt the text by DES then embed by counting the equivalent position for each secret message and cover text [9]. The characters for the secret message is converted to its binary, then it will replaced by the ASCII characters [10]. two-letter word: regarding to the location of two letters, the secret message will be embedded [11]. SEFS method: upper-case letter, punctuations and white space [12].	High security [9]. Robust and High capacity because of hiding Eight bits in one character [10]. High capacity and robust [11]. Better capacity, and robust [12].	Low capacity because the interactivity is very simple [9]. Security should be increased [10, 11].	By reviewing the previous researches, it was noticed that by increasing the capacity, the security will be decreased because the quality will be low.
Arabic-Based	Diacritic will be presented if the secret bit is 1, else if the secret bit is 0, the diacritic will be removed [13, 14]. Used the isolated letters to hide the secret message [15]. Looking for the not connected letters to hide the secret message [16]. Using HARAKAT to hide the text in the reverse Fatha [17].	Better capacity and the method is easy to be implemented [13, 14]. High capacity and robust [15]. Higher capacity than [15], because of high ratio for the characters, also the proposed method has a robust scheme [16]. High capacity, robust [17].	Weak robust [13, 14]. Losing the information in case of retyping [17].	
Chinese-Based	Multi-keywords carrier-free text steganography method based on part of speech tagging has been proposed [18]. Coverless plain text steganography based on characters' features has been proposed to find the common features of the characters to be represented in binary (0 and 1) [19].	High capacity [18, 19].	Security should be improved by applying encryption [18]. By generating a long sentences, there will be a poor readability (low security) [19].	

Table 2. Review on feature-based technique used (letter-based)

Techniques used	Proposed methods/technique	Advantages	Disadvantages	Review
SEFT [20]	Changing the font for the selected character. It needs three characters from the cover text to hide one character from the secret message.	High capacity and robust [20].	Some fonts take large size when replace it with their similarity [20].	
XORSTEG [21]	Work with the binary of the character and XOR the first and the last word in the same paragraph.	High capacity [21].		
CALP, CURVE, VERT [5]	Comparing between CAPL, CURVE, and VERT techniques in term of capacity and embedding time.	CALP has the highest hiding capacity. Also, the higher size of the secret text, the longer embedding time.	VERT has the lowest capacity.	The capacity is increased by using letter-based method.
Polish text steganography [22]	Hiding "0" bit in the pointed letter and "1" bit in the un-pointed letter.	Robust, secured, reliable.	Capacity has not been considered by the author.	
Hypertext markup [23]	Using HTML tags and attributes to hide the secret message.	High capacity.	Time complexity should be decreased.	

2.2. Word-rule based

The implementation is based on hiding the characters horizontally or vertically on word-rule based techniques. It has two techniques which are line-shift coding (which conceals the secret message vertically) and word-shifting technique (which hides the secret message horizontally in the text). As shown in Table 3, the performance for embedding processing will be high as long as the capacity performance is low. Moreover, it is noticed that there is an inverse relationship performance between security and embedding capacity.

Table 3. Review on word-rule based technique used

Technique used	Proposed methods/technique	Advantages	Disadvantages	Review
Line-shift code [3]	Converting the secret message into ASCII value then converted to binary. Convert the binary bit with the number of shifted rows. Random binary bit will be added if the binary bit is less than the number of shifted rows.	High robustness and high capacity	Low performance	
Combination (Line-shift + word-shift) [24]	The secret message converted to ASCII code then to binary. Then, the bits are stored in blocks of 2 bits to embed them into the over file.	High capacity	Weak robustness, because by deleting a space, the hidden data will be damaged	The capacity is increased by reducing the performance of the embedding technique.
Word-shift [25-27]	Hide the secret message 0 in the unchanged space, and 1 bit in the changed space between the neighbors.	Can be used with PDF documents		
	Changing the document by shifting the location of the word horizontally in the same line.	High security	Low capacity	
	Encrypt the message using AES then convert the text to binary, after that embed the bits into the white space.	High security	Time consuming and low capacity	

3. PERFORMANCE METRIC ON TEXT STEGANOGRAPHY

The performance metrics for the text steganography are assisted in evaluating the performance in embedding processing, which are embedding time, capacity, imperceptibility, robustness, security, availability, confidentiality, and integrity. Table 3 has elaborated the performance metric used in text steganography that evaluates the embedding performance of text steganography techniques. Each metric has its usefulness that helps to ensure the quality of the technique. Moreover, Table 4 has mentioned the techniques used for text steganography by using different performance metrics.

