Modern Education
and Computer Science
PRESS

A State-of-the-Art Survey of Coverless Text Information Hiding

Shahbaz Ali
School of Computer Science, Shaanxi Normal University, Xi'an, China
E-mail: ali.shahbaz@snnu.edu.cn
ORCID ID: 0000-0002-3989-1811

Abstract—Information plays a pre-eminent role in people's routine lives. It provides people with facts about abundant topics of interest. Information can be represented by a variety of communicative media such as text, video, audio, and image, out of which text is the most common medium for representing the information. In the digital era, the information can easily be imitated, exchanged and distributed from one place to another in an instant. Thus, it is incredibly essential to hide the confidential information so that it couldn't be accessed by unauthorized means. The traditional information hiding techniques require a designated carrier to hide the secret information, which ultimately introduces some modifications in the carrier. As a result, it is quite hard for the existing traditional methods to escape from the steganalysis. In contrast to conventional information hiding techniques, the term 'coverless information hiding' has been coined by the researchers lately, which doesn't require a designated carrier to conceal the secret information. Hence, the technique of coverless information hiding can efficiently resist the attacks of steganalysis. This paper presents a state-of-the-art survey of coverless text information hiding by discussing the current scope of the aforementioned technique comprehensively. The existing coverless text information hiding methods are compared and contrasted by various vital aspects such as embedding capacity, algorithm efficiency, ability to resist steganalysis, and methods' theoretical and real-world significance. Moreover, some future aspects of coverless text information hiding are highlighted at the end.

Index Terms—Coverless Information Hiding, Data Security, Steganography, Steganalysis, Digital Content.

I. INTRODUCTION

The importance of information in one's everyday life cannot be denied. With the help of information, one can know about the whole story of an individual, situation, happening, etc. [1, 2]. That is to say; the information answers a variety of questions. The content of the information can be of several forms such as audio, video, image, and text. Unlike other forms of communicative media, the unique nature of text makes it one of the most convenient ways to represent the information. Moreover, in the computer age, it has become much easier to send out the information from one place of the world to another in a very short time [3]. However, in the digital age, it is also effortless to imitate, exchange, and distribute the information, which introduces digital data confidentiality concerns. As a result, the confidential information can easily be accessed by unauthorized means [4, 5]. For that reason, it is indispensable to conceal the secret information so that it could be transmitted from the sender to the receiver in a secure way without grabbing the attention of cybercriminals.

In order to deal with the problems of digital data privacy, a number of information hiding techniques have been proposed by the researchers in the past to ensure the security of the confidential information [6-10]. For these existing information hiding methods, a designated carrier signal is needed to hide the sensitive information so that the presence of the secret information could be covered up. The selected carrier can be of several types such as image, text, audio, and video [11]. For performing the operation of information hiding, the unnecessary part of an image, text, audio or video carrier is utilized to hide the secret information. Although the existing methods of information hiding somehow cover up the presence of the hidden information, they, by some means, introduce some modifications in the designated carrier, which is a major drawback of traditional information hiding methods. The modified carrier is not an ideal carrier to carry out the confidential information, and it cannot resist all types of steganalysis attacks. Hence, in this way, there is always a chance that the secret information could be accessed or destroyed by cybercriminals [12].

The term 'coverless information hiding' has been coined by the researchers lately to overcome the drawback of the existing traditional information hiding techniques [12, 13]. The word 'coverless' should not be confused with the absence of the carrier signal. What is meant by 'coverless information hiding' here is that it does not require any designated carrier to embed the secret information. In fact, the original carrier signal already contains the confidential information. As a result, this technique of hiding the secret information does not modify the original carrier signal, which makes the

coverless information hiding techniques more robust in resisting the current steganalysis attacks [14, 15]. Although the present coverless information hiding methods are efficient enough to withstand the steganalysis attacks, to get the most out of coverless information hiding, there is a need to improve the algorithm efficiency and embedding capacity.

This paper provides readers with an up-to-the-minute review of coverless text information hiding techniques and discusses the current scope of coverless text information hiding more straightforwardly. The present coverless text information hiding methods are compared and contrasted by various vital aspects such as the efficiency of the coverless information hiding algorithm, secret information embedding capacity, researchers' proposed methods' ability to withstand the steganalysis attacks, and the theoretical and practical significance of the coverless text information hiding techniques proposed by the researchers so far. Additionally, some future aspects of coverless text information hiding are highlighted at the end.

