iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** VII **Month of publication:** July 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 5 Issue VII, July 2017- Available at www.ijraset.com

Digital Text Watermarking and its Application to the Sindhi Language

Shahbaz Ali[1], Liping Shao[2]
[1]*Research Scholar, School of Computer Science, Shaanxi Normal University, Xi'an, China*
[2]*Associate Professor, School of Computer Science, Shaanxi Normal University, Xi'an, China*

Abstract*: Because of its unique nature, the digital data can easily be exchanged, reproduced and distributed in the blink of an eye. However, this creates easiness for the cyber criminals to misuse the digital data, which ultimately introduces the problem of data ownership. Digital watermarking is one of the ascendant technologies of information hiding which helps solve the data ownership problem. There are several types of digital watermarking, out of which digital text watermarking is very distinctive because the text has a very limited capacity to embed the secret mark, so it is quite challenging to protect and secure the digital text. This paper presents the up-to-the-minute information on the technology of digital text watermarking and its application to the Sindhi language, which is an Indo-Aryan language and recognized as the official language of Sindh (the second largest province of Pakistan). Unlike English, the writing system of Sindhi is completely different, so the application of digital text watermarking to the Sindhi language becomes even more challenging. In this study, a few important aspects of digital text watermarking such as main techniques of text watermarking and conventional attack approaches are discussed in detail to provide readers with the complete know-how of digital text watermarking.

Keywords: Digital Text Watermarking, Information Hiding, Data Security, Sindhi Language, Data Ownership, Steganography

I. INTRODUCTION

In today's world, the sharing of information has become much easier. No matter how huge the amount of information, it can easily be transmitted in the blink of an eye, and this became achievable when the technology of the internet was invented [1], [2]. The Internet has helped exchange, reproduce and distribute the digital data in less than a second. Despite the fact that the internet technology has benefited humanity in a lot of ways, it created easiness for the cyber criminals to misuse the important information [3]. This misuse of the information leads to the creation of the data ownership problem. As a consequence, it becomes very challenging to protect and secure the essential information. Digital watermarking is one of the dominant technologies of information hiding which efficiently solves the problem of data ownership [4]. According to different forms of the digital content, there are various types of digital watermarking such as digital text watermarking, digital audio watermarking and digital video watermarking [5], [6].

Digital text watermarking is a form of digital watermarking which is used to protect the critical digital text by embedding a secret mark into that digital text [7]. Unlike digital audio and video watermarking, digital text watermarking offers a very low capacity to hide the secret mark because of the unique characteristics of the digital text, so special care has to be taken in this case to protect and secure the important digital text in the best possible way. Every language of the world has its unique writing system so protecting and securing the essential text becomes a very challenging task [8]. Unlike English, the writing system of the Sindhi language is based on the Perso-Arabic script so the application of digital text watermarking to the Sindhi language requires special attention. Fig. 1 shows a text written in English and Sindhi.

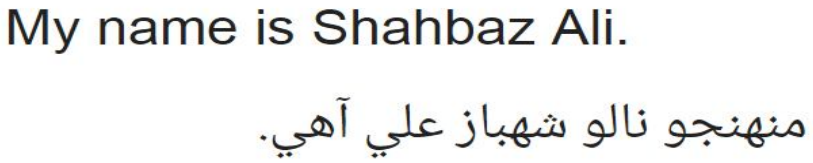


Fig. 1 Text written in English and Sindhi

This paper provides readers with the up-to-the-minute information on the technology of digital text watermarking and its application to the Sindhi language. Some significant aspects such as main techniques of digital text watermarking and conventional attack approaches are elucidated in a simpler and clearer way, which helps readers to have the complete know-how of digital text watermarking.

II. DIGITAL TEXT WATERMARKING: A COMPLETE OVERVIEW

Digital text watermarking is a distinctive type of watermarking which is used to protect and secure the important digital text from being reproduced and distributed in an illegal way. It is used to verify the ownership of the copyrights of the digital text, and this is achieved by embedding a hidden watermark into the digital text whose sole purpose is to check the originality and integrity of that text [9]. Fig. 2 shows the basic working principle of digital text watermarking.