Figure 1 has illustrated the percentage of the performance metrics of text steganography which are capacity, imperceptibility, robustness, security, availability, confidentiality, integrity, and embedding time. The main parameters evaluated were capacity at 29% used, followed by security at 24% used. Besides, robustness and embedding time are followed with 16% used and 14% used respectively. Meanwhile, the performance metrics such as confidentiality, availability, and integrity have the minimum percentage that used by researchers at 4%, 3% and 3% respectively.

Table 4. Performance metric in text steganography used by researchers

Parameter Metric	Resource	Purpose
Embedding time	[3, 5, 23, 26-32]	Calculates the time of the technique that the embedding process consumes.
Hiding capacity	[3, 6, 11, 17, 18, 20, 22, 25, 27, 30, 33-43]	It is the maximum number of bits that could be embedded in the cover medium.
Imperceptibility	[11, 17, 37, 44]	Hiding the message in a way that the human vision cannot perceive the difference in the stego medium.
Robustness	[2, 3, 12, 15, 16, 22, 24, 38, 45-48]	It measures how is the technique is difficult to be destroyed by the attacks or the manipulating processes.
Security	[2, 6, 8, 10, 11, 13-15, 18, 21-23, 26, 29, 40, 47-49]	It protects the secret message from the internal and external attacks.
Availability	[18, 26]	It is the ability of the authorized person to get access to the hidden data.
Confidentiality	[27, 34, 45]	It is a measurement to ensure that the secret data will not be available to unauthorized people.
Integrity	[31, 34]	It ensures that the data should not be altered by unauthorized people during the transmission.

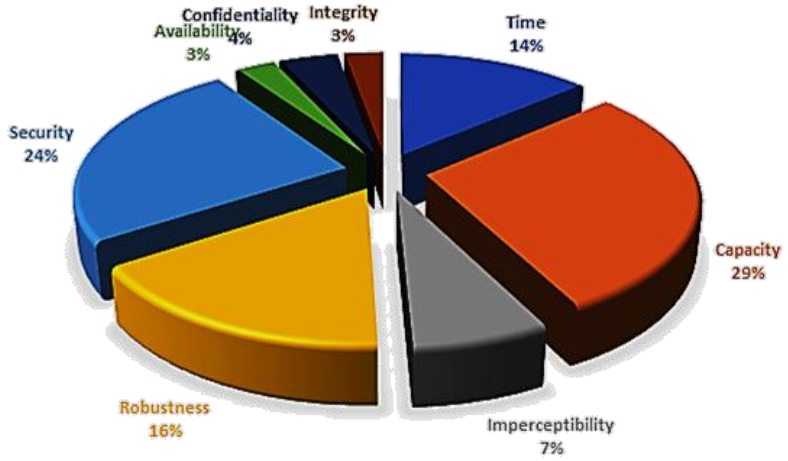

Figure 1. Percentage of performance metric used

4. DISSCUSSION AND CONCLUSION

This paper has presented an analysis review of text steganography techniques. Text steganography has two main techniques, which are feature-based and word-rule based. Each technique has some enhanced technique used in order to improve the embedding performance for text steganography. It is focused that the main technique used among the researchers was feature-based at 76% due to its simplicity system in implementation and its security strength, as shown in Figure 2. Based on Figure 3, there are common parameter metrics used to evaluate text steganography, which are security performance at 34% used, capacity performance at 24% used, robustness performance at 23% used, and embedding time performance at 19% used.

As a conclusion, it is found that the most method used for text steganography is feature-based. It is because of its security and simplicity in text steganography implementation. As well, this paper has mentioned the performance metrics used in evaluating the feature-based techniques which are security, capacity, robustness, and embedding time.

