Bulletin of Electrical Engineering and Informatics
Vol. 10, No. 1, February 2021, pp. 493~500
ISSN: 2302-9285, DOI: 10.11591/eei.v10i1.2518

Traid-bit embedding process on Arabic text steganography method

Roshidi Din[1], **Reema Ahmed Thabit**[2], **Nur Izura Udzir**[3], **Sunariya Utama**[4]
[1.4]School of Computing, College Arts and Sciences, Universiti Utara Malaysia, Kedah, Malaysia
[2.3] Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Selangor, Malaysia

Article Info

Article history:

Received Jan 31, 2020
Revised Apr 9, 2020
Accepted May 17, 2020

Keywords:

Arabic text
Embedded performance
Integrated technique
Stegnaography

ABSTRACT

The enormous development in the utilization of the Internet has driven by a continuous improvement in the region of security. The enhancement of the security embedded techniques is applied to save the intellectual property. There are numerous types of security mechanisms. Steganography is the art and science of concealing secret information inside a cover media such as image, audio, video and text, without drawing any suspicion to the eavesdropper. The text is ideal for steganography due to its ubiquity. There are many steganography embedded techniques used Arabic language to embed the hidden message in the cover text. Kashida, Shifting Point and Sharp-edges are the three Arabic steganography embedded techniques with high capacity. However, these three techniques have lack of performance to embed the hidden message into the cover text. This paper present about traid-bit method by integrating these three Arabic text steganography embedded techniques. It is an effective way to evaluate many embedded techniques at the same time, and introduced one solution for various cases.

Corresponding Author:

Roshidi Din,
School of Computing, UUM College Arts and Sciences,
Universiti Utara Malaysia,
06010, Sintok, Kedah, Malaysia.
Email: roshidi@uum.edu.my

1. INTRODUCTION

Nowaday, our communications are hardly private. Intruders in communication and technology can get information in a readable and understandable form of system. The information disclosed by Intruders to others might be used to lunch attack or altered against the person or organizations [1]. The unauthorized people are increasingly interested in other people's conversations, especially with the Internet being an open system. Hence, added counter measures must be taken to ensure privacy rights. One of the solutions used to tackle this problem is by using steganography method.

Steganography is the art and science that deals with hiding secret messages in order to protect the existence of the messages being detected by human senses. It is also a sub-discipline of information hiding. It is often mistaken for Cryptography, even though both are used to protect valuable information [2]. The difference between them is that steganography is the study of hiding information to conceal its existence, while Cryptography is the art of cryptogram or secret writing, involving various methods or embedded technique to ensure the protection of message contents [3]. This depicts that the use of steganography makes anyone who is looking at the object storing secret information not to expect the existence of a message in the object, therefore dismissing thoughts of decrypting the object [4]. Steganography consists of two main focus categories; the technical steganography and linguistic steganography [5]. Technical steganography is a method of hiding information in digitally invisible codes and other media such as image, video, and audio.

Meanwhile, linguistic steganography, also known as natural language steganography, is the art of using the natural language to conceal a secret message that is divided to be two main focuses; linguistic steganography and text steganography. In addition, linguistic steganography entails encoding messages based on order linguistically altered or modified cover document. Meanwhile, text steganography is based on manipulating lines, characters, spaces, or other features of the message. Text steganography is said to be the most challenging approach in steganography. This is due to the fact that text file, as the cover media, normally possesses redundant data for hiding information in a very small quantity [6, 7]. Thus, a method from text steganography category, is the main focus of the study. Text steganography methods are mostly applied to English texts. However, a few text steganography methods are applied to other languages [8]. Only a few researches have been conducted on information hiding in Arabic texts [9]. Therefore, the focus of this study will concentrate on Arabic text steganography and its method developed by researchers.