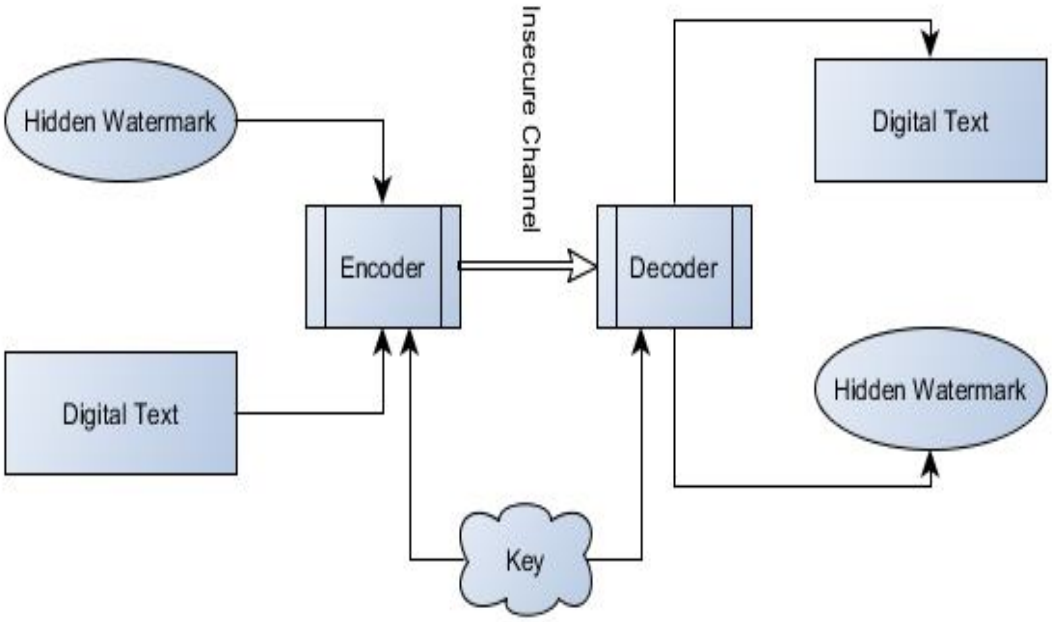


Fig. 2 Working Principle of Digital Text Watermarking

With the help of a text watermarking encoding algorithm, the hidden watermark is embedded into the digital text and then the encoded data is transmitted over an insecure channel. When the encoded data reaches at the decoding part, it is decoded with the help of a key and the hidden watermark is obtained successfully at the end.

A. Digital Text Watermarking Techniques

There are several techniques of digital text watermarking. A few of which are given as:

1) Word-Shift Coding Based Text Watermarking: In this technique of digital text watermarking, the words of the digital text are relocated in the horizontal position which results in the expansion of the space, and then it is utilized for embedding the hidden watermark into the text [10]. In this technique, the words closest to the words being relocated are kept in the original state, which ultimately helps in the decoding process at the end. Fig. 3 shows an example of word-shift coding based text watermarking in which two words “Computer” and “Normal” in the second line are relocated horizontally.

School of Computer Science, Shaanxi Normal University

School of Computer Science, Shaanxi Normal University

Fig. 3 Word-Shift Coding Based Text Watermarking

2) Feature Coding Based Text Watermarking: In this technique of digital text watermarking, several features of the text such as pixels and the length of the characters' end lines are modified to embed the hidden watermark into the text [11]. Every language has its unique features which can be utilized for the purpose of feature coding based text watermarking.

3) Line-Shift Coding Based Text Watermarking: In line-shift coding based text watermarking technique, the line of the digital text is repositioned in the vertical position to embed the hidden watermark into the text [12]. The lines closest to the line being repositioned are kept in the original state. The lines maintained in the original state help in the decoding process. The line-shift coding based text watermarking example is shown in Fig. 4 in which the middle line is displaced vertically.