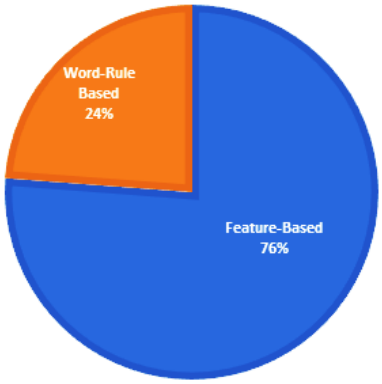

Figure 2. Performance of technique used in feature-based and word-rule based

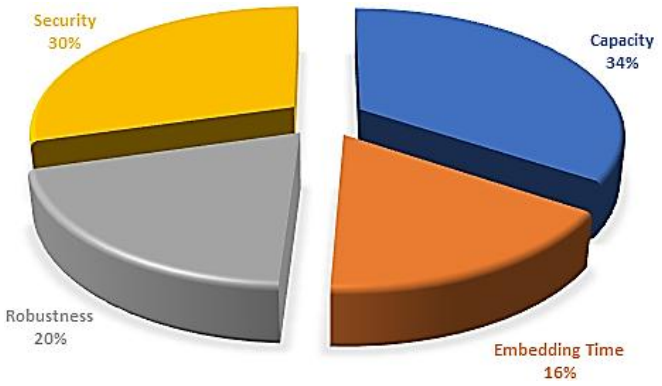

Figure 3. Percentage on top performance metrics used

ACKNOWLEDGEMENTS

We would like to thank Dean, School of Computing, Universiti Utara Malaysia and director of Awang Had Salleh Graduate School (AHSGS) for their moral support to the achievement of this work. A special thanks to Lebanese French University, Kurdistan Region, Iraq, for providing a special support to finalize this work.

REFERENCES

[1] R. Din and S. Utama, "Critical Review of Verification and Validation Process in Feature-Based Method Steganography," in *Int. Conf. E-Commerce*, 2017, pp. 15-19.
[2] S. S. Iyer and K. Lakhtaria, "New robust and secure alphabet pairing Text Steganography Algorithm," *Int. J. Curr. Trends Eng. Res.*, vol. 2, no. 7, pp. 15-21, 2016.
[3] H. T. Ciptaningtyas, R. Anggoro, and M. B. A. Fadhillah, "Text Steganography on Sundanese Script using Improved Line Shift Coding," in *2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)*, 2018, pp. 254-261.
[4] S. Utama, R. Din, and M. Mahmuddin, "The Performance Evaluation of Feature-Based Technique in Text Steganography," *J. Eng. Sci. Technol.*, vol. 12, pp. 169-180, 2017.
[5] R. Din, R. Bakar, S. Utama, J. Jasmis, and S. J. Elias, "The evaluation performance of letter-based technique on text steganography system," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 1, pp. 291-297, 2019.
[6] R. Kumar, S. Chand, and S. Singh, "An Email based high capacity text steganography scheme using combinatorial compression," in *2014 5th International Conference-Confluence The Next Generation Information Technology Summit (Confluence)*, 2014, pp. 336-339.
[7] M. V. Nasab and B. M. Shafiei, "Steganography in programming," *Aust. J. Basic Appl. Sci.*, vol. 5, no. 12, pp. 1496-1499, 2011.
[8] H. K. Tayyeh, M. S. Mahdi, and A. S. Ahmed AL-Jumaili, "Novel steganography scheme using Arabic text features in Holy Quran," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 3, p. 1910, 2019.
[9] M. P. Uddin, M. Saha, S. J. Ferdousi, M. I. Afjal, and M. A. Marjan, "Developing an efficient solution to information hiding through text steganography along with cryptography," *2014 9th Int. Forum Strateg. Technol. IFOST 2014*, 2014, pp. 14-17.
[10] S. Chaudhary, M. Dave, and A. Sanghi, "Text steganography based on feature coding method," in *ACM Int. Conf. Proceeding Ser.*, vol. 12-13-Augu, 2016, pp. 5-8.
[11] S. S. Baawi, M. R. Mokhtar, and R. Sulaiman, "New text steganography technique based on a set of two-letter words," *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 22, pp. 6247-6255, 2017.
[12] S. A. El Rahman, "Text steganography approaches using similarity of English font styles," *Int. J. Softw. Innov.*, vol. 7, no. 3, pp. 29-50, 2019.
[13] M. L. Bensaad and M. B. Yagoubi, "High capacity diacritics-based method for information hiding in Arabic text," in *2011 Int. Conf. Innov. Inf. Technol. IIT 2011*, 2011, pp. 433-436.
[14] M. L. Bensaad and M. B. Yagoubi, "Boosting the Capacity of Diacritics-Based Methods for Information Hiding in Arabic Text," *Arab. J. Sci. Eng.*, vol. 38, no. 8, pp. 2035-2041, 2013.
[15] A. A. Mohamed, "An improved algorithm for information hiding based on features of Arabic text: A Unicode approach," *Egypt. Informatics J.*, vol. 15, no. 2, pp. 79-87, 2014.
[16] A. A. Obeidat, "Arabic text steganography using Unicode of non-joined to right side letters," *J. Comput. Sci.*, vol. 13, no. 6, pp. 184-191, 2017.

