A Reversible Watermarking Technique for Ownership Protection and Data Recovery

Kanchan S. Rahinj[1], Dipak R. Patil[2]
[1]*Student M.E IT, AVCOE sangamner, Maharashtra, India*
[2]*Assistant Professor, Department of Information technology AVCOE Sangamner, India*

ABSTRACT

Watermarking is commonly used technique in which, information is enforced to prove ownership on different types of data such as image, audio, video, text and relational database. There are many watermarking techniques for databases. Social network data are being mined for extracting knowledge and patterns. Such data are composed by different researchers and organizations and this are usually also shared via different channels. These data is usually big in volume because there are millions of social network users all over the world. Ownership protection of such data with huge volume becomes relevant. Digital watermarking is a more demanding solution than any other technique for ensuring rights protection and integrity of the original data sets. The leading objective is to devise a reversible watermarking technique for the social network data to prove ownership rights and also provide a mechanism for data recovery. Robustness of the proposed technique is evaluated through attack analysis using experimental study

Keyword : - *Database, digital watermarking, reversible watermarking, integrity of relational data.….*

1. INTRODUCTION

The use of social networks is increasing day by day. Recent statistics show that there were more than 32 million social network users worldwide [1]. The recent trends of the people all over the world indicate that this number would keep on increasing. The social networks' users perform different actions such as sharing their activities and news about their events, sharing news, sharing their personal data, and discussing various topics. Consequently, they are generating "Big Data" for social network computing, i.e., an essential component of Cyber, Physical, and Social Computing (CPSCom). To extract and analyze useful information from these data, data mining and other knowledge extraction algorithms are applied. Now a day's internet is offering wide range of web services that includes database as a service, digital repositories and libraries, online decision support system, e-commerce etc. As a result there are some problems occurs such as forgery, piracy, illegal redistribution, ownership claiming etc. The solution on these problems is Digital Watermarking. Digital Watermarking is the technique that used to protect digital data by hiding some information into original data.

The term "Digital Watermark" was introduced by Andrew Tirkel and Charles Osborne in December 1992. Andrew Tirkel, Charles Osborne and Gerard Rankin demonstrate the first successful embedding and extraction of a steganographic spread spectrum watermark. A watermark is considered to be some type of information that is embedded into original data for tamper detection, localization, ownership proof, traitor tracing etc [4].

Initially, most of work of watermarking is on still images audio and video, but now a day's watermarking of relational database becoming popular because of its increasing utility in many real life applications. The idea to make safe a database of map information (represented as a graph) by digital watermarking technique was initially coined by Khanna and Zane. Agrawal et al. proposed the scheme of digital watermarking for relational database [3].

2. LITERATURE SURVEY

From our knowledge, no one technique has been planned for modelling usability constraints for watermarking data mining datasets. But there are some techniques define bellows.

2.1 Technique of Watermarking Numeric Attributes in a Database:

In the work of Agrawal et al. the _rst well famous technique for watermarking numeric attributes in the database has been proposed. In that technique, message authenticated code (MAC) is calculated by helping of a secret key to dentify the candidate tuples.[6]

1. In these, it is one of the Watermarking techniques of numerical data.

2. This Technique is highly dependent on a secret key.A

3. It Uses markers to trace tuples to hide watermark bits. And it hides that watermark bits in the least significant bits.

Disadvantages:

1. No condition of multi-bit watermark and all operations are dependent only on the secret key.
2. No resilient to alteration attacks. Least Significant Bits (LSBs) can be easily manipulated by simple numerical alterations that are Shift LSB bits to the right/left.
3. Requires the attendance of a primary key in the watermarked relation.
4. Does not handle other usability constraints such that Category preserving usability constraints