2. RESEARCH METHOD

A few studies have been conducted on the hiding of information in the Arabic texts. The following is a list of various methods used by studies in utilizing the Arabic text, and reported as shown in Figure 1. In the earlier approach for steganography of information in Persian and Arabic texts, data are hidden in the Arabic texts by displacing the points of Arabic letter vertically as shown in Figure 2 [10]. A challenging problem for algorithms is the retyping process which removes all the hidden data. Another suggested technique [11] is in text steganography for Arabic multipoint letters. The algorithm deals with two (2) bits for every multipoint alphabet. Their embedded technique was combined with vertical point shifting in order to improve the amount or quantity of the hidden data. Their study proposed a solution to this challenge to reduce any new font format changes. This was achieved by unifying all the data leading to a homogenous file.

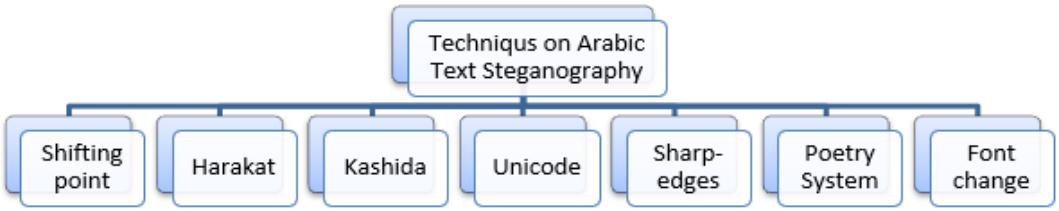

Figure 1. Embedded techniques on Arabic text steganography

The proposed study [12] has presented an approach that makes use of eight varying diacritical symbols in Arabic for the purpose of hiding binary bits in the original cover medium. The embedded information is extracted and revealed by reading the diacritics from the text, and converting it back to a binary representation. Arabic texts which are fully diacritized are utilized as cover media. Another work [13] has designed an Arabic text steganography embedded technique by means of utilizing two (2) Diacritics (Kasrah and Fathah) for the purpose of information hiding as a novel version refining the single (only Fathah) Diacritic approach used before. This use of diacritics (Harakat) for security purposes is conceived as an important approach for text written originally with Diacritics such as historical books or religious Arabic. The study [14] proposed an algorithm for text steganography in the Arabic language. This proposed technique, the extension (Kashida in Arabic) was added to words to represent secret bit 'one', and is not added to words without the extension (Kashida) to represent the secret bit 'zero'. It is worth noting that alphabet extensions have no effect on the writing content or the message content. The main drawbacks of this approach are it captures the reader's attention; increase file size, and changes the text apparent look.

The study utilized the isolated alphabets as hidden keys in the Arabic texts which is written in the Unicode format [15]. The concealing program searches for these alphabets in the word are presented in the carrier text. Hence, data can be hidden in the carrier text by using the isolated alphabets without being noticed in the target word. To simplify the algorithm's complexity, consideration is given to the isolated letters at the beginning of the words, and also at the end of the words even though this is not applicable to all the isolated letters in the words. Another technique which takes advantage of the uniqueness of the Arabic characters is the possession of numerous sharp-edges [16]. The varying number of sharp-edges represents the possibility to conceal the bit 1 and 0 (as secret bits). Alphabets that are having one (1) sharp-edge can conceal the secret bit using two (2) conditions that are hiding bit 0 or 1 at the position of sharp-edges. Meanwhile, if the sharp-edges are two (2), the option for the position of secret bits is in four conditions; 11, 10, 00 or 01.

This approach still delivers the maximum capacity of concealed information, even though not every alphabet in the cover-text is fully utilized. Besides, the study that proposed this technique is used in the Arabic poetry system in hiding secret bits [17].

Since there is a representation of binary units embedded in each Arabic poem, it can be utilized as a means of hiding secret bits. For the purpose of increasing the capacity of the proposed embedded technique, Diacritics and Kashida approaches have been utilized in some cases. The study included a redundant embedded technique only in cases where the secret bit is contrast to the binary bit of the corresponding alphabet in the cover poem. Also, the study [18] conceals information by preparing and arranging the hidden message using font changing feature of end dots that appear in sentences of a word file. The Persian alphabets in a word file were examined to check the font letters which resulted in the selection of alphabets that are not connected to other alphabets that come before and after the character. This results in a contrast in the hidden information, because the demarcation of alphabets can be noticed in text.