School of Computer Science, Shaanxi Normal University

School of Computer Science, Shaanxi Normal University

School of Computer Science, Shaanxi Normal University

Fig. 4 Line-Shift Coding Based Text Watermarking

4) Syntactic and Semantic Based Text Watermarking: In syntactic based text watermarking technique, the syntactic structures of the text are utilized, and with the application of syntactic transformation, the hidden watermark is embedded into the digital text. For using this type of technique, a syntactic tree is needed on which the process of syntactic transformation is applied to embed the hidden watermark. In semantic based text watermarking method, the semantic structures of the text are utilized for embedding the hidden watermark into the digital text. There are various types of semantic structures such as grammar rules, the content of the text, nouns, spellings, structure of the sentences and verbs. By using the proper synonyms of some specific words in the text, the watermark can be embedded secretly [13].

5) Structure Based Text Watermarking: In structure based text watermarking technique, the structure of the digital text is utilized to embed the hidden watermark into the text. Several parameters are used in this type of text watermarking, and a few of which are localization, sequence, the rate of recurrence and components of the text. In structure based text watermarking, the original state of the digital text is preserved [14].

B. Conventional Attacks Approaches against Digital Text Watermarking

The technology of digital text watermarking is still immature, and this helps cyber criminals to launch several types of attacks against digital text watermarking. A few types of conventional attacks against text watermarking are shown in Fig. 5.

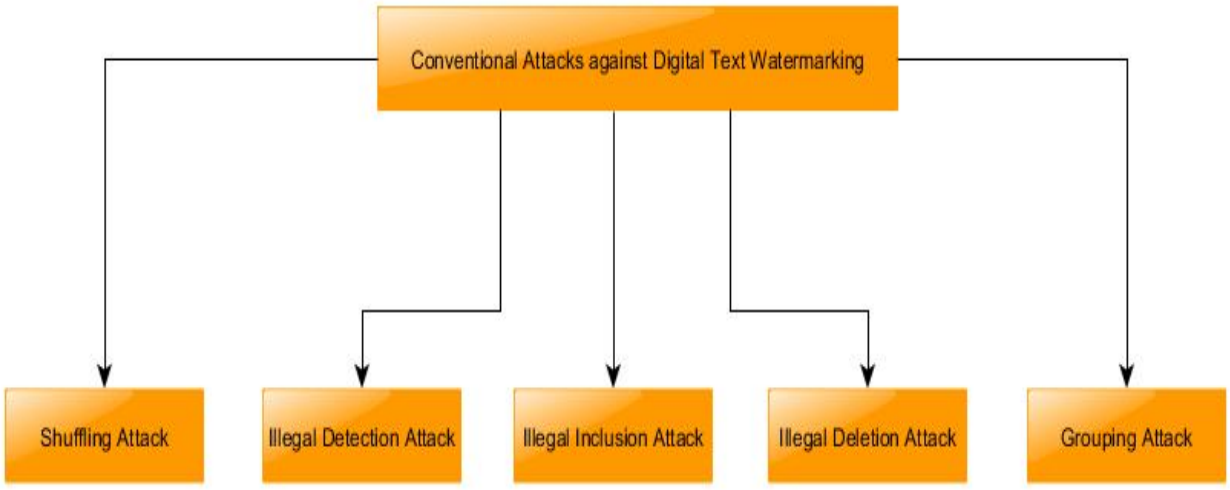


Fig. 5 Conventional Attacks against Digital Text Watermarking

In shuffling attack type, cyber criminals try to jumble up some parts of the text which result in the destruction of the hidden watermark. In detection attack, an attempt is made by cyber criminals to locate and access the watermark hidden into the digital text.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 5 Issue VII, July 2017- Available at www.ijraset.com

In inclusion attack, cyber criminals try to include some false parts into the original digital text to convey the bogus information. In deletion attack, an attempt is made by cyber criminals to delete some parts of the original digital text to conceal the identity of the original owner of that text. In grouping attack, cyber criminals try to launch a combined attack which may include tampering, text deletion, false text insertion and text shuffling [5].

III.DIGITAL TEXT WATERMARKING AND SINDHI

The Sindhi language is an Indo-Aryan language and recognized as the official language of Sindh (the second largest province of Pakistan). Unlike English, the writing system of Sindhi is based on the Perso-Arabic script. Thus, the application of digital text watermarking to the Sindhi language is quite challenging and requires special attention. The alphabet of Sindhi has 52 letters. Fig. 6 shows the alphabet of the Sindhi language.