II. ELUCIDATION OF COVERLESS INFORMATION HIDING

Coverless information hiding is one of the hottest topics in computer science research these days. Because of the limitations of traditional information hiding methods, coverless information hiding has grabbed the attention of researchers [16, 17]. Unlike all of the existing ways of information hiding, coverless information hiding technique does not require any designated carrier signal to conceal the presence of the hidden information. The original carrier signal already holds the confidential information. Hence, this keeps the originality of the carrier signal, which was not possible in the traditional information hiding techniques. Moreover, the method of coverless information hiding can resist steganalysis attacks, and this makes coverless information hiding as one of the most suitable techniques to hide and protect the secret information in a way that it could not be accessed in an unauthorized way. The concept of coverless information hiding can be better understood by referring to Figure 1. For instance, the secret information to be transmitted from the sender to the receiver is in the binary form, and its value is 00011000, and the carrier signal is a grayscale image with an intensity value of 24. As 24 is a decimal number and the binary value of 24 is 00011000, hence, in this way, the image carrier already contains the secret information. Moreover, the originality of the image carrier is preserved, which makes it robust against steganalysis attacks.
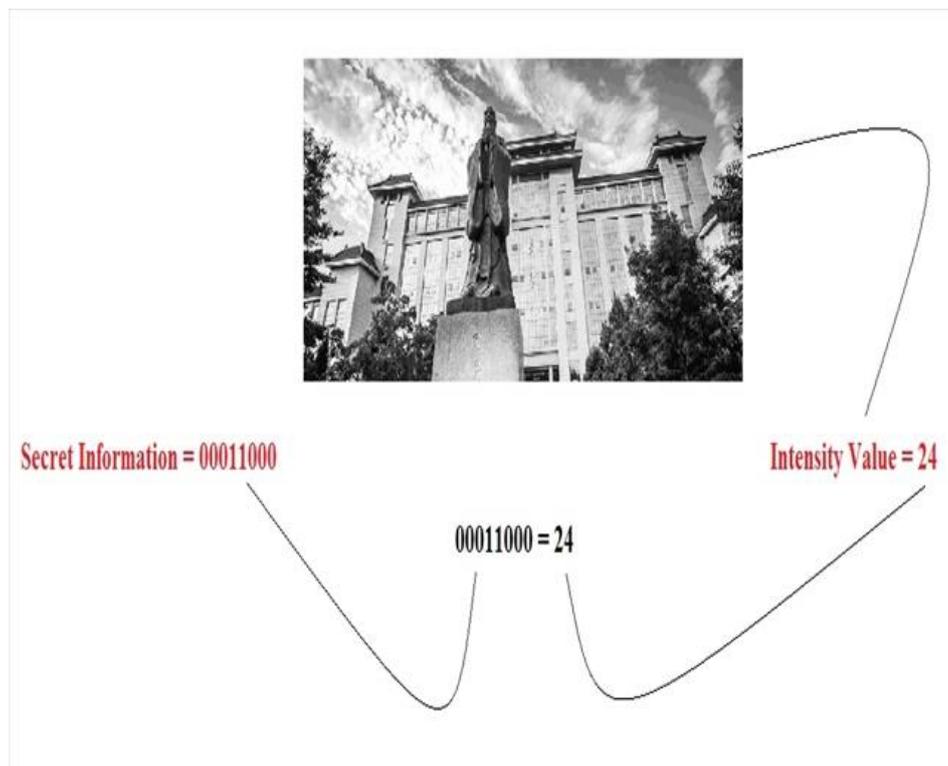

Fig.1. Coverless Information Hiding Demonstration

In spite of the fact that in computer science research, coverless information hiding is a new concept, it is already in practice in people's routine lives. An acrostic poem is one of the best examples, which proves that the idea of coverless information hiding is not a new thing for people. An acrostic is a unique riddle of words, poem or any other composition in which a selection of letters in each line creates a word or words [18]. Figure 2 shows an acrostic poem written by Lewis Carroll. The information which is hidden in this acrostic poem is 'Alice Pleasance Liddell' – person name.

A Boat, Beneath A Sunny Sky

A boat, beneath a sunny sky
Lingering onward dreamily
In an evening of July -

Children three that nestle near,
Eager eye and willing ear,
Pleased a simple tale to hear -

Long has paled that sunny sky:
Echoes fade and memories die:
Autumn frosts have slain July.