3. DEVELOPMENT OF TRIAD-BIT METHOD

This paper proposed a novel method called triad-bit method. Triad-bit method integrates three steganography embedded techniques that use the Arabic text as a cover text to embed the hidden message. The three chosen steganography embedded techniques are Kashida, Shifting Point and Sharp-edges. In the Kashida (extension), the technique uses the extension that is added to words to hide the secret bit 'one', and not added to words to hide the secret bit 'zero'. In shifting point technique the displacing of point vertically or horizontally represent the secret bit 'one' and the not displacing of point use to represent the secret bit 'zero'. In the sharp-edges technique, alphabets that are having one sharp-edge can conceal the secret bit using two conditions that is hiding bit 'zero' or 'one' at the position of sharp-edges. These techniques are selected because they have high capacity with lack performance of embedded process. The triad-bit method evaluates the three steganography techniques in order to overcome the lack performance of embedding process. The embedding process is used by the sender to embed the hidden message into the cover text. This process is divided into three stages; preparation required inputs, selection stego key and Triad-bit embedding process. The output of the embedding process is the stego text which will be used by the receiver to retrieve the hidden message. There are three stages of embedding process in the implementation model of traid-bit methods as shown Figure 2.

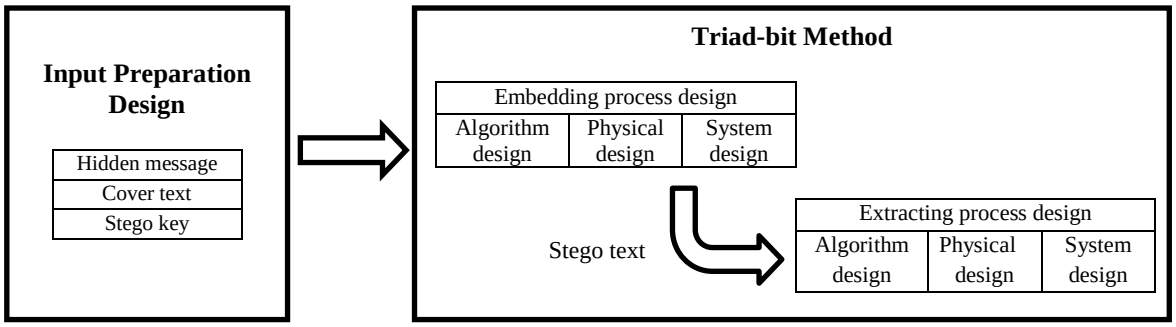

Figure 2. Embedding flow process of triad-bit method

3.1. Preparation required inputs

The sender needs to determine and prepare the required inputs by loading the hidden message from a file or typing it directly. Then, it converts the hidden message from a string data type to binary code as a bit stream to be able to encode into the cover text. After that, the sender will load the cover text file or type it directly. The cover text will be an Arabic text and the method will utilize the characteristics of Arabic text to hide the bit stream.

3.2. Selection stego key

The second stage; the selected embedding algorithm will be determined at this stage by choosing one stego key from the three stego keys. The Triad-bit method involves three stego keys (single-bit1, single-bit2 and stream-bit) based on the three Arabic text steganography embedded techniques (Kashida, Shifting Point, and Sharp-edges). The selection of stego key is an important part to determine the flow of the embedding process. The method will be terminated in case there is no selection of the stego key.

3.3. Triad-bit embedding process

This process is embedding the secret bit stream by integrating the three Arabic steganography embedded techniques (Kashida, Shifting Point, and Sharp-edges). There are two embedding flow processes based on the existence of the secret key. The first flow is Sharp-edges embedding process in which the secret key is compulsory and the binary code of a converted hidden message can hide a bit stream such as '0' or '01' or '1011' for a character. Meanwhile, the second flow is a single-bit embedding process in which the secret key is not required and the binary code of a converted hidden message can hide a single bit such as '0' or '1'.