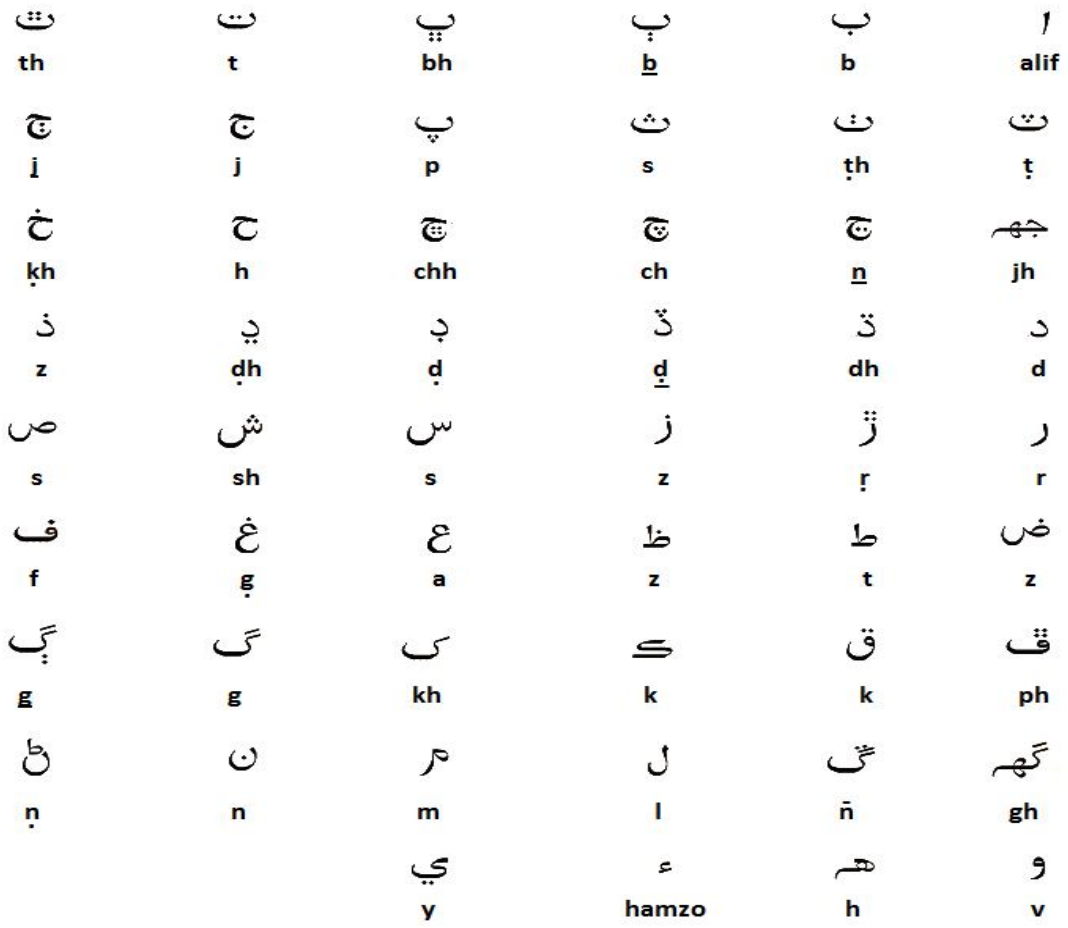


Fig. 6 Sindhi Alphabet

The writing system of Arabic, Persian, Urdu and Sindhi is based on the Perso-Arabic script. Moreover, many words when written are actually formed by connecting several letters to each other. However, as compared to Arabic, Persian and Urdu, the writing system of Sindhi is quite unique. Out of 52 letters of Sindhi alphabet, 34 letters have dots. Furthermore, the Sindhi language has a quite large stock of unique letters which are not found in Arabic, Persian and Urdu. These distinctive features of the writing system of Sindhi makes Sindhi a unique language.