2.2 Rights Protection for Relational data:

Sion et al. presented marker tuples based watermarking technique for relational databases, but these techniques are inapplicable to data mining datasets because they do not aim at preserving the knowledge contained in the dataset. Protecting rights over relational data is of still increasing interest, especially considering areas where responsive, valuable content is to be outsourced. A simple and better example is a data mining application, where data is put up for sale in pieces to parties specialized in mining it. Different opportunities for rights protection are presented, each one with its own advantages and drawbacks. Enforcement by authorized means is usually not so effective in preventing theft of patented works, unless improved by a digital counter-part, for example watermarking. The main reason of Digital Watermarking is to keep certain content from unauthorized duplication and sharing by enabling provable ownership over the content. It has conventionally relied upon the availability of a great noise domain within which the object can’t be unchanged while retaining its essential properties A big challenge of watermarking is to embed an indelible mark in the object such that,

1. The insertion of the mark does not destroy the cost of the object that is the object is still useful for the intended purpose and

2. It is difficult for an adversary to eliminate or alter the mark away from detection without destroying the value of that object.

2.3 Watermarking Technique for a Partitioning Based Database:

Shehab et al. projected a partitioning based database watermarking technique. They modeled the process of watermark embedding as a constraint optimization problem and tested genetic algorithm (GA) and pattern search (PS) optimizers. They select PS because it is capable to optimize in concurrent But this method requires defining usability constraints manually and does not account for preserving the knowledge enclosed in the data mining datasets.[8,]

In comparison, the focus of our current work is on developing a formal model to define usability constraints for watermarking of data mining datasets in such a way that the watermark is not only robust but the knowledge contained in the dataset is also preserved. Furthermore, we also provide a mechanism to logically group the dataset into groups such that high ranked features might also be watermarked during watermarking. This is a significant enhancement because if only low ranked features are watermarked, an attacker can launch malicious attacks on low ranked features only without compromising the data quality to a great extent. In this context, our data grouping approach enables a data owner to embed a watermark in high ranked features as well while still satisfying the usability constraints imposed by our formal model. Last but not the least; we have significantly enhanced our recently proposed information-preserving watermarking scheme [10] for data mining datasets in such a way that it can now watermark any type of features numeric, nonnumeric.

3. PROPOSED SYSTEM:

Digital watermarking is a more demanding solution than any other technique for ensuring rights protection and integrity of the original data sets. The objective of this paper is to devise a reversible watermarking technique for the social network data to prove ownership rights and also provide a mechanism for data recovery. Robustness of the proposed technique is evaluated through attack analysis using experimental study.

The reversible watermarking of social network datasets has some modules such as, preprocessing, watermark encoding, watermark decoding, and data recovery.

