International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 02 Issue: 08 | Nov-2015 www.irjet.net p-ISSN: 2395-0072

# Secure Data Recovery with Enhancement of Reversible Watermarking

**Mrs.Divyamala.A [1] , Mr.Raja.R [2], Mr.Baskaran.G [3],Ms.Zaibunnisa.S [4]**

[1]Student, Dept.of comp.sci.,Dhanalakshmi Srinivasan Engineering College,Tamilnadu,India

[2]Assistant Professor, Dept.of comp.sci.,Dhanalakshmi Srinivasan Engineering College,Tamilnadu,India

[3]Assistant Professor,Dept.of comp.sci., Dhanalakshmi Srinivasan Engineering College,Tamilnadu,India

[4]Student, Dept.of comp.sci., Dhanalakshmi Srinivasan Engineering College,Tamilnadu,India

----------**********----------

***Abstract -****Advancement in information technology is playing an increasing role in the use of information systems comprising relational databases. These databases are used effectively in collaborative environments for information extraction; consequently, they are vulnerable to security threats concerning ownership rights and data tampering. Watermarking is the process of hiding in different types of data formats like image, audio, video and text. In existing system they used, data encryption standard algorithm. It displays poor security and produces lot of errors. The partitions of the data to ensure minimum distortion; therefore, recover original data with degraded data quality and lack robustness. In this system the data hiding and recovery process it produce lot of errors in the timing for image recovery. So anyone can attack the data . To overcome this type of attack, there are two different types of algorithms discrete wavelet decomposition and chaos algorithm are used in proposed system. The Discrete Wavelet Transform is used to remove the noises and allocate the space for digital data. This algorithm improves security on data hiding and data recovery. Experimental studies prove the effectiveness of Robust Reversible Watermarking (RRW) against malicious attacks and show that the proposed technique out performs existing ones.*

***Key Words:*** *Reversible watermarking, genetic algorithm, data recovery, data quality, robustness, numerical data.*

## 1. INTRODUCTION

In the digital world of today, data is excessively being generated due to the increasing use of the Internet and cloud computing [1]. Data is stored in different digital formats such as images, audio, video, natural language texts and relational data. Relational data in particular is shared extensively by the owners with research communities and in virtual data storage locations in the cloud. The purpose is to work in a collaborative environment and make data openly available so that it is useful for knowledge extraction and decision making. Take the case of Walmart—a large multinational retail corporation that has made its sales database available openly over the Internet so that it may be used for the purposes of identifying market trends through data mining.

However these openly available datasets make attractive targets for attacks. For example there are documented attack incidents where data containing personal information related to customers using certain Walmart video services was stolen. According to a survey related to the security of outsourced customer data [4], it is reported that 46 percent of organizations do not consider security and privacy issues while sharing their confidential data. Therefore, 64 percent organizations have to face data loss repeatedly. Similarly, data breaches in the health care and medical domain are increasing alarmingly. Therefore it is imperative, that in shared environments such as that of the cloud, security threats that arise from un-trusted parties and relational databases need to be addressed along with the enforcement of ownership rights on behalf of their owners.

Reversible watermarking tries to overcome the problem of data quality degradation by allowing recovery of original data along with the embedded watermark information. This paper proposes one such reversible watermarking technique that keeps the data useful for knowledge discovery. Data modifications are allowed to such extent that the quality of the data before embedding watermark information and after extracting, is acceptable for knowledge extraction process.

## 2. SYSTEM MODEL

In this section we consider the existing system design and the proposed system.

## 2.1 Existing System

- The existing system data encryption standard is used for security.
- The partitions of the data to ensure minimum distortion; therefore, recover original data with degraded data quality and lack robustness.
- In this system the data hiding and recovery process it produce lot of errors in the timing for image recovery.

**Problem Identified:**

- The recover original data with degraded data quality and lack robustness.
- The original Image can be retrieved from the encrypted image after extracting or removing the data hidden in the image.
- It produces lot of errors in the timing for image recovery.

## 2.2 Proposed System

In our system we proposed reversible watermarking techniques is required that ensures, (i) Watermark encoding and decoding by accounting for the role of all the features in knowledge discovery; and, (ii) original data recovery in the presence of active malicious attacks. In this system, a robust and semi-blind reversible watermarking (RRW) technique for numerical relational data has been proposed that addresses the above objectives.

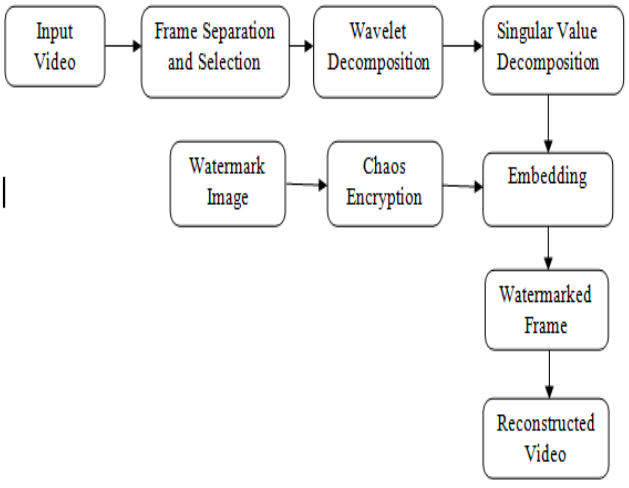


**Fig-1: Embedding Process**

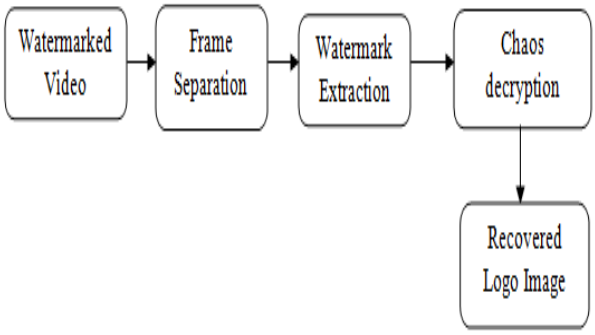


**Fig-2: Extraction Process**

**Benefits:**

- In RRW is used to generate a parameter that controls the data distortions to make sure that the data quality remains intact after watermarking.
  - The semi-blind nature of the technique allows robustness against heavy attacks and also for regeneration of the original dataset after watermark decoding.
  - Original data recovery in the presence of active malicious attacks.
  - High hiding Capacity & High Robustness Watermark will be well protected from Invariant features.

## 3. DESIGN CONSTRUCTION

This Section consists of the following module design to form the ring routing protocol. These are to be explained in this section.

### 3.1 Frames Separation

The Video Format which converts to number of frames which can allocate and can be store into images. The Frames separation format proceeds for embedding process.

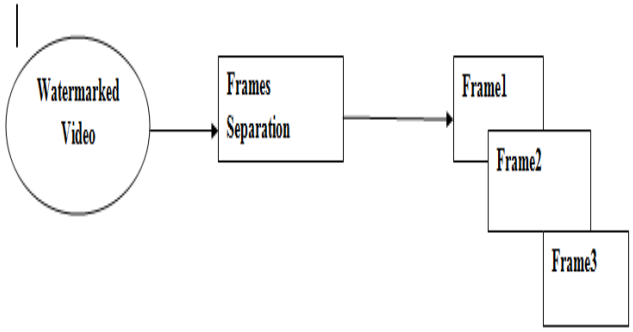

**Fig-3: Frames Separation**

## 3.2 Wavelet Decomposition

The wavelet transform involves projecting a signal onto a complete set of translated and dilated versions of a mother wavelet. The strict definition of a mother wavelet will be dealt with later so that the form of the wavelet transform can be examined first. For now, assume the loose requirement that has compact temporal and spectral support (limited by the uncertainty principle of course), upon which set of basic functions can be defined. LWT decomposes the image into different subbands images, namely, LL, LH, HL, and HH for embedding the messages in the pixel coefficients of subbands. Lifting scheme is a technique to convert DWT coefficients to Integer coefficients without losing information. LL subbands contain the significant part of the spatial domain image. High-frequency sub band contains the edge information of input image. These coefficients are selected as reserved space foe hiding the text data. The secret text data is embedded into the wavelet coefficients of high frequency subbands because it is non sensitive to human visual system.

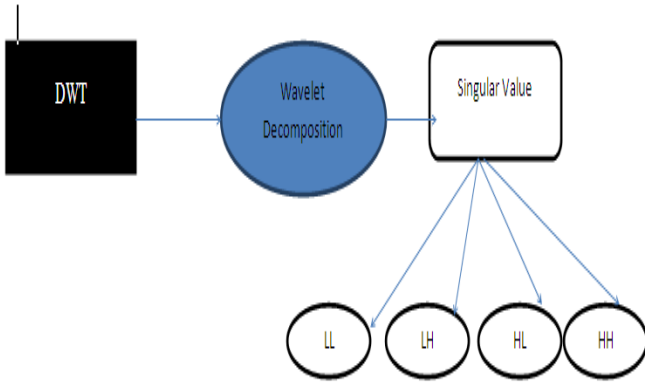


**Fig-4: Wavelet Decomposition**

## 3.3 Embedding

The Logo image can be embedded within frames which can be decomposed. The singular value decompose occur here and transformation performed.

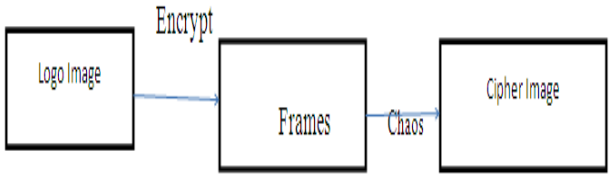


**Fig-5: Embedding**

## 3.4 Validation

It can be validate and produce exact Mean Noise Error value and Signal to Noise ratio for DE noising and reconstruction can be send to receiver.

## 3.5 Extraction

The Reconstructed Chaos Encryption made and Decryption performed using that technique and result can be obtained recovered logo image.

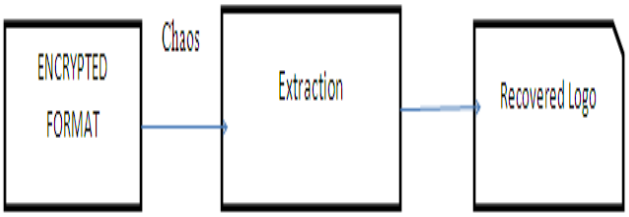


**Fig-6: Extraction**

## 4.Results and Discussion

Experiments are conducted on Intel Core i3 with CPU of 2.40 GHz and RAM of 2 GB. For brevity, heart disease medical dataset, containing more than 300 tuples is selected. RRW was evaluated for: 1. investigating effect on the data quality of the underlying data. 2. robustness against malicious attacks; and 3.restoration of the original data. The data recovery, watermark detection accuracy and effect of RRW on data quality are evaluated using the case study of a heart disease medical dataset. A small set of tuples from the same dataset are also used as

an example to illustrate the entire procedure step by step.

Robustness of RRW is demonstrated through an extensive attack analysis. Our results have shown 100 percent accuracy in both watermark detection and data recovery. The experiments performed, demonstrate data recovery in best case as well as in worst case scenarios where Mallory tries to insert, alter and delete 10, 20, 30, 40, 50 percent and up-to 100 percent of the data and the results are plotted in the graphs below. After such attacks, RRW recovered 100, 65.02 and 50.50 percent tuples with 50 percent insertion, alteration and deletion attacks on data respectively.

In another experiment, 37 percent of tuples have been recovered with 100 percent distortion in data. The effect of RRW on the statistical measures reported in the GA fitness function as in Equation (4) is analyzed with the original as well as watermarked data. Results are compared with effect on the statistical measures of the PEEW technique before and after watermarking. RRW is also compared with DEW, GADEW and PEEW techniques for watermark detection accuracy with subset insertion, subset alteration and subset deletion attacks. In all these scenarios, RRW technique has shown better results.

D (Original Data) for Watermarking

| A1 | A2 | ... | Class Labels |
|---|---|---|---|
| 63.0 | M | ... | 1 |
| 67.0 | F | ... | 0 |
| 37.0 | M | ... | 3 |
| 41.0 | M | ... | 4 |

**Table 1: Original Data for Watermarking**

The computational time of RRW is ðl _ R _ AÞ where l is the watermark length, R is the total number of tuples in the dataset and A is the feature selected for watermarking. The number of tuples are usually much larger as compared to the number of features in databases and the watermark length l; so, A <<R and l <<R. Therefore, for large databases, (R termed as 0n0) the time complexity of RRW for watermark insertion and detection .For datasets involving large number of features or large number of tuples, the data owner may use a separate machine, with high computation power, for watermarking the datasets. This is a small price to pay for an improved sense of privacy against data theft (false claim of ownership can be tackled by watermark encoding and decoding). Computational time of PEEW is also OnÞ as this technique also needs to process all the tuples.

DEW marks less number of tuples based on whether they meet a certain threshold; however, it still needs to process all tuples leading to a time complexity. For GADEW, the time complexity is C<<G, where C is the number of chromosomes, G is the number of generations involved in calculating the total cost, R is the number of tuples and A is the number of features. In general this is again OnÞ.

Mutual Information of Selected Feature Before and After Watermarking

| Sr no | Name of features | $MI_O$ | $MI_W$ | $\Delta_{MI}$ |
|---|---|---|---|---|
| A1 | Age | – | – | – |
| A2 | Sex | 0.1175 | 0.1175 | 0 |
| A3 | cp | 0.3919 | 0.3919 | 0 |
| A4 | trestbps | 1.9705 | 1.9705 | 0 |
| A5 | chol | 3.8996 | 3.8996 | 0 |
| A6 | fbs | 0.1113 | 0.1113 | 0 |
| A7 | restecg | 0.1763 | 0.1763 | 0 |
| A8 | thalach | 3.0698 | 3.0698 | 0 |
| A9 | exang | 0.1396 | 0.1396 | 0 |
| A10 | oldpeak | 0.4373 | 0.4373 | 0 |
| A11 | slope | 0.1925 | 0.1925 | 0 |
| A12 | ca | 0.4259 | 0.4259 | 0 |

**Chart 2 :Mutual Information**

An optimum value of b ¼ 0.29 is calculated by using an optimization scheme. This value might be different for different databases. We have performed sufficient number of experiments to find the most reliable set of parameters for the genetic algorithm. The detailed set of GA parameters found reliable are given in Table 2. The preprocessing phase, takes approximately 1.7 milliseconds for computing an optimum value and a watermark string for the entire medical dataset . It demonstrates the time incurred in this phase for increasing dataset sizes (number of tuples) where l was kept fixed at 16-bits and the number of features were 14.

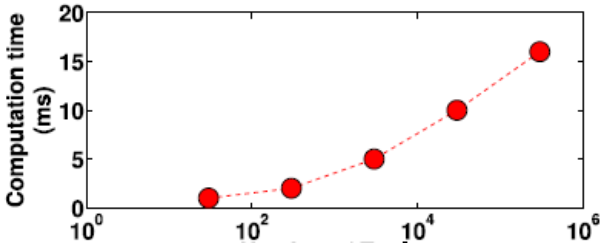


**Chart-3: Genetic Algorithm**

## 5. CONCLUSION

Watermarking techniques make changes in the data to such an extent that data quality gets compromised. Reversible watermarking techniques are used to cater to such scenarios because they are able to recover original data from watermarked data and ensure data quality to some extent. Irreversible watermarking techniques make changes in the data to such an extent that data quality gets compromised. Reversible watermarking techniques are used to cater to such scenarios because they are able to recover original data from watermarked

data and ensure data quality to some extent. However, these techniques are not robust against malicious attacks particularly those techniques
that target some selected tuples for watermarking. In this paper, a novel robust and reversible technique for watermarking numerical data of relational databases is presented. The main contribution of this work is that it allows recovery of a large portion of the data even after being subjected to malicious attacks. RRW is also evaluated through attack analysis where the watermark is detected with maximum decoding accuracy in different scenarios. A number of experiments have been conducted with different number of tuples attacked. The results of the experimental study show that, even if an intruder deletes, adds or alters up to 50 percent of tuples, RRW is able to recover both the embedded watermark and the original data. RRW is compared with recently proposed state-of-the-art techniques such as DEW, GADEW and PEEW to demonstrate that RRW outperforms all of them on different performance merits.

## REFERENCES

[1] Y.-C. Liu, Y.-T. Ma, H.-S. Zhang, D.-Y. Li, and G.-S. Chen, "A method for trust management in cloud computing: Data coloring by cloud watermarking," Int. J. Autom. Comput., vol. 8, no. 3, pp. 280–285, 2011.
[2] I.J.Cox, M. Miller, J. Bloom, and M. Miller, Digital Watermarking. Burlington, MA, USA: Morgan Kaufmann, 2001.
[3] P. W. Wong, "A public key watermark for image verification and authentication," in Proc. IEEE Int. Conf. Image Process., 1998, vol. 1, pp. 455–459.
[4] P. W. Wong and N. Memon, "Secret and public key image watermarking schemes for image authentication and ownership verification," IEEE Trans. Image Process., vol. 10, no. 10, pp. 1593–1601, Oct. 2001.
[5] F. A. Petitcolas, "Watermarking schemes evaluation," IEEE Signal Process. Mag., vol. 17, no. 5, pp. 58–64, Sep. 2000.
[6] J. T. Brassil, S. Low, and N. F. Maxemchuk, "Copyright protection for the electronic distribution of text documents," Proc. IEEE, vol. 87, no. 7, pp. 1181–1196, Jul. 1999.
[7] R. Agrawal and J. Kiernan, "Watermarking relational databases," in Proc. 28th Int. Conf. Very Large Data Bases, 2002, pp. 155–166.
[8] R. Sion, M. Atallah, and S. Prabhakar, "Rights protection for categorical data," IEEE Trans. Knowl. Data Eng., vol. 17, no. 7, pp. 912–926, Jul. 2005.
[9] S. Subramanya and B. K. Yi, "Digital rights management," IEEE Potentials, vol. 25, no. 2, pp. 31–34, Mar.-Apr. 2006.