International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

COPYRIGHT PROTECTION USING DATAMINING ANALYSIS

Rinta Jolly

Student, Dept. of Dual Degree Computer Applications, Sree Narayana guru institute of science and technology paravur, Kerala, India

---

Abstract: *These days, with the broad utilization of web everywhere throughout the world, the vast majority of the administration, open and private information are progressively being distributed on web. The security of these online archives is the need of great importance and should be managed direly. Information assurance from malignant assaults and abuse has become a significant issue. Different sorts of information, including pictures, recordings, sound and content reports, have warranted the improvement of various techniques for their insurance. Cryptography, advanced marks and steganography are the most notable advances used to secure information. During the most recent decade, advanced watermarking innovation has likewise been used as a choice to forestall media imitation and altering or adulteration to guarantee both copyright and validation. Much work has been done to secure pictures, recordings and sound however just a couple of Algorithm have been considered for content archive insurance with computerized watermarking. In any case, our review saw that accessible content watermarking calculations are neither powerful nor intangible and as such stay unbound strategies for security. Henceforth, research to improve the exhibition of content watermarking calculations is required. The computerized libraries offer successful approaches to get to instructive materials, government e-archives, budgetary reports, online networking substance and numerous others. Anyway content origin and alter recognition of all these advanced content reports require uncommon consideration. The proposed strategy achieved an elevated level of subtlety where pinnacle sign to commotion proportion (PSNR) values are somewhere in the range of 64.67% and 71.03%, and likeness (SIM) rate is somewhere in the range of 99.92% and 99.99%. The proposed strategy is vigorous and opposes from arranging assaults and limit of the proposed procedure is likewise improved when contrasted with the past strategies.*

***Key Words*: Cryptography, Encryption, Decryption, Neural Network, Soft Computing, Security, Algorithm.**

1. INTRODUCTION

On the other hand, corresponding to the negative aspects, the misuse of these technologies raises many issues like copyright protection and data manipulation. Due to the advanced technologies such as high-speed computer networks, and Internet, etc. It is necessary to secure digital contents and protects them against unauthorized copy. Internet of Things (IoT) and cloud has received significant support from governments and research institutes around the world. Data is the crucial element in smart cities which sustains the infrastructure of data and helps people to gain access to digital contents. A secret message is placed inside a digital content without compromising valuable data. This secret information is used later for ownership identification. Digital watermarking is categorized into text watermarking, image watermarking, audio water marking and video watermarking. Most of the research has focused on image, audio, and video. Currently, text watermarking has received popularity due to large numbers of the text document are produced and shared. The advanced watermark is installed with the excess bearer of the content, and it should have better power, great visual effects and immense watermark limit. Consequently, we have to make a tradeoff among the three to accomplish better outcomes.

By and large, the ebb and flow content watermarking calculations can be isolated into three classes: One depends on the content structure, another depends on the common language preparing, and the third depends on the picture handling. Phonetic based methodologies utilize the semantic highlights of the content to insert a watermark.

For instance, by supplanting certain words or sentences in the content to install the watermark without contacting the semantics of the content substance. The extent of this paper is in this way restricted to the advanced watermarking of content reports, the investigation is sorted out as follows: The main segment characterizes the significance of computerized content watermarking though the general idea and design of content watermarking frameworks are portrayed in the subsequent segment.

It ought to likewise be versatile to various content arrangements and ought to have high data conveying limit. It ought to be adequately applied to print/computerized evidence.

2. EXISTING SYSTEM

An advanced watermark is a sort of marker clandestinely inserted in a clamor lenient sign, for example, sound, and video or picture information. It is regularly used to recognize responsibility for copyright of such sign. "Watermarking" is the way toward concealing advanced data in a transporter signal; the shrouded data should, yet doesn't have to, contain a connection to the bearer signal. Computerized watermarks might be utilized to confirm the legitimacy or trustworthiness of the bearer signal or to show the character of its proprietors. It is unmistakably utilized for following copyright encroachments and for banknote validation. Watermarks are recognizable proof imprints delivered during the paper making process. The primary watermarks showed up in Italy during the thirteenth century, however their utilization quickly spread across Europe. They were utilized as a way to distinguish the paper creator or the exchange society that produced the paper. The imprints frequently were made by a wire sewn onto the paper form. Watermarks keep on being utilized today as producer's imprints and to forestall fabrication. Past deals with content watermarking could be classified in three principle classes including an image-based approach, a syntactic methodology, and a semantic methodology.

Watermark embeds in text image through customizing the interline or word gaps in the middle of lines and words 27, 28. The syntactic structure of text is made and used in syntactic approach to embed watermark bits through some transformation like passivation, clefting, and topicalization 29-31.Finally, in semantic approach, synonyms 32, 33, acronyms 34, 35, words spelling, pre-supposition 36, 37, and text meanings as text's semantics are considered as watermark in text. The algorithms used in text watermarking by using binary text image are not strong against retyping. In syntactic approaches, researchers mixed algorithms with the natural language processing (NLP). These algorithms are more effective, but the main issue is slow research progress in NLP and inefficient syntactic analyzers.

3. PROPOSED SYSTEM

Current security challenges for IoT that need to be sorted out are presented, and it shows that data integrity, security, privacy, automation, updating, a common frame-work, and encryption capabilities are the main challenges. In IoT, text documents integrity and security issues are exist in a modern digital world. A large number of text documents are generated daily and shared through IoT. MS used in text watermarking by using binary. Due to advanced technologies, these documents can be easily copied and redistributed. IoT has unlimited benefits, but on the contrary, illegal use of these documents creates a problem for the original. Nowadays, a number of ways have been used by hackers to infect or access the information. Digital text document protection is a crucial issue for researchers in the modern world. , we use of digital libraries, Internet technologies, mobile phones, e-commerce, and iPods are a fast and easy way of broadcasting information [8]. However, the security and privacy of digital content are difficult to handle. In this case, it is necessary to provide protection to digital materials that are traveling over the internet.

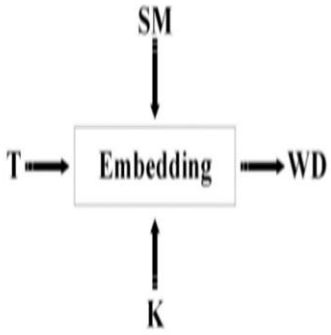
4. METHODOLOGY

An advanced watermark is a sort of marker clandestinely inserted in a clamor lenient sign, for example, sound, and video or picture information. It is regularly used to recognize responsibility for copyright of such sign. "Watermarking" is the way toward concealing advanced data in a transporter signal; the shrouded data should, yet doesn't have to, contain a connection to the bearer signal. Computerized watermarks might be utilized to confirm the legitimacy or trustworthiness of the bearer signal or to show the character of its proprietors. It is unmistakably utilized for following copyright encroachments and for banknote validation. Watermarks are recognizable proof imprints delivered during the paper making process. The primary watermarks showed up in Italy during the thirteenth century, however their utilization quickly spread across Europe. They were utilized as a way to distinguish the paper creator or the exchange society that produced the paper. The imprints frequently were made by a wire sewn onto the paper form. Watermarks keep on being utilized today as producer's imprints and to forestall fabrication. Past deals with content watermarking could be classified in three principle classes including an image-based approach, a syntactic methodology, and a semantic methodology.

Watermark embeds in text image through customizing the interline or word gaps in the middle of lines and words 27, 28. The syntactic structure of text is made and used in syntactic approach to embed watermark bits through some transformation like passivation, clefting, and topicalization 29-31.Finally, in semantic approach, synonyms 32, 33, acronyms 34, 35, words spelling, pre-supposition 36, 37, and text meanings as text's semantics are considered as watermark in text. The algorithms used in text watermarking by using binary text image are not strong against retyping. In syntactic approaches, researchers mixed algorithms with the natural language processing (NLP). These algorithms are more effective, but the main issue is slow research progress in NLP and inefficient syntactic analyzers.

4.1 Watermark Embedding

- The secret message is encrypted using steganography algorithm.
- The encrypted message is shifted into the next phase, where the watermark is generated.
- The encrypted message is converted into binary and then numbers.