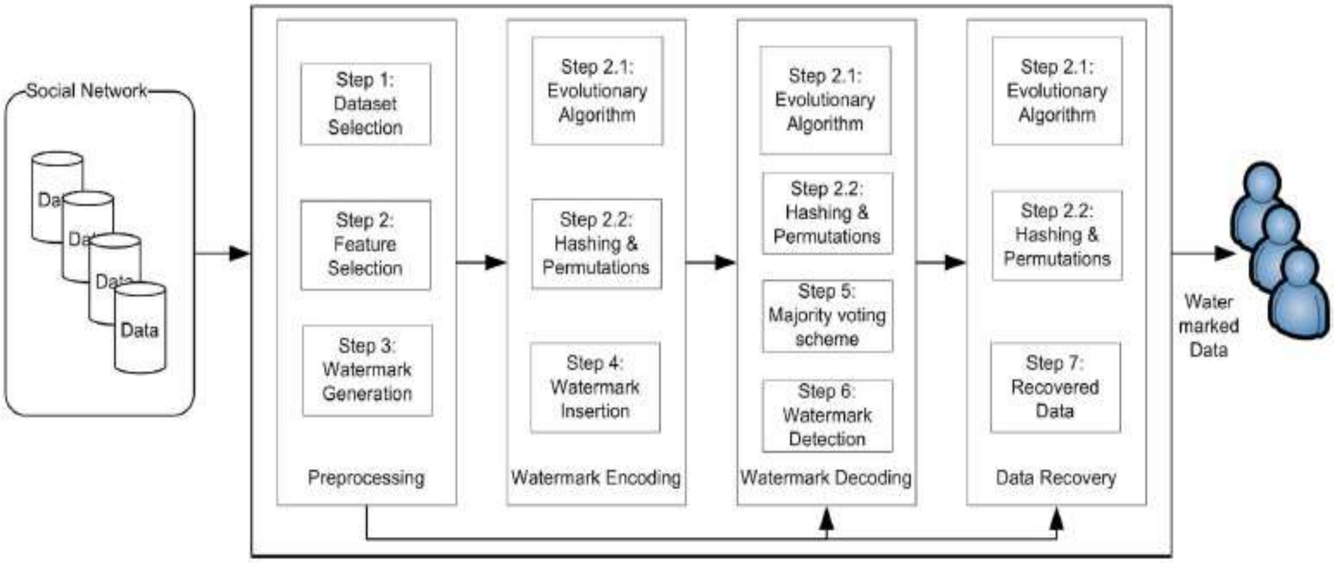

Social Network
Data
Step 1: Dataset Selection
Step 2: Feature Selection
Step 3: Watermark Generation
Preprocessing
Step 2.1: Evolutionary Algorithm
Step 2.2: Hashing & Permutations
Step 4: Watermark Insertion
Watermark Encoding
Step 2.1: Evolutionary Algorithm
Step 2.2: Hashing & Permutations
Step 5: Majority voting scheme
Step 6: Watermark Detection
Watermark Decoding
Step 2.1: Evolutionary Algorithm
Step 2.2: Hashing & Permutations
Step 7: Recovered Data
Data Recovery
Water marked Data

Fig. 1Main Architecture of the proposed technique

3.1 Preprocessing

The preprocessing includes 1) data selection; 2) feature selection; and 3) watermark creation. First, the dataset to be watermarked is selected. Next, the numeric or nonnumeric feature is chosen for watermarking. Then, a watermark is generated through a pseudorandom sequence generator to encode the selected feature of the selected dataset.

3.2 Watermark Encoding

After that selecting the feature from the dataset, two further steps are performed for each set of selected feature before encoding watermark. In the first step, an evolutionary algorithm, i.e., genetic algorithm (GA), is used to create an optimum value to be embedded in the numeric type of dataset for ensuring robust watermark detection. In the second step, Hashing and permutations are created for the nonnumeric type of dataset to make sure reversible watermarking. After calculating a seeded watermark in step 3, the watermark is embedded in each type of data in step 4. As shown in main architecture.

3.3 Watermark Decoding

In watermark detection from the watermarked data, first, the preprocessing steps, i.e., hashing and permutation, are performed again for selected type of feature. Next a majority voting scheme is used to detect the watermark from the marked dataset on the basis of the number of 1s and 0s. Finally, the watermark is extracted from the whole dataset to prove ownership.

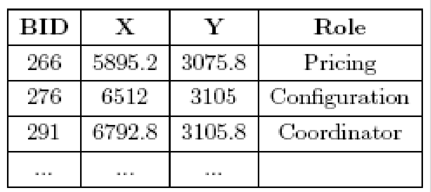
3.4 Data Recovery

In data recovery GA, hashing, and permutation steps are performed again for selected type of feature. Data are recovered from the marked data of the selected feature type through employing GA, hashing, and permutation after detecting the

4.Result Analysis

The proposed technique has been evaluated for providing: 1) reversible watermarking; and 2) robustness against malicious attacks. For brevity, experiments have been reported with a Badge dataset containing the performance and dynamics of a real-world organization. A relatively small watermark that consists of 8 bits is used in the analysis.

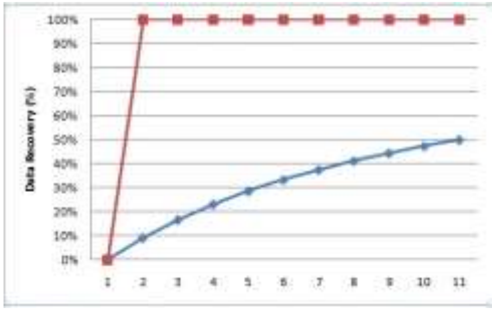
The dataset consists of four features, including 1) BID (badges identified by unique numbers assigned to the employees); 2) and 3) x and y (locations of employees cubicles and anchor nodes); and 4) roles of employees. The dataset has been shown in Table with only three records for brevity.



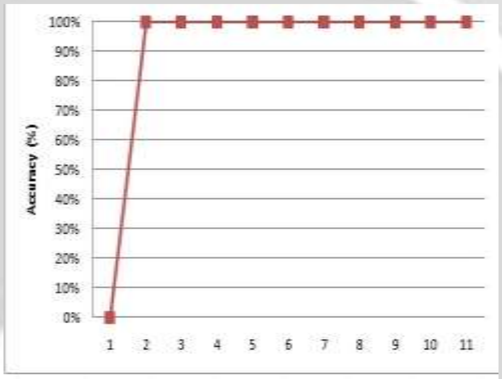
BID	X	Y	Role
266	5895.2	3075.8	Pricing
276	6512	3105	Configuration
291	6792.8	3105.8	Coordinator
...	...	...	

4.1Performance

The proposed technique handles big data in a systematic way with logical grouping. For instance, for nonnumeric data encoding, similar records for the same employees get hashed to same values and make logical groups; thus, the data size does not increase. Moreover, the data owner has more interest in acquiring ownership rights and data recovery. She gets her data watermarked only once, and watermarking is usually done offline; this is to say that it is done on the machine of the data owner and data are delivered to the recipient only after completing the embedding process. Thus, she can afford large computation time for providing ownership rights and data recovery. For datasets involving a large number of features or a large number of rows Big Data, the data owner may use a separate machine, with high computation power, to watermark the datasets. This might incur some cost but gain the owner more security (false claim of ownership can be tackled by watermark encoding and decoding).



Data Recovery rate



Data Accuracy

5.CONCLUSION

The proposed technique presented a mechanism for providing ownership rights over the digital data through digital watermarking. The proposed watermarking technique is not only robust but it also ensures original data recovery after watermark decoding. A formal method has been used to prove the effectiveness of the system. Experimental study has been performed for the evaluation of the proposed watermarking technique against the defined threat model.

6. REFERENCES

[1] A.Z.Tirkel, G.A. Rankin, R.M. Van Schyndel, W.J.Ho, N.R.A.Mee, C.F.Osborne, "Electronic Water Mark", *DICTA 93, Macquarie University*. pp. 666-673,1993.

[2] Raju Halder, Shantanu Pal, Agostino Cortesi, "Watermarking Techniques for Relational Databases: Survey, Classification and Comparison", *Journal of Universal Computer Science,* vol. 16, no. 21 , pp. 3164-3190, 2010.

[3] Khanna, S. and Zane, "Watermarking maps: hiding information in structured data", In *Proceedings of the 11th annual ACM-SIAM symposium on Discrete algorithms (SODA '00),* pp. 596–605, 2000.

[4] Agrawal, R. and Kiernan, J., "Watermarking relational databases" ,In *Proceedings of the 28th international conference on Very Large Data Bases (VLDB '02),* pp.155–166, 2002

[5] Agrawal, R., Haas, P. J., and Kiernan, J., "A system for watermarking relational databases", In *Proceedings of the 2003 ACM SIGMOD international conference on Management of data (SIGMOD '03),* pp 674–674, 2003.

[6] Agrawal, R., Haas, P. J., and Kiernan, J., "Watermarking relational data: framework, algorithms and analysis" , *The VLDB Journal,* 12 pp. 157–169, 2003

[7] Lafaye, J., "An analysis of database watermarking security", In *Proceedings of the 3rd International Symposium on Information Assurance and Security (IAS '07),* pp 462–467, 2007.

[8] Qin, Z., Ying, Y., Jia-jin, L., and Yi-shu, L. "Watermark based copyright protection of outsourced database", In *Proceedings of the 10th International Database Engineering and Applications Symposium (IDEAS'06)*, pp. 301–308, 2006.

[9] V. Khanduja and O. Verma, "Identification and proof of ownership by watermarking relational databases," *Int. J. Inf. Electron. Eng.*, vol. 2, no. 2, pp. 274–277, Mar. 2012

[10] W. Yanmin and G. Yuxi, "The digital watermarking algorithm of the relational database based on the effective bits of numerical field," in *Proc. WAC*, pp. 1–4, 2012.