Because of the distinctive nature of the writing system of Sindhi, many techniques of digital text watermarking can be applied to protect and secure the digital text. In Fig. 7 the technique of line-shift coding is applied, and the third line is displaced vertically.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 5 Issue VII, July 2017- Available at www.ijraset.com

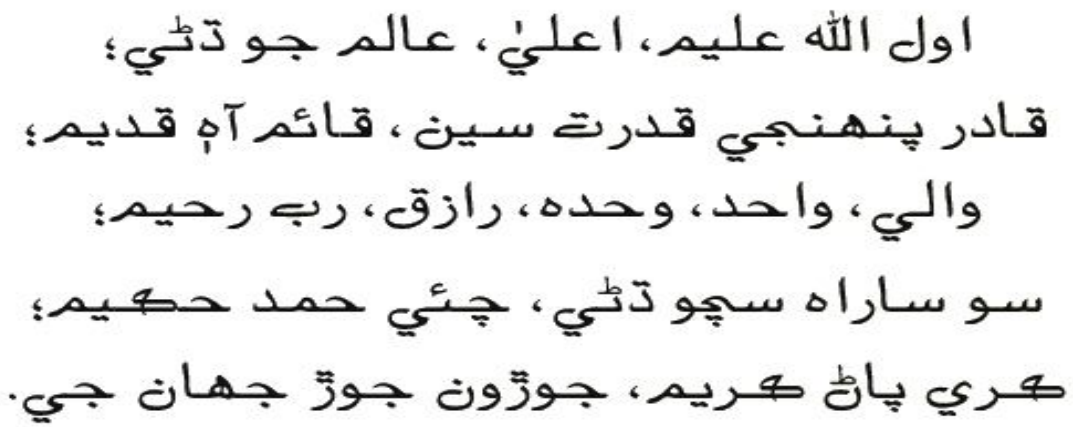


Fig. 7 Line-Shift Coding applied to Sindhi

In Fig. 8, the technique of word-shift coding is applied on the second line of the text of Sindhi, and three words (fourth, ninth and twelfth when reading in the order from right to left) are relocated horizontally.

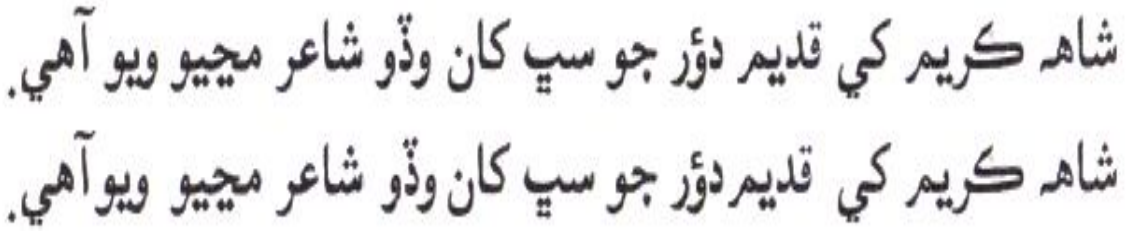


Fig. 8 Word-Shift Coding applied to Sindhi

In Fig. 9, the technique of feature coding is applied on two lines (second and fourth) of the text of Sindhi and different features of the text such as fatah (slash like symbol on the top of the letters), the length of the letters and dots of the letters are changed. The traditional fatah is changed into the reverse fatah, and the vertical position of the dots and the space between dots is changed slightly. Furthermore, the length of a few letters is increased and decreased slightly.