3.4. The metric evaluation in triad-bit embedding process

The three techniques are integrated in one implementation system. It is using the same cover text in each technique to hide the hidden message. The cover text that embeds a hidden message is evaluated by using the embedding process of triad-bit method based on five metrics evaluation performances. The first metric performance is to evaluate the capability of stego text by comparing the size of stego text and the size of expected text. Then, the second metric usage ratio is to evaluate the used character during the embedding process. The capacity percentage for used character is using the following equation:

$$Usage\ Ratio = \left[\frac{Total\ of\ used\ character}{Total\ character\ of\ cover\ text}\right] x\ 100\%$$

$$= \frac{\sum_{i=1}^{1<m<n} a_1}{\sum_{i=1}^{1<m<1n} b_1}\ x\ 100\%$$

After that, the third metric performance is usability ratio that is referring to how many characters can be used to embed the hidden message. The capacity percentage for usable character is calculated using the following equation:

$$Usability\ Ratio = \left[\frac{Total\ of\ usabale\ character}{Total\ character\ of\ cover\ text}\right] x\ 100\%$$

$$= \frac{\sum_{i=1}^{i<m<n} a_1}{\sum_{i=1}^{1<m<1n} b_1}\ x\ 100\%$$

The fourth metric performance is to evaluate the fitness performance of triad-bit method. Finally, metric performance is to evaluate the running time of embedding process of triad-bit method. By using these metric performances, this study tries to compare three chosen techniques by looking at which technique is having the highest percentage in the embedding process.

4. EXPERIMENTAL RESULT

In this paper, the prototype namely traid-bit steganography system (TBSS) has been developed by C# programming language through Visual Studio 2013 IDE as a tool in order to evaluate this technique based on several aspects such as capability performance, usage ratio, usability, fitness performance and running time performance.

4.1. Capability of stego text size

The capability of stego text size is used to compare the size of stego text and the expected stego text. This analysis is very important for steganographer to understand the capability of a stego text that is produced by the system. The result from this experiment is illustrated in Figure 3. It is clearly shows that size comparison between the stego text system and the expected stego text in Traid-bit method. Even though both texts have a steady increase, the stego text overruled the expected stego text, making it reach a higher level than the latter.

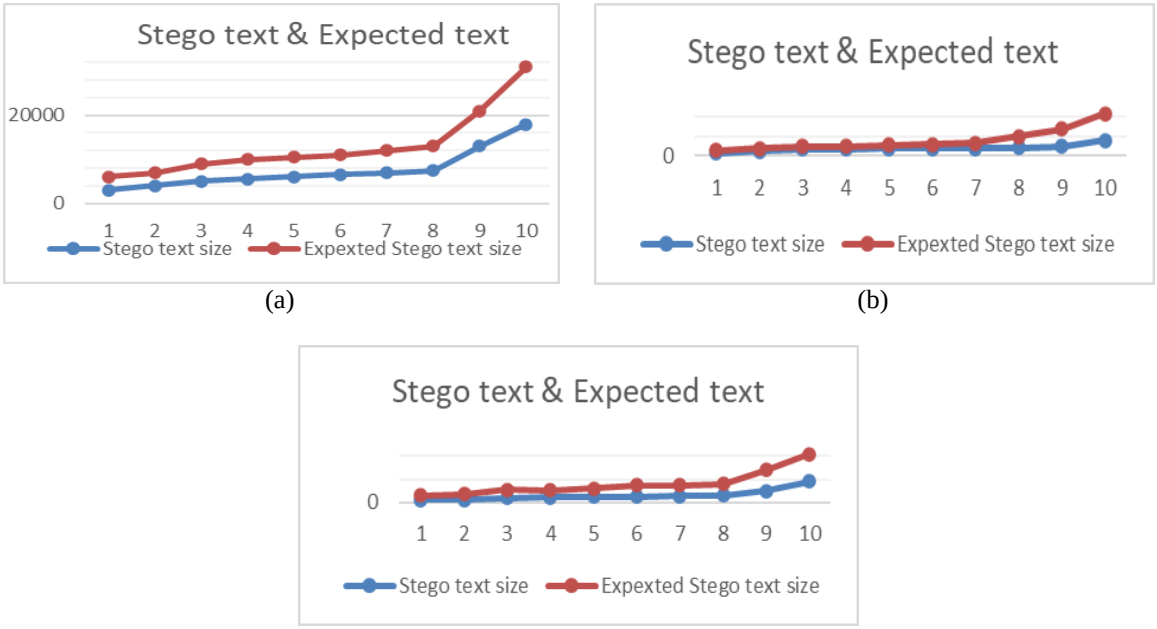

Figure 3. Capability of stego text size, (a) Dataset 1, (b) Dataset 2, (c) Dataset 3

4.2. Usage of cover text

Usage ratio is used to determine the total characters of a cover text that can be used to embed the hidden message [19]. This analysis is very important for steganographer to understand the capability of a cover text to hide a hidden message. The results from Kashida, Shifting Point and Sharp-edges technique are illustrated in Figure 4. It is illustrated that the proposed method used less character from the cover text which are 8% by using Kashida technique, 9% by using Shifting Point technique and 4% by using Sharp-edges technique.

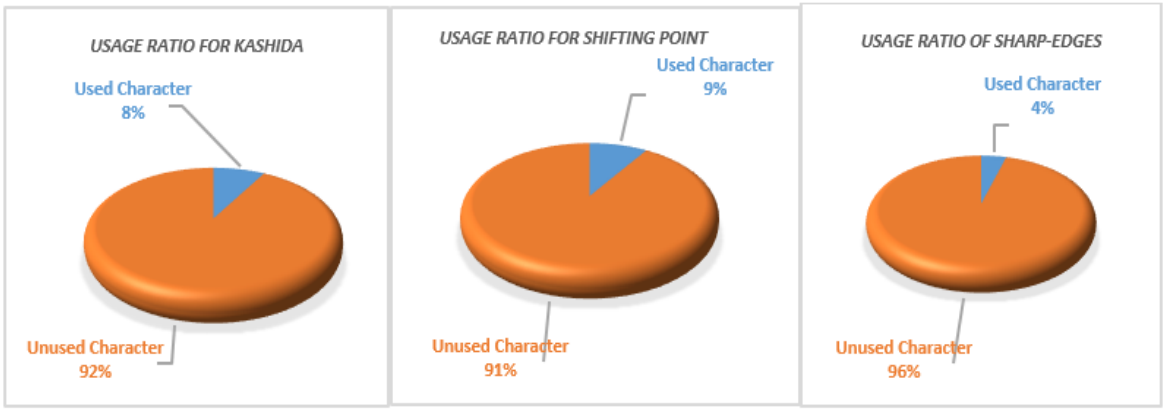

Figure 4. Cover text character usage for data hiding using traid-bit method

4.3. Usability of cover text

The usability ratio is used to determine the total characters of a cover text that are able to be utilized in embedding the process of the hidden message [12, 20]. This analysis is very important for a steganographer to understand the capability of the cover text to hide a hidden message. The results from Kashida, Shifting Point and Sharp-edges technique are illustrated in Figure 5. It is shows that Traid-bit method has high useable character on account of the Sharp-edges technique. It can utilize most of Arabic characters to embed the hidden message.