Still she haunts me, phantomwise,
Alice moving under skies
Never seen by waking eyes.

Children yet, the tale to hear,
Eager eye and willing ear,
Lovingly shall nestle near.

In a Wonderland they lie,
Dreaming as the days go by,
Dreaming as the summers die:

Ever drifting down the stream -
Lingering in the golden gleam -
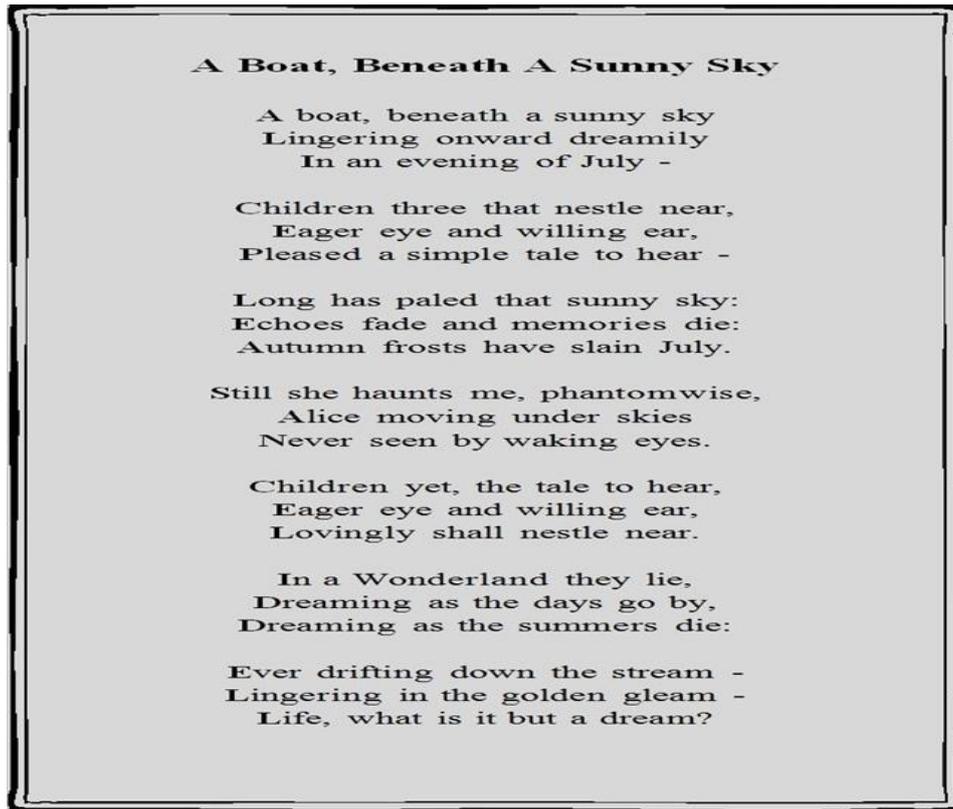Life, what is it but a dream?

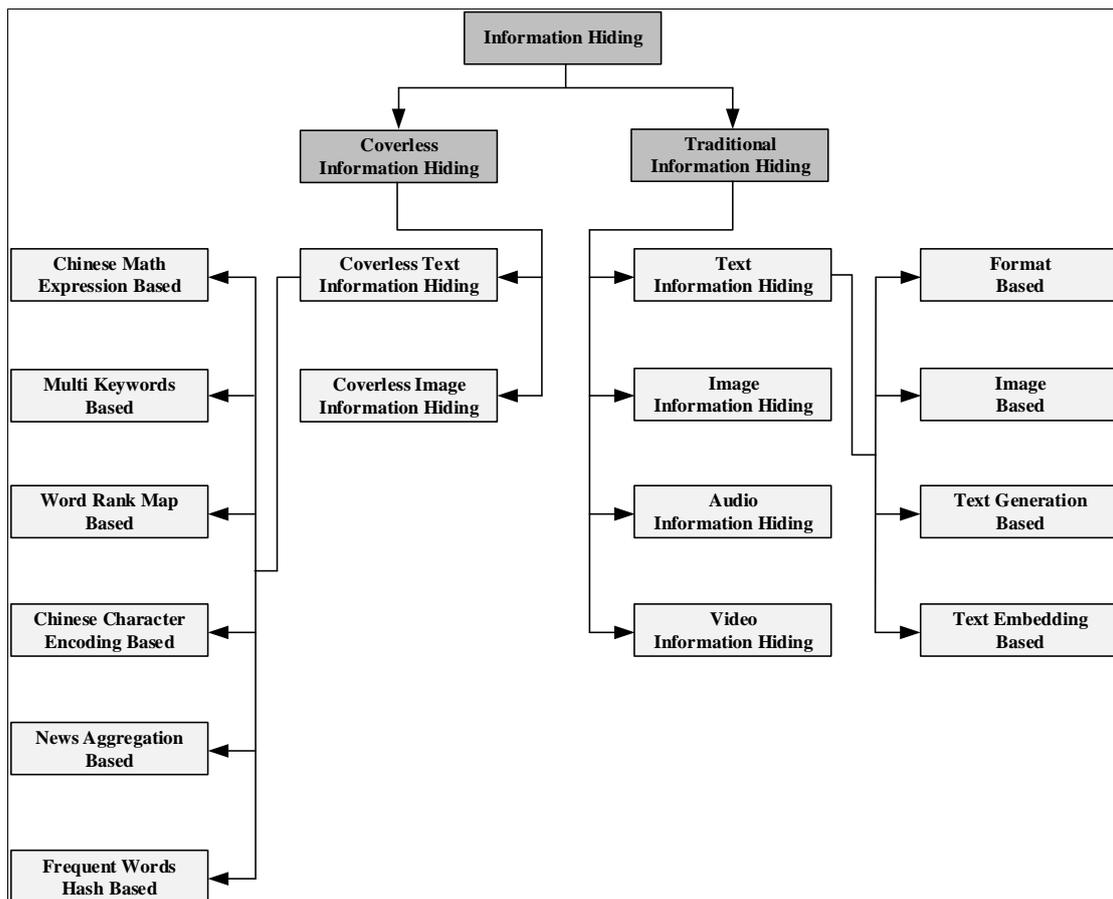Fig.2. An Acrostic Poem by Lewis Carroll

Fig.3. The Nomenclature of Information Hiding

Although any communicative medium can represent the content of secret information, the text is the most common medium to represent the information. The unique nature of text makes it one of the most convenient ways to represent the information [19]. Because of its newness in computer science, coverless information hiding has two main categories so far: coverless text information hiding and coverless image information hiding. Figure 3 shows the nomenclature of information hiding with a focus on text information hiding.

III. Coverless Text Information Hiding

Since the birth of coverless information hiding field, a number of information hiding techniques have been developed lately for both text-based and image-based coverless information hiding. Coverless text information hiding is the subcategory of coverless information hiding in which the secret information is hidden in a text without modifying the carrier signal. Thus, the originality of the carrier signal is preserved, which makes it more robust in resisting the steganalysis attacks. The detailed elucidation of the techniques of coverless text information hiding developed so far is given as:

A. Chinese Math Expression Based Coverless Text Information Hiding

Coverless text information hiding based on the Chinese mathematical expression is the first method in the area of coverless information hiding proposed by Chen et al. [12]. This method utilizes the concept of representing the Chinese characters in the form of a mathematical expression in which the components of the characters of Chinese language are the operands, and the spatial relations between the Chinese characters' components are denoted by the operators [20]. The process of hiding the secret information using the Chinese mathematical expression involves a couple of steps. The first step of the developed method is related to the stego-vector generation in which a stego-vector or encryption vector is generated from the available secret information. After that, a piece of text is obtained from a bigger text database in the second step. The retrieved text is actually a normal text which includes the encryption vector generated from the secret information. The final step is related to the information transmission. The sender can send the normal texts (obtained in the second step) to the receiver, and the receiver can apply the inverse process to retrieve the hidden confidential information from the normal texts sent by the sender. The obtained normal text is normal for both the receiver and other people. Hence, this preserves the originality of the carrier signal.

In spite of the fact that the developed method of information hiding can resist the attacks of steganalysis and preserves the carrier originality, this technique has some limitations. The developed method offers a very low capacity to embed the secret information. The developed technique is able to hide only 1 keyword in a 1-kilobyte file, and the mean capacity offered is only 1. If the keyword's average length is 2, the method is able to hide 10.08 bits, which is relatively low. The total embedding capacity related information of the Chinese mathematical expression based coverless text information hiding (CME-CTIH) is shown in Table 1.

Table 1. Secret Information Embedding Capacity of CME-CTIH

Method	Mean Capacity	Maximum Keywords Hiding Capacity (in one text)	Maximum Information Hiding Capacity
CME-CTIH	1	1 keyword	10.08 bits

In addition, the receiver is unable to know the total number of keywords in each piece of text. Another limitation of this method of coverless text information hiding is that it requires a huge text database in advance, which wastes away the computational resources and degrades the performance.

B. Multi Keywords Based Coverless Text Information Hiding

Zhou et al. developed the technique of coverless text information hiding based on the multiple keywords [13]. This method hides not only the secret information but also the number of keywords in the created text database. The main difference between coverless information hiding based on the Chinese mathematical expression and this method is that the stego-text retrieved by utilizing this technique includes both the confidential information and the number of keywords. As a result, the receiver will have knowledge of the total number of keywords contained in the received stego-text, which was not possible in the Chinese mathematical expression based coverless text information hiding method. Moreover, this technique offers slightly better secret information embedding capacity compared to the previous one. The developed method can hide 1.57 keywords in a 1-kilobyte file. If the keyword's average length is 2, the developed method can hide 15.82 bits of secret information. Table 2 shows the total embedding capacity related information of the multi-keywords based coverless text information hiding (MK-CTIH).

Table 2. Secret Information Embedding Capacity of MK-CTIH

Method	Mean Capacity	Maximum Keywords Hiding Capacity (in one text)	Maximum Information Hiding Capacity
MK-CTIH	1.57	1.57 keywords	15.82 bits

Although the embedding capacity offered by this method is slightly better compared to the previous one, it is not very high. Additionally, one piece of text contains insufficient numbers, which decreases the overall success rate. Furthermore, this technique requires a huge text

database, which ultimately squanders the computational resources.

C. Word Rank Map Based Coverless Text Information Hiding

By using the concept of word rank map, Zhang et al. developed the technique of coverless text information hiding [21]. The word rank map based coverless text information hiding method takes a couple of steps to hide the secret information. The first step is related to the generation of stego-vectors from the confidential information by using the rank map of words, which is calculated by the statistical analysis of the text database. After the stego-vector generation, some normal texts are obtained from the text database in the second step. The normal texts obtained contain the secret information in the form of stego-text. In the final step, the sender sends the normal texts containing the stego-text to the receiver without introducing modifications in the original carrier signal. The secret information receiver can apply the inverse process to retrieve the hidden confidential information.

The developed method is not difficult to implement and can robustly resist the existing steganalysis attacks, but the embedding capacity offered by this method is inadequate. In one piece of English text, this technique can hide only one word. Moreover, this technique like the previously mentioned techniques requires a huge text database, which affects the performance.

D. Chinese Character Encoding Based Coverless Text Information Hiding

The Chinese character encoding based coverless text information hiding method has been proposed by Chen et al. to hide the secret information securely without modifying the carrier signal [22]. This technique of coverless text information hiding creates the binary number system based tags to find the confidential information, and the Chinese characters transform these tags. The receiver then can select some specific tags with the help of an independent secret key to retrieve the hidden information.

The developed technique provides enhanced security, and the success rate of this method is more than 95%. However, there is still need for improving the secret information embedding capacity and the overall performance.

E. News Aggregation Based Coverless Text Information Hiding

Liu et al. proposed the news aggregation based coverless text information hiding method to embed the secret information in the news (available online) [23]. The process of information hiding using this technique involves a few steps. First, the algorithm converts the secret message that is to be sent to the receiver into an integer of large size. Secondly, the integer is hidden into a piece of news, which is available online. The integer hiding process is carried out by shuffling the news headlines order. Finally, the receiver can use the secret key to retrieve the confidential information hidden in the news text sent by the sender.

Table 3. Secret Information Embedding Capacity of NA-CTIH

Method	No. of News Headlines	No. of Division Blocks	Maximum Information Hiding Capacity (in a 1-kilobyte file)
NA-CTIH	20	2	43 bits

This method can resist the present steganalysis attacks in a robust way. Moreover, the introduced method offers a significant capacity to embed the secret information. If the number of division blocks is 2, the developed method can hide 43 bits in a 1-kilobyte file containing 20 news headlines. Table 3 shows the maximum embedding capacity related information of the news aggregation based coverless text information hiding (NA-CTIH).

In addition, the introduced method does not require a huge text database in advance, which saves the computational resources to some extent. However, there is still room to increase the secret information embedding capacity.

F. Frequent Words Hash Based Coverless Text Information Hiding

The technique of coverless text information hiding based on the idea of frequent words hash is another method proposed by Zhang et al. in the field of coverless text information hiding [17, 21]. In this technique, Zhang et al. have used the concept of hashing along with the word rank map to hide the secret information without introducing modifications in the carrier signal. The frequent words hash based coverless text information hiding method takes a couple of steps to hide the confidential information. The first step is related to the creation of a text database. In the second step, the algorithm of this method calculates the frequent words distance and the word rank map. The rank map of the words is calculated by the statistical analysis of the text database created in the first step, while the frequent words distance or frequent words hash is calculated by utilizing the concept of the hamming distance. In the third and final step, the frequent words hash and the word rank map (calculated in the second step) are utilized to retrieve a piece of normal text. At the receiver end, the secret information hidden in the normal text sent by the sender can be extracted by applying the inverse process.

The technique of coverless text information hiding based on the frequent words hash offers enhanced security, and the algorithm of the developed method is simple to implement. However, to further improve the theoretical and real-world significance of the developed technique, there is still need for saving the overall computational resources used by the developed method and improving the capacity to embed the secret information.

IV. Conclusions

Coverless text information hiding is a promising approach to hide the secret information in a text without modifying the carrier signal. Moreover, the techniques of coverless text information hiding can robustly resist the steganalysis attacks. This paper comprehensively surveyed coverless text information hiding, which is one of the hottest topics in computer science research these days. The current scope of the aforementioned technique was discussed in detail. Additionally, the existing methods of coverless text information hiding were compared and contrasted by a number of crucial aspects such as the secret information embedding capacity, the robustness of the methods developed by the researchers, algorithm efficiency, and the theoretical and real-world significance of the contemporary coverless text information hiding methods.

Despite the fact that the contemporary methods of coverless text information hiding have secured and protected the confidential information to a great extent, there is still room for improvement. The existing techniques do not provide a significant capacity to embed the secret information. Moreover, these methods need a huge text database in advance to hide the confidential information. Few of the present methods of coverless text information hiding are a bit difficult to implement, while on the other hand; the accuracy of few techniques is not that good. Therefore, the methods with improved embedding capacity, better accuracy, and sound quality are needed to be developed and implemented to flawlessly secure and protect the secret information.

References

[1] Deb, S.: Information Technology, Its Impact on Society and Its Future. *Adv. Commun.* 4, 25-29 (2014).

[2] Maggiolini, P.: Information Technology Benefits: A Framework. In: *Carugati, A., Rossignoli, C. (eds.) Emerging Themes in Information Systems and Organization Studies*, pp. 281-292. Physica-Verlag HD, Heidelberg (2011).

[3] Hura, G.S.: The Internet: global information superhighway for the future. *Comput. Commun.* 20, 1412-1430 (1998).

[4] Kim, W., Jeong, O., Kim, C., So, J.: The dark side of the Internet: Attacks, costs and response. *Inf. Syst.* 36, 675-705 (2011).

[5] Ali, S.: Steganography and Digital Watermarking as Promising Approaches to Information Hiding: A State-Of-The-Art Review. *Int. J. Res. Appl. Sci. & Eng. Technol.* 5, 313-318 (2017).

[6] Maxemchuk, N.F., Liu, T.Y., Tsai, W.H.: Electronic document distribution. *AT&T Technol. J.* 73, 73-80 (1994).

[7] Huang, D., Yan, H.: Inter word distance changes represented by sine waves for watermarking text images. *IEEE Transactions Circuits Syst.* 11, 1237-1245 (2001).

[8] Por, L.Y., Ang, T.F., Delina, B.: WhiteSteg: a new Scheme in information hiding using text steganography. *WSEAS Transactions Comput.* 7, 735-745 (2008).

[9] Huang, Y.F., Tang, S., Yuan, J.: Steganography in Inactive Frames of VoIP Streams Encoded by Source Codec. *IEEE Transactions Inf. Forensics & Secur.* 6, 296-306 (2011).

[10] Bodo, Y., Laurent, N., Laurent, C., Dugelay, J.L.: Video Waterscrambling: Towards a Video Protection Scheme Based on the Disturbance of Motion Vectors. *EURASIP J. Adv. Signal Process.* 14, 2224-2237 (2004).

[11] Nematollahi, M.A., Vorakulpipat, C., Rosales, H.G.: Digital Watermarking: Techniques and Trends. *Springer, Singapore* (2017).

[12] Chen, X., Sun, H., Tobe, Y., Zhou, Z., Sun, X.: Coverless Information Hiding Method Based on the Chinese Mathematical Expression. In: Huang, Z., Sun, X., Luo, J., Wang, J. (eds.) Cloud Computing and Security, ICCCS 2015, Nanjing, August 2015. *Lecture Notes in Computer Science*, vol. 9483, pp. 133-143. Springer, Cham (2015).

[13] Zhou, Z., Mu, Y., Zhao, N., Wu, Q.M.J., Yang, C-N.: Coverless Information Hiding Method Based on Multi-keywords. In: Sun, X., Liu, A., Chao, H.C., Bertino, E. (eds.) Cloud Computing and Security, ICCCS 2016, Nanjing, July 2016. *Lecture Notes in Computer Science*, vol. 10039, pp. 39-47. Springer, Cham (2016).

[14] Zhou, Z., Sun, H., Harit, R., Chen, X., Sun, X.: Coverless image steganography without embedding. In: Huang, Z., Sun, X., Luo, J., Wang, J. (eds.) Cloud Computing and Security, ICCCS 2015, Nanjing, August 2015. *Lecture Notes in Computer Science*, vol. 9483, pp. 123-132. Springer, Cham (2015).

[15] Zhou, Z., Cao, Y., Sun, X.: Coverless information hiding based on bag-of-words model of image. *J. Appl. Sci.* 2016; 34: 527-536 (2016).

[16] Yuan, C., Xia, Z., Sun, X.: Coverless image steganography based on SIFT and BOF. *J. Internet Technol.* 18, 435-442 (2017).

[17] Zhang, J., Huang, H., Wang, L., Lin, H., Gao, D.: Coverless Text Information Hiding Method Using the Frequent Words Hash. *Int. J. Netw. Secur.* 19, 1016-1023 (2017).

[18] Miller, M.: Acrostic Poems ... and some prose. Seattle, WA, *USA: CreateSpace Independent Publishing Platform*, (2011).

[19] Ali, S., Shao, L.: Digital Text Watermarking and its Application to the Sindhi Language. *Int. J. Res. Appl. Sci. & Eng. Technol.* 5, 944-949 (2017).

[20] Sun, X.M., Chen, H.W., Yang, L.H., Tang, Y.Y.: Mathematical representation of a Chinese character and its applications. *Int. J. Pattern Recognit. & Artif. Intell.* 16, 735-747 (2002).

[21] Zhang, J., Shen, J., Wang, L., Lin, H.: Coverless text information hiding method based on the word rank map. In: Sun, X., Liu, A., Chao, H.C., Bertino, E. (eds.) Cloud Computing and Security, ICCCS 2016, Nanjing, July 2016. *Lecture Notes in Computer Science*, vol. 10039, pp. 145-155. Springer, Cham (2016).

[22] Chen, X., Chen, S., Wu, Y.: Coverless information hiding method based on the Chinese character encoding. *J. Internet Technol.* 18, 91-98 (2017).

[23] Liu, C., Luo, G., Tian, Z.: Coverless Information Hiding Technology Research Based on News Aggregation. In: Sun, X., Chao, H.C., You, X., Bertino, E. (eds.) Cloud Computing and Security, ICCCS 2017, Nanjing, June 2017. *Lecture Notes in Computer Science,* vol. 10602, pp. 153-163. Springer, Cham (2017).

Author's Profiles

Shahbaz Ali earned the B.Eng. in Electrical (Telecoms) Engineering from Sukkur IBA University, Sukkur, Pakistan under the PM's National ICT Scholarship in 2014. He is currently a Research Scholar and pursuing the M.S. in Computer Software and Theory from Shaanxi Normal University, Xi'an, China under the Chinese Govt Scholarship. He has published a few research papers in the international journals. His research interests include Cryptology and Information Security.

How to cite this paper: Shahbaz Ali,"A State-of-the-Art Survey of Coverless Text Information Hiding", International Journal of Computer Network and Information Security(IJCNIS), Vol.10, No.7, pp.52-58, 2018.DOI: 10.5815/ijcnis.2018.07.06