[17] J. Memon, K. Khowaja, and H. Kazi, "Evaluation of Steganography for Urdu/Arabic Text," *J. Theor. Appl. Inf. Technol.*, 2015.
[18] Y. Liu, J. Wu, and G. Xin, "Multi-keywords carrier-free text steganography based on part of speech tagging," in *ICNC-FSKD 2017 - 13th Int. Conf. Nat. Comput. Fuzzy Syst. Knowl. Discov.*, 2018, pp. 2102-2107.
[19] K. Wang and Q. Gao, "A Coverless Plain Text Steganography Based on Character Features," *IEEE Access*, vol. 7, pp. 95665-95676, 2019.
[20] W. Bhaya, A. M. Rahma, and D. AL-Nasrawi, "Text steganography based on font type in MS-word documents," *J. Comput. Sci.*, vol. 9, no. 7, pp. 898-904, 2013.
[21] T. Acharjee, A. Konwar, R. K. Ram, R. Sharma, and D. Goswami, "XORSTEG: A new model of text steganography," in *2016 International Conference on Communication and Electronics Systems (ICCES)*, vol. 10, no. 3, 2016, pp. 1-4.
[22] S. Khan, B. Abhijitha, R. Sankineni, and B. Sunil, "Polish text steganography method using letter points and extension," in *Proceedings of 2015 IEEE International Conference on Electrical, Computer and Communication Technologies, ICECCT 2015*, 2015, pp. 1-5.
[23] S. Mahato, D. K. Yadav, and D. A. Khan, "A Modified Approach to Text Steganography Using HyperText Markup Language," in *2013 Third International Conference on Advanced Computing and Communication Technologies (ACCT)*, 2013, pp. 40-44.
[24] S. Roy and M. Manasmita, "A novel approach to format based text steganography," in *Proceedings of the 2011 International Conference on Communication, Computing & Security - ICCCS '11*, 2011, p. 511.
[25] L. Li, L. Huang, X. Zhao, W. Yang, and Z. Chen, "A statistical attack on a kind of word-shift text-steganography," in *Proc. - 2008 4th Int. Conf. Intell. Inf. Hiding Multimed. Signal Process. IIH-MSP 2008*, 2008, pp. 1503-1507.
[26] H. Singh, K. Singh, and K. Saroha, "A Survey on Text Based Steganography," in *Proceedings of the 3rd National Conference; INDIACom-2009*, 2009.
[27] A. Altigani and B. Barry, "A hybrid approach to secure transmitted messages using advanced encryption standard (AES) and Word Shift Coding Protocol," in *Proc. - 2013 Int. Conf. Comput. Electr. Electron. Eng. 'Research Makes a Differ. ICCEEE 2013*, 2013, pp. 134-139.
[28] S. Tyagi, R. Dwivedi, and A. Saxena, "A High Capacity PDF Text Steganography Technique Based on Hashing Using Quadratic Probing," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 3, pp. 192-202, 2019.
[29] R. Din *et al.*, "Evaluating the Feature-Based Technique of Text Steganography Based on Capacity and Time Processing Parameters," *Adv. Sci. Lett.*, vol. 24, no. 10, pp. 7355-7359, Oct. 2018.
[30] S. Utama, R. Din, and M. Mahmuddin, "Critical analysis on steganography technique in text domain," in *Proc. 5th Int. Cryptol. Inf. Secur. Conf. 2016, Cryptol. 2016*, no. December, 2016, pp. 150-157.
[31] R. Din, S. Utama, and A. Mustapha, "Evaluation Review on Effectiveness and Security Performances of Text Steganography Technique," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 11, no. 2, p. 747, Aug. 2018.
[32] R. Din and S. Utama, "Analysis review of feature-based method in term of verification and validation performance," *J. Telecommun. Electron. Comput. Eng.*, vol. 10, no. 2-4, pp. 173-177, 2018.
[33] Z. Yang, Y. Huang, and Y. J. Zhang, "A Fast and Efficient Text Steganalysis Method," *IEEE Signal Process. Lett.*, vol. 26, no. 4, pp. 627-631, 2019.
[34] A. Naharuddin, A. D. Wibawa, and S. Sumpeno, "A High Capacity and Imperceptible Text Steganography Using Binary Digit Mapping on ASCII Characters," in *Proceeding - 2018 Int. Semin. Intell. Technol. Its Appl. ISITIA 2018*, 2019, pp. 287-292.
[35] N. Wu, P. Shang, J. Fan, Z. Yang, W. Ma, and Z. Liu, "Coverless Text Steganography Based on Maximum Variable Bit Embedding Rules," *J. Phys. Conf. Ser.*, vol. 1237, p. 022078, 2019.
[36] A. Darbani, M. M. Alyannezhadi, and M. Forghani, "A New Steganography Method for Embedding Message in JPEG Images," in *2019 IEEE 5th Conf. Knowl. Based Eng. Innov. KBEI 2019*, 2019, pp. 617-621.
[37] D. Majercak, V. Banoci, M. Broda, G. Bugar, and D. Levicky, "Performance evaluation of feature-based steganalysis in steganography," in *Proc. 23rd Int. Conf. RADIOELEKTRONIKA 2013*, 2013, pp. 377-382.
[38] S. S. Iyer, "Practical Evaluation and Comparative Study of Text Steganography Algorithms," *Int. J. Adv. Eng. Res.*, vol. 3, no. 4, 2017.
[39] E. Satir and H. Isik, "A compression-based text steganography method," *J. Syst. Softw.*, vol. 85, no. 10, pp. 2385-2394, Oct. 2012.
[40] R. B. Krishnan, P. K. Thandra, and M. S. Baba, "An overview of text steganography," in *2017 4th Int. Conf. Signal Process. Commun. Networking, ICSCN 2017*, 2017, pp. 0-5.
[41] A. A. A. Gutub and K. A. Alaseri, "Refining Arabic text stego-techniques for shares memorization of counting-based secret sharing," *J. King Saud Univ. - Comput. Inf. Sci.*, 2019.
[42] N. Wu, P. Shang, J. Fan, Z. Yang, W. Ma, and Z. Liu, "Research on Coverless Text Steganography Based on Single Bit Rules," *J. Phys. Conf. Ser.*, vol. 1237, p. 022077, 2019.
[43] N. Alghamdi and L. Berriche, "Capacity investigation of Markov chain-based statistical text steganography: Arabic language case," *ACM Int. Conf. Proceeding Ser.*, pp. 37-43, 2019.
[44] S. S. Baawi, M. R. Mokhtar, and R. Sulaiman, "Enhancement of Text Steganography Technique Using Lempel-Ziv-Welch Algorithm and Two-Letter Word Technique," in *International Conference of Reliable Information and Communication Technology*, vol. 843, Springer International Publishing, 2019, pp. 525-537.
[45] M. Khairullah, "A novel steganography method using transliteration of Bengali text," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 31, no. 3, pp. 348-366, 2019.

[46] L. Lingjun, H. Liusheng, Y. Wei, Z. Xinxin, Y. Zhenshan, and C. Zhili, "Detection of word shift steganography in PDF document," in *Proc. 4th Int. Conf. Secur. Priv. Commun. Networks, Secur.*, 2008, pp. 22-25.
[47] A. Gutub and K. Alaseri, "Hiding Shares of Counting-Based Secret Sharing via Arabic Text Steganography for Personal Usage," *Arab. J. Sci. Eng.*, no. 0123456789, 2019.
[48] S. P. Rajput, K. P. Adhiya, and G. K. Patnaik, "An Efficient Audio Steganography Technique to Hide Text in Audio," in *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, 2017, pp. 1-6.
[49] Y. Liu, T. Yang, and G. Xin, "Text steganography in chat based on emoticons and interjections," *J. Comput. Theor. Nanosci.*, vol. 12, no. 9, pp. 2091-2094, 2015.

BIOGRAPHIES OF AUTHORS

Farah Qasim Ahmed Alyousuf is an assistant lecturer in Lebanese French University in Department of Information Technology, Kurdistan Region, Erbil, Iraq. PhD candidate in Information Technology, School of Computing (SOC), Awang Had Salleh Graduate School (AHSGS), Universiti Utara Malaysia, Sintok, Kedah, Malaysia since January 2019.

Assoc. Prof. Dr. Roshidi Din received his Bachelor of Information Technology and Master of Science in Information Technology degrees from Universiti Utara Malaysia (UUM) in 1996 and 1999 respectively. He later completed his Ph.D from Universiti Sains Malaysia (USM) in 2015. He is currently at the School of Computing, UUM. His current research interests are more on the application of Discrete Mathematics in various areas especially in Information Security, Steganography and Steganalysis, and Natural Language Steganology.