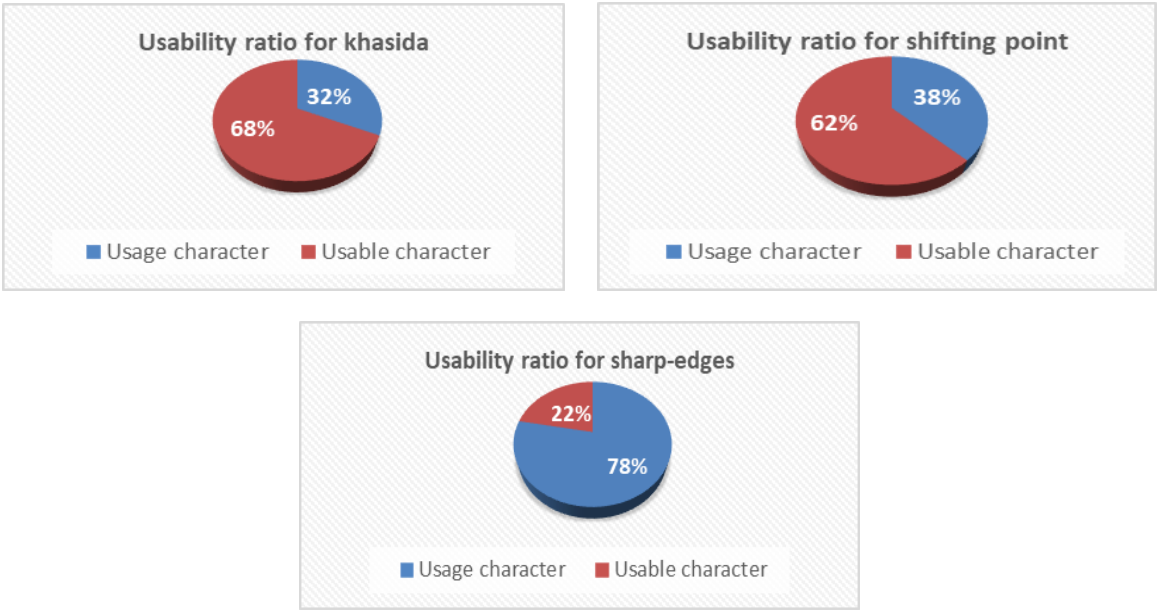

Figure 5. Cover text character usability for data hiding using traid-bit method

4.4. Fitness performance of stego key

A basic step in this stage is to understand the fitness performance of the stego keys used. Therefore, the fitness performance of stego keys used such as Kashida, Shifting Point and Sharp-edges are examined. The metrics performance for stego keys used in this research is adopted from several researchers in steganography works [21-23] such as capacity ratio, embedded fitness ratio and saving space ratio. A capacity ratio of Kashida, Shifting Point and Sharp-edges showed an increment over the whole data set. On the other hand, the embed fitness ratio of single-bit '2' declined slightly compared to the single-bit '1' and Sharp-edges techniques. In comparison, the saving ratio of Shifting Point has a higher ratio than the Shifting Point and the stream-bit

4.5. Running time

Running time is the period of time when the code of the system is executed [24, 25]. It provides the information to steganographer about the embedding performance. In this part the TBSS system calculates the code Traid-bit method. The comparison of running time for Traid-bit method is shown in Figure 6. The running time to embed the hidden message by using the Shifting Point technique is very high and not acceptable compared to Kashida technique. Similarly, the Sharp-edges running time to embed the hidden message is low regardless of the font type and the extension inserted.