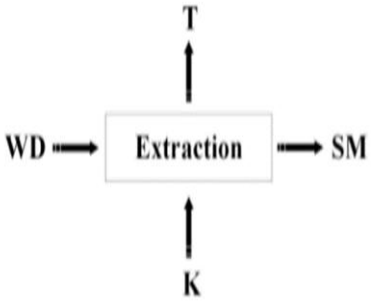


(Watermark embedding)

Where "SM" denotes the secret message, "T" represents the original document, "WD" is a watermarked document and "K" denotes Key.

4.2 Watermark Extraction

- Watermark extraction or verification process extract the watermark (secret information) from the atermarked document.
- Watermark extraction is the converse procedure of watermark inserting.



(Watermark extraction)

5. FUTURE WORK

In the future, the proposed solution will be extended for the printed text documents copyright protection. Research work has to be focused on ensuring robustness, integrity and accuracy of text documents. Hence we conclude that new watermarking algorithms, which are computationally inexpensive and robust, need to be developed to ensure online safety of text documents. In future more work is required to be done in the area of security of text watermarking. More secure methods need to be developed to ensure security of text documents over internet.

6. CONCLUSION

A robust and secure watermarking algorithm is proposed to authenticate the digital contents in smart cities. The performance of the proposed technique is compared and verified with the previous techniques to confirm the imperceptibility, security, robustness, and capacity. Several techniques have been proposed in this field, but still need a technique which is applicable for the cloud, IoT devices, and smart cities. Through experiments, the proposed algorithm is highly imperceptible and achieve about 99.99 similarity factor. After applying formatting attacks such as cut, copy, paste, font size, font color, and alignment proposed algorithm proves that it is robust and to legate most of possible attacks and watermark is extracted with high accuracy. The capacity of the proposed algorithm is also increased as compared with previous techniques. In the cloud computing environment, the proposed technique gives the same results which are suitable in smart cities to ensure the security of text documents.

ACKNOWLEDGEMENT

In the name of almighty, I would like to extend my heartfelt thanks to our HoD Mrs.Kavitha C.R, Department of a Dual Degree Master of Computer Applications for the helps extended to me throughout my course of my study. I am deeply grateful to my guide Mrs. Sandhya TJ.Assistant Professor, Department of a Dual Degree Master of Computer Applications for the valuable guidance

REFERENCES

[1]. S. G. Rizzo, F. Bertini, and D. Montesi, ''Content-preserving text watermarking through Unicode homo glyph substitution,'' in Proc. 20th Int. Database Eng. Appl. Symp., 2016, pp. 97–104.
[2] .O. Tayan, M. N. Kabir, and Y. M. Alginahi, ''A hybrid digital-signature and zero-watermarking approach for authentication and protection of sensitive electronic documents,'' Sci. World J., vol. 2014, Aug. 2014, Art. No. 514652.
[3]. K. Thong or and T. Amornraksa, ''Digital image watermarking for printed and scanned documents,'' Proc. SPIE, vol. 10420, Jul. 2017, Art. no. 104203O.
[4]. Y. Liu, Y. Zhu, and G. Xin, ''A zero-watermarking algorithm based on merging features of sentences for Chinesetext,''J.Chin.Inst.Eng.vol.38, no. 3, pp. 391–398, Apr. 2015.
[5]. M. Yingjie, L. Huiran, S. Tong, and T. Xiaoyu, ''A zero-watermarking scheme for prose writings,'' in Proc. Int. Conf. Cyber-Enabled Distrib. Comput. Knowl. Discovery (CyberC), Oct. 2017, pp. 276–282.
[6].A.Taha,A.S.Hammad,andM.M.Selim,''Ahighcapacityal gorithmfor information hiding in Arabic text,'' J. King Saud Univ.-Comput. Inf. Sci., to be published.
[7]. F. M. Ba-Alwi, M. M. Ghilan, and F. N. Al-Wesabi, ''Content authentication of english text via Internet using zero watermarking technique and Markov model,'' Int. J. Appl. Inf. Syst., vol. 7, no. 1, pp. 25–36, 2014.
[8].Y.Zhang,H.Qin,andT.Kong,''Anovelrobusttextwaterm arkingforword document,'' in Proc. 3rd Int. Conger. Image Signal Process. (CISP), vol. 1, Oct. 2010, pp. 38–42.