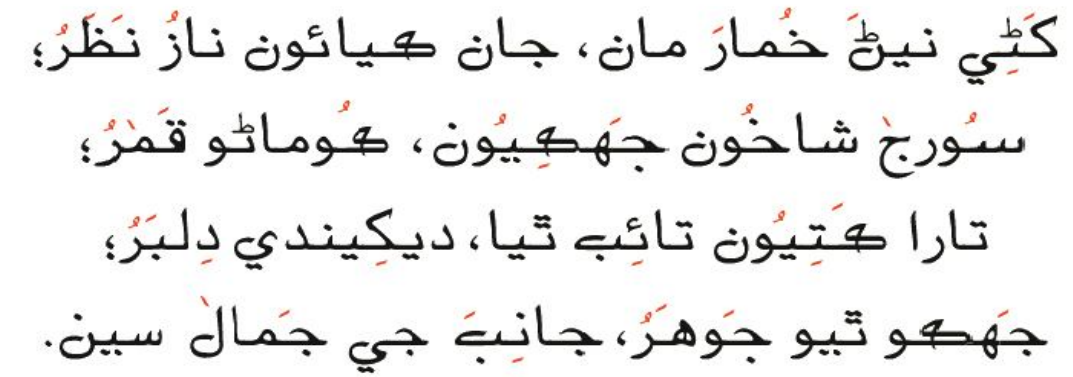


Fig. 9 Feature Coding applied to Sindhi

Every technique of digital text watermarking has some merits and demerits. Word-shift coding and feature coding based text watermarking techniques are not strong enough to survive under different attacks launched by cyber criminals. However, line-shift coding based text watermarking method has the ability to withstand various attacks. Syntactic and semantic based watermarking techniques have proved to be efficient watermarking techniques to protect and secure the digital text. Moreover, in syntactic and

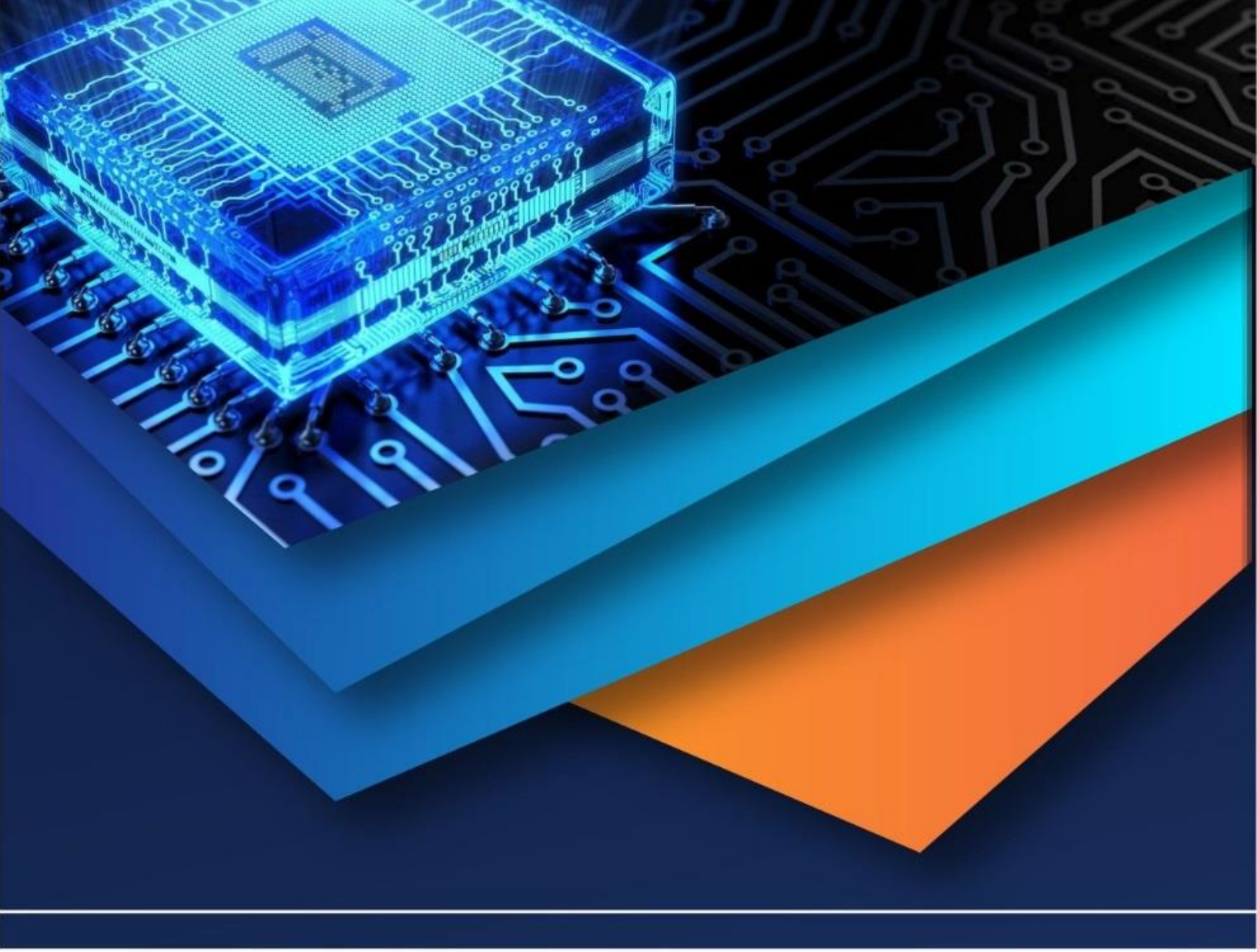
semantic based watermarking techniques, the core characteristics of the digital text are well-kept-up. Better results are achieved when several text watermarking techniques are combined in several ways.