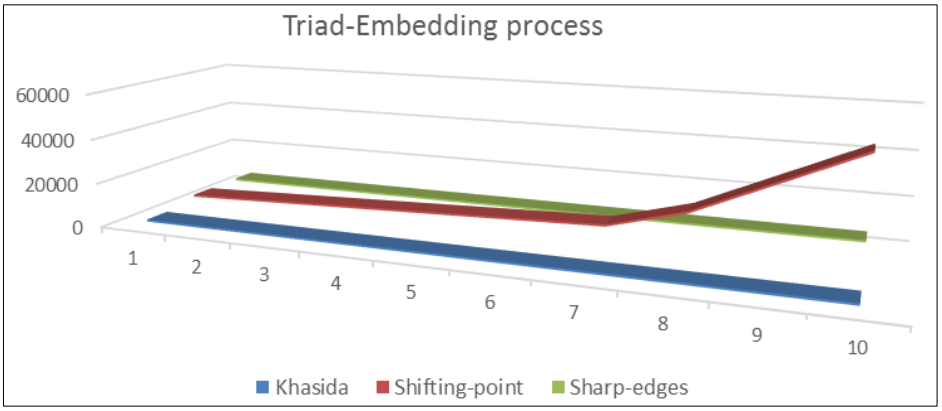

Figure 6. Running time of traid-bit method

5. CONCLUSION

This paper present about the development of an integrated embedded technique called Triad-bit method which integrated the Kashida, Shifting Point and Sharp-edges. This integrated embedded technique enables the steganographer to utilize the Arabic steganography method that is capable to analyze and evaluate the embedding performance. The experiment for embedding performance is done through several parameters such as size capability, usage ratio, usability ratio, fitness performance and running time. The study stated that the size capability of proposed method is improved because the system stego text size is lower than the expected stego text size. In addition, the proposed method used less character from the cover text such as 8% by using Single-bit1 technique, 9% by using Shifting Point technique and 4% by using Sharp-edges technique. The study found that the usability ratio of the proposed method is high especially by using Sharp-edges technique (78%). The running time for the proposed method is the embedding time based on the used embedded technique. Therefore, the proposed method provides an accurate analysis and evaluation for embedding performance in Arabic text steganography embedded technique.

ACKNOWLEDGEMENTS

We would like thank to members of School of Computing, Universiti Utara Malaysia and Universiti Utara Malaysia for their moral support for the realization of this work.

REFERENCES

[1] S. Bhattacharyya, I. Banerjee, and G. Sanyal, "A survey of steganography and steganalysis technique in image, text, audio and video as cover carrier," *Journal of Global Research in Computer Science*, vol. 2, no. 4, pp. 1-16, 2011.
[2] R. Din and A. J. Qasim, "Steganography analysis techniques applied to audio and image files," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 4, pp. 1297-1302, 2019.
[3] M. Kumar and A. J. Singh, "Understanding steganography over cryptography and various steganography techniques," *International Journal of Computer Science and Mobile Computing*, vol. 4, no. 3, pp. 253-258, 2015.
[4] B. Osman, A. Yasin and M. N. Omar, "Hiding message in text steganography using RGB color in random location," *TEST Engineering & Management*, vol. 81, pp. 6030-6037, 2019.
[5] S. Malik and W. Mitra, "Hiding information - A survey," *Journal of Information Sciences and Computing Technologies*, vol. 3, no. 3, pp. 232-240, 2015.
[6] J. A. Memon, K. Khowaja, and H. Kazi, "Evaluation of steganography for Urdu/Arabic text. pace pacing and clinical electrophysiology," *J. of Theoretical and Applied Information Technology*, vol. 4, no. 3, pp. 232-237, 2008.
[7] R. Din, Rosmadi 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.
[8] M. V. Nasab and B. M. Shafiei. "Steganography in programming," *Australian Journal of Basic and Applied Science*s, vol. 5, no. 1, pp. 1496-1499, 2011.
[9] M. Khairullah, "A novel text steganography system using font color of the invisible characters in Microsoft word documents," *2009 Second International Conference on Computer and Electrical Engineering*, pp. 482-484, 2009.
[10] A. Odeh, A. Alzubi, Q. B. Hani, and K. Elleithy, "Steganography by multipoint Arabic letters," *2012 IEEE Long Island Systems, Applications and Technology Conference (LISAT)*, pp. 1-7, 2012.
[11] M. H. Shirali-Shahreza and M. Shirali-Shahreza, "A new approach to Persian/Arabic text steganography," *5th IEEE/ACIS International Conference on Computer and Information Science and 1st IEEE/ACIS International Workshop on Component-Based Software Engineering, Software Architecture and Reuse (ICIS-COMSAR'06)*, pp. 310-315, 2006.
[12] A. Odeh and K. Elleithy, "Steganography in Arabic text using zero width and kashida letters", *International Journal of Computer Science and Information Technology*, vol. 4, no. 3, pp. 1-11, 2012.
[13] M. A. Aabed, S. M. Awaideh, A. M. Elshafei, and A. A. Gutub, "Arabic diacritics based steganography," *2007 IEEE International Conference on Signal Processing and Communications*, pp. 756-759, 2007.
[14] E. M. Ahmadoh and A. A-A. Gutub, "Utilization of two diacritics for Arabic text steganography to enhance performance," *Lecture Notes on Information Theory*, vol. 3, no. 1, pp. 42-47, 2015.
[15] F. Al-Haidari, A. Gutub, K. Al-Kahsah, and J. Hamodi, "Improving security and capacity for Arabic text steganography using 'Kashida' extensions," *2009 IEEE/ACS International Conference on Computer Systems and Applications*, pp. 396-399, 2009.
[16] A. A Mohamed, "An improved algorithm for information hiding based on features of Arabic text: A unicode approach," *Egyptian Informatics Journal*, vol. 15, no. 2, pp. 79-87, 2014.
[17] N. A. Roslan, R. Mahmod, and N. I. Udzir, "Sharp-Edges method in Arabic text steganography," *Journal of Theoretical and Applied Information Technology*, vol. 33, no. 1, pp. 32-141, 2011.
[18] E. A. Khan, "Using Arabic poetry system for steganography," *Asian Journal of Computer Science and Information Technology*, vol. 4, no. 6, pp. 55-61, 2014.
[19] S. B. Ardakani, A. M. Latif, and K. Mirzaie, "Presentation a new method for steganography in Persian text of an electronic document," *Fen Bilimleri Dergisi (CFD)*, vol. 36, no. 4, pp. 700-707, 2015.
[20] B. Osman, A. Yasin, and M. N. Omar "Analysis review of alphabet-based techniques in text steganography," *Journal of Telecommunication, Electronic and Computer Engineering*, vol. 8, no. 10, pp. 109-115, 2016.

[21] P. Singh, R. Chaudhary, and A. Agarwal, "A novel approach of text steganography based on null spaces," *IOSR Journal of Computer Engineering*, vol. 3, no. 4, pp. 11-17, 2012.
[22] L. Y. Por, T. F. Ang, and B. Delina, "WhiteSteg: A new scheme in information hiding using text steganography," *WSEAS Transactions on Computers*, vol. 7, no. 6, pp. 735-745, 2008.
[23] B. Osman, A. Yasin and M. N. Omar, "A new scheme of representing a secret message using second quotient remainder theorem in text steganography", *International Journal of Engineering & Technology* vol. 7, no. 4.19, pp. 83-88, 2018.
[24] B. Osman, R. Din, and M. R. Idrus, "Capacity Performance of Steganography Method in Text Based Domain," *ARPN Journal of Engineering and Applied Sciences*, vol. 10, no. 3, pp. 1345-1351, 2015.
[25] S. S. Iyer and K. Lakhtaria, "Practical evaluation and comparative study of text steganography algorithms," *International Journal of Advance Engineering and Research Development*, vol. 3, no. 4, pp.277-283, 2016.

BIOGRAPHIES OF AUTHORS

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.

Reema Ahmed Thabit received her Bachelor of Computer Science & Engineering from Aden University in 2004. She completed the Master of Science in Information Technology by mix mode in 2017 from Universiti Utara Malaysia (UUM). Currently, she is a PhD student UPM, Faculty of Computer Science and Information Technology at Universiti Putra Malaysia (UPM).

Nur Izura Udzir currently works as an Associate Professor at Universiti Putra Malaysia. Department of Computer Science, BSK (UPM), M.Sc. (UPM), Ph. D (York) in teaching Computer Programming Operating System (since 2007 to date) (undergraduate). Her research interests in Access Control, Intrusion Detection Systems, Computer Security, Coordination in Distributed Systems, Computer security, Secure operating systems, Access control, Distributed systems, Intrusion detection systems, Capability-based coordination in open distributed systems and Neural Network for information.

Sunariya Utama received his Bachelor of Information Technology from Universiti Utara Malaysia (UUM) in 2012. He completed the Master of Science in Information Technology by research in 2017 from UUM. Currently, he is a PhD student in Information Technology, School of Computing (SOC), Awang Had Salleh Graduate School (AHSGS), Universiti Utara Malaysia, Sintok, Kedah, Malaysia since December 2017.