IV. CONCLUSIONS

The technique of digital text watermarking helps protect and secure the digital text. This paper presented the contemporaneous information on digital text watermarking and elucidated some significant aspects such as primary techniques of digital text watermarking and conventional attack approaches to help readers have the complete know-how of digital text watermarking. Moreover, digital text watermarking was applied to Sindhi (an Indo-Aryan language whose writing system is based on the Perso-Arabic script), and the robustness of the contemporary text watermarking methods was evaluated in detail. It is concluded that the present text watermarking methods have played a significant part in securing the digital text. However, this area needs further research, and there is a need to design and implement more real-world and robust text watermarking algorithms for protecting and securing the digital text flawlessly.

REFERENCES

[1] S. Ali, "Steganography and Digital Watermarking as Promising Approaches to Information Hiding: A State-Of-The-Art Review," International Journal for Research in Applied Science & Engineering Technology, vol. 5, no. 7, pp. 313-318, 2017.
[2] G. S. Hura, "The Internet: global information superhighway for the future," Computer Communications, vol. 20, no. 16, pp. 1412-1430, 1998.
[3] W. Kim, O. Jeong, C. Kim, and J. So, "The dark side of the Internet: Attacks, costs and response," Information Systems, vol. 36, no. 3, pp. 675-705, 2011.
[4] M. Thapa and S. K. Sood, "On Secure Digital Image Watermarking Techniques," Journal of Information Security, vol. 2, pp. 169-184, 2011.
[5] M. A. Nematollahi, C. Vorakulpipat, and H. G. Rosales, "Digital Watermarking: Techniques and Trends," Springer Topics in Signal Processing, Singapore, Springer, vol. 11, 2017.
[6] L. Robert and T. Shanmugapriya, "A Study on Digital Watermarking Techniques," International Journal of Recent Trends in Engineering, vol. 1, no. 2, pp. 223-225, 2009.
[7] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker, "Digital Watermarking and Steganography," The Morgan Kaufmann Series in Multimedia Information and Systems, Massachusetts, Morgan Kaufmann Publishers, 2nd ed., 2007.
[8] A. A. Khodami and K. Yaghmaie, "Persian Text Watermarking," Proceedings of the 7th Pacific Rim conference on Advances in Multimedia Information Processing, vol. 4261, pp. 927-934, 2006.
[9] S. Katzenbeisser and F. A. P. Petitcolas, "Information Hiding Techniques for Steganography and Digital Watermarking," Artech House Computer Security Series, Boston, Artech House Publishers, 1st ed., 1999.
[10] J. A. Memon, K. Khowaja, and H. Kazi, "Evaluation of Steganography for Urdu/Arabic Text," Journal of Theoretical and Applied Information Technology, vol. 4, no. 3, pp.232-237, 2008.
[11] M. Kaur and K. Mahajan, "An Existential Review on Text Watermarking Techniques," International Journal of Computer Applications, vol. 120, no. 18, pp. 29-32, 2015.
[12] M. S. Memon and A. Shah, "A Novel Text Steganography Technique to Arabic Language Using Reverse Fatha," Pakistan Journal of Engineering, Technology & Science, vol. 1, no. 2, pp. 106-113, 2011.
[13] R. A. Alotaibi and L. A. Elrefaei, "Arabic Text Watermarking: A Review," International Journal of Artificial Intelligence & Applications, vol. 6, no. 4, pp. 1-16, 2015.
[14] X. Rui, C. XiaoJun, and S. Jinqiao, "A Multiple Watermarking Algorithm for Texts Mixed Chinese and English," Proceedings of the International Conference on Information Technology and Quantitative Management, vol. 17, pp. 844-851, 2013.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY