NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY [IJIERT]
ISSN: 2394-3696
VOLUME 4, ISSUE 8, Aug.-2017

LITERATURE REVIEW ON WATERMARKING TECHNIQUES OF RELATIONAL DATABASES

KANCHAN S. RAHINJ
Student M.E IT, AVCOE sangamner, Maharashtra, India, kanchanrahinj@ gmail.com

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. These paper describe, History of watermarking, Applications of digital watermark for relational databases, Types of attacks on watermark database, and most important existing watermarking techniques. Social network data are being mined for extracting knowledgeable 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. In this context, 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 objective of this paper is to study work done in watermarking on relational data until now.

KEYWORDS: database, digital watermarking, reversible watermarking, integrity of relational data.

INTRODUCTION

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].

Generally, there are two main phases of database watermarking,

1) Watermark Embedding
2) Watermark Verification

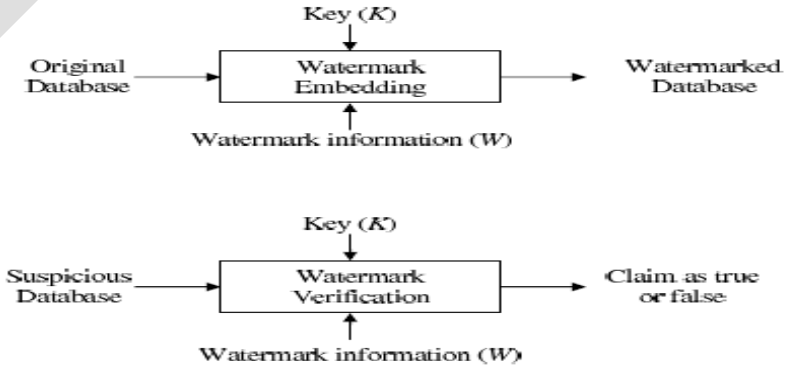

Fig. 1 Basic Watermarking Technique

In watermark embedding, a private key K, which is only known to the owner, is used to embed watermark W into the original database then the output of these phase is publically available. The verification is performed to verify the ownership of suspicious database. In this phase, suspicious database is taken as input and with the help of private key K the embedded watermark is extracted. Figure 1 shows the basic database watermarking technique [2].

APPLICATIONS OF DIGITAL WATERMARK FOR RELATIONAL DATABASES

There are some applications of Digital Watermark for relational databases, such as

A Ownership Assertion: Digital watermark can be used for ownership assertion. In this application Alice can embed a watermark into database R by using private parameter i.e. secret key. Then she can make data publically available. If she notices the relation S published by Mallory has been pirated from her relation R. the set of tuples and attributes in S can be a subset of R

B Fingerprinting: Fingerprinting aims to identify a defector in the applications where content of database is publicly available all over the network, the content vendor would like to depress unauthorized duplication and sharing by embedding a distinct watermark or fingerprint in every copy of the database content. If, unauthorized copies of the database are found, then the origin of the copy can be determined by retrieving the fingerprint.

C Fraud and Tamper Detection: If database content is used for very critical applications like commercial transactions or medical applications, then it is important to make sure that the content was originated from a specific source and that it had never been changed, manipulated or falsified. This can be achieved by embedding a watermark in the original data of the database [2].

CLASSIFICATION OF WATERMARKING TECHNIQUES

In this paper, I try to cover the details of various watermarking techniques. There are many techniques in watermarking. Here I make survey on classification of watermarking techniques based on:

1) **WATERMARK INFORMATION:** Different watermarking methods embed different types of information into the database for example image, text, sound etc. in watermarking system information which is used to embed is very important therefore classification is based on information which is used to embed.

2) **DISTORTION:** in these, watermarking either distortion based or distortion free. In distortion based watermarking slight changes in original data may occur while embedding watermark. The change must be tolerable and should not make data useless. And in distortion free watermark, watermark embedding phase is independent on types of attribute. And any distortion is not happen in original data.

3) **COVER TYPE: i**n this, watermarking is classified based on type of cover i.e. type of attributes in which watermark is embedded.

4) **GRANULARITY LEVEL:** this is another classification parameter of watermarking. Watermarking can be performed by modifying or inserting information at bit level or higher level such as character level, attribute level and tuple level.

5) **VERIFIABILITY:** verification process may be deterministic or probabilistic in nature. It can be performed blindly; also it can be performed publicly or privately.

6) **INTENT:** various schemes of watermarking are designed various purposes namely, integrity, temper detection, ownership proof, traitor detection localization etc. [2].

TYPES OF ATTACKS ON WATERMARK DATABASE

Fragile watermarking means the digital watermarking for integrity verification. For detection, modification and localize embedded watermark should be fragile. Digital watermarking used for copyright protection is

called robust watermarking; embedded watermark should be robust against various attacks which aim at removing watermark [2].
Some intentional or unintentional attacks can damage and erase the watermark. These attacks are as follows:

A. BENIGN UPDATES

Any watermark relation tuples or data processed in this case, result of that marked tuples may be added, deleted or updated which may cause embedded watermark not detectable or may remove embedded watermark. This is unintentionally performed processing.

B. VALUE MODIFICATION ATTACK

1) Bit attack: By changing and altering one or more bits in watermarked data this attack attempts to destroy. In this case usefulness of data is important. The completely useless after more alternation. This attack takes place randomly so called as randomisation.
2) Rounding Attack: By rounding all values of attribute Mallory may try to lose the marks contained in a numeric attribute. Rounding attacks depends on estimation of number of bit position involved. Attack gets unsuccessful because of underestimation and overestimation may cause the data useless.
3) Transformation: it is attack in which numeric values are linearly transformed. For example, Mallory may convert the data to different unit of measurement. Mallory increase on trust among user.

C. SUBSET ATTACK

Mallory may think that a subset of tuples or attributes of watermark relation. He may hope that watermark has been lost after attacking.

D. SUPERSET ATTACK

New tuples or attributes are added to watermark database, which affects the correct detection of watermark.

TYPES OF WATERMARK

A) Division Based on human perception:
There are two types
1) Visible Watermark: This watermark can be clearly seen by users. It changes signal. Watermarked signal and original signal is not same.
2) Invisible Watermark:: this watermark not seen by users. Output signal and original signal is same

B) Division Based on Application: This is divided into three types
1) Fragile Watermark: These watermarks are so sensitive. They can be destroyed with small amount of change in watermark signal.
2) Semi Fragile Watermark: These watermark are destroy if changes in watermark signal exceed a predefined user threshold. If threshold is equal to zero, then it used as fragile watermark.
3) Robust Watermark: This cannot destroy easily. Therefore this method is usually used in copyright protection.

C) Division Based on Level of Information Required to Detect the Embedded Data
In this there are three types
1) Blind Watermark
2) Semi-Blind Watermark
3) Non Blind Watermark

D) Division Based on Knowledge of the User on the Presence of Watermark
This is sub-divided into two types of watermarking
1) Steganographic Watermark: The user is not aware of the presence of the watermark.
2) Non-steganoghaphic Watermark: The user is aware of the presence of the watermark

EXISTING WATERMARKING TECHNIQUES

In this Section, there is detailed about watermarking techniques proposed so far. Here we sort out the techniques based on (i) whether marking introduces any distortion, (ii) the type of the underlying data (cover) in which watermark information is embedded, and (iii) Type of the watermark information to be embedded. Depending on whether the marking introduces any changes in the original data of the database, the watermarking techniques can be divided into two: Distortion-based and Distortion-free.

A) DISTORTION-BASED WATERMARKING

(A)WATERMARKING BASED ON NUMERICAL DATA TYPE ATTRIBUTE

The watermarking schemes proposed by Agrawal et al. also known as AHK algorithm is depend on numeric data type attribute and marking is made at bit-level. The vital idea of these schemes is to guarantee that some bit positions for some of the attributes of some of the tuples in the relation contain specific values. This bit pattern assigns the watermark. The tuples, attributes inside a tuple, bit positions in an attribute, and this specific bit values at are algorithmically determined by private parameters γ, ν, ξ and K which are only known to the owner of the relation. The parameters γ, ν, ξ and K stand for number of tuples to mark, number of attributes, number of least significant bits and secret key respectively. [4], [5], [6].

The cryptographic MAC function H(K||H(K||r.P)) where r.P is the primary key of the tuple r and || represents concatenation operation, is used to verify candidate bit positions. The HASH function H(K||r.P) is used to find out the bit values to be embedded at those positions. The selection of MAC and HASH is due to the one-way functional characteristics and less collision probability [4].

In AHK algorithm authors apply pseudorandom sequence generator (e.g., Linear Feedback Shift Register) as a substitute of HASH and MAC to identify the marking bits and mark positions. The security and robustness of this scheme relies on these parameters which are totally private to the owner. The algorithm used for watermark detection is blind and probabilistic in nature. The relation is measured as pirated if the matching pattern is present in at least τ tuples, where τ depends on the actual number of tuples marked and a preselected value α, called the significance level of the test. Examine that the achievement of watermark detection phase depends on the fixed order of attributes. Re-sort of attributes' order may defer to the detection phase almost infeasible. Although the main statement of this scheme is that the relation has primary key whose value does not change, they also propose an alternative to treat a relation without primary key. [5], [6].

Lafaye describes the security analysis for the AHK algorithm in that he analyzes the security and robustness in two conditions: (i) Multiple Keys Single Database (MKSD): A single database is watermarked several times using different secret keys and sold to different users, and (ii) Single Key Multiple Databases (SKMD): Different databases are watermarked by using a single secret key. An attempt of arbitrary attack on a watermarked content obtained by the AHK algorithm, it may be successful when randomize the ξth least significant bits of all tuples of the relation. On the other hand, this attack is highly invasive since most values of the relation are impacted by the attack. The locations guessed based on MKSD and SKMD can be used to build a better focused attack [7].

Qin et al propose an improvement over the Agrawal and Kiernan's scheme. As a replacement for using hash function, they use chaotic random series based on the Logistic chaos equation having two properties i.e. the non-repetitive iterative operation and the sensitiveness to initial value. It ignores the inherent limitation of collision of Hash function. The choice of bits of LSB for embedding watermark meets the equirements of data range and data precision of each attribute, rather than simply to use a same ξ for all attributes. So the error caused by watermark is decreased significantly [4], [8].

Among the most current works, Gupta et al. propose a reversible watermarking scheme which is the modified version of Agrawal and Kiernan's one. In this method, at the detection phase, the original unwatermarked version of the database can be recovered along with the ownership proof. The operation first extracts a bit called OldBit from the integer portion of the attribute value before replacing it by the watermark bit and inserts it in the little portion of the attribute value. Therefore, the watermark bit can be recovered during detection and the attribute can be restored to its unmarked value by replacing the watermark bit with the original bit OldBit extracted from the small part. They also propose a different

algorithm to overcome any attempt of additive or secondary attack which relies on the apparent fact that the database relation must be watermarked by the actual vendor before Mallory [9].

The watermarking technique embeds random digits in between 0 to 9 at LSB positions of the candidate attributes for some algorithmically selected tuples. At the time of embedding phase, the tuples are securely partitioned into groups by using the cryptographic hash function and only the first m which is equal to the length of the owner's watermark groups are considered. The assessment whether to mark ith ($1 \leq i \leq m$) group depends on the ith bit of the owner's watermark, while the selection of the tuples in a group is based on a secret key (which is different from that used during partitioning) also the information at second LSB positions of the numeric candidate attributes. Lastly, for the selected tuples random numbers (between 0 and 9) are embedded at LSB positions in the attribute values of those tuples. Observe that even if the owner has a watermark of length m, it is not actually embedded. Rather, it is used to identify some valid groups to embed the random values which act as embedded watermark information. The detection phase determines the existence of mark in a group if the maximum amount frequency for a value between 0 and 9 for that group exceeds a threshold [10].

(B) IMAGE AS WATERMARK INFORMATION.

Wang et al. explain an image-based watermarking scheme a scrambled image based on Arnold transform with scrambling number d is used as a watermark instead of embedding original image. Because Arnold transform of an image has the periodicity P, the result which is obtained in the extraction phase can be recovered from the scrambled form to the original after (P – d) iterations. In embedding phase, the original image of size $N \times N$ is first transformed into scrambled image which is then represented by a binary string bs of length $L = N \times N$. Secondly, all tuples in the relation are grouped into L groups. The hash value which is computed by using tuple's primary key, secret key and order of the image, determines the group in which every tuple belongs. Lastly, the ith bit of bs is embedded into the algorithmically selected bit position of the attribute value for those tuples in ith group that assure a particular criteria. The detection phase imitates majority voting technique. Though, the security of this scheme improves as it relies on the secret key and the scrambling number d and the order of the image N [11].

Instead of embedding scrambled image, the watermarking technique embeds the original image by initially converting it into a bit flow (EMC, Encrypted Mark Code) of certain length, and then by following same algorithmic steps as in [10]. The just two differences are that (i) the watermark insertion technique in [10] assumes single fixed attribute to mark for all tuples where [11] does not, and (ii) during choice of bit positions, the order of the image is not measured. At last after marking, it checks the usability of the data with respect to the intended use. If acceptable, the change is committed, otherwise rolled back [12].

Another watermarking system to embed image in BMP format is presented. In this watermark embedding phase, the BMP image is sub-divided into two parts i.e. header and image data. An error correction approach of BCH (Bose-Chaudhuri-Hocquenhem) coding is used to encode the image data part into watermark. Depending on the tuple's ID value that is calculated by performing hashing function parameterized with tuple's primary key and BMP header, each tuples are allocated to k distinct subsets, where k is the watermark length. Lastly, all of the k bits of the watermark is used to mark each of the k subsets of tuples. During the marking, the elected least significant bit positions of selected attributes of some specific tuples satisfying significant criteria are altered. The selection of bit positions rely on HASH or MAC function parameterized with BMP header, tuple's primary key and other parameters like number of least significant bits in the attribute value. Monitor that the selected bit positions are not set to the watermark bits directly but rather, are set to mask bits which are computed from both the hash value and the watermark bit together [13].

The image-based fragile watermarking technique aims at maintaining honesty of the database and uses support vector regression (SVR) to train high correlation attributes to produce the SVR predicting function for embedding watermark into particular numeric attributes. This system consists of three phases: (i) Training Phase: Select training tuples and get trained SVR predicting function; (ii) Embedding Phase: All tuples in the relation are used to insert image as watermark where the number of watermark bits is designed to be equal to the number of tuples. Each numeric attribute value Ci of ith tuple ti is estimated using the SVR prediction function f resulting $C^{-}i = f(ti)$. Depend on the ith watermark bit bi (obtained after

converting the image into bit flow), the value of Ci is customized by $C^{-}i + 1$ or $C^{-}i - 1$; (iii) Tamper Detection: The trained SVR predicting function is used to create the predicted value for each tuple and compared with the value contained in the database. The difference of these values calculates the watermark information and can make sure whether database is tampered or not. Yet, the limitation of this scheme is that it can identify the modification which takes place in the objective attribute set only. This technique work good in the situation where the tuples in the table are independent but highly correlated between the attributes [14].

(C) SPEECH AS WATERMARK INFORMATION.

Wang et al. propose the use the owner's speech to create distinct watermark. The preparation of watermark from the speech made up of several stages: compression of speech signal to compressed the watermark, speech signal enhancement to eliminate noise in frequency domain, speech signal transfer into bit stream, and finally, watermark production by using the message of the copyright of the holder and the product of the converted speech signal. The bit-level marking is performed at the watermark embedding phase by following the same algorithmic steps as in image-based technique of [12],[15].

(D) GENETIC ALGORITHM BASED WATERMARK SIGNAL.

The authors intend Genetic Algorithm-based technique to generate watermark signal, focusing on the optimization issue. They follow the same algorithmic framework as of [12].

(E) CONTENT CHARACTERISTICS AS WATERMARK INFORMATION.

The watermarking techniques are performed based on the content of the database itself. The watermark embedding phase extracts some bits, called local characteristic, from the characteristic attribute A1 of tuple t and embeds those bits in the watermark attribute A2 of the same tuple. The choice of tuples depends on the generated random value which is in between 0 and 1 is less than the embedded proportion α of the relational databases and the non-NULL requirement of characteristic attribute value. In the watermark detection phase, similar procedure is followed to get the local characteristic of the characteristic attribute and compared against the last bits of watermark attribute [17].

The authors propose a fragile watermarking method that can verify the honesty of database relation. In the proposed system, all tuples in a database relation are initially securely divided into groups and sorted. In each group, there are two types of watermarks to be embedded: attribute watermark W1 and tuple watermark W2 which consists of γ watermarks of length ν and ν watermarks of length γ respectively where γ and ν are the number of attributes in a tuple and average number of tuples in each group respectively. W1 and W2 are formed by extracting bit sequence from the hash value. For attribute watermark W1, the hash value is created according to the message authentication code and the same attribute of all tuples in the similar group, at the same time for tuple watermark W2, it is produced from the same message authentication code and all attributes of the same tuple. Examine that, in both the embedding and detection phases, they ignore the least two significant bits of all attributes of numeric type except the primary key when calculating hash values. The attribute watermark is embedded in LSB, where the tuple watermark is embedded at next to the LSB level. In that manner, the embedded watermarks actually form a watermark grid, which helps to detect, localize and characterize modifications [18].

(F) OTHER MEANINGFUL WATERMARK INFORMATION

The partitioning of tuples in most of the techniques is based on hashing. Huang et al. in its place, propose the use of well-known techniques (e.g. k-means algorithm) to cluster the tuples into some of the same classes. The insertion of the watermark bit is based on the comparison of the parity of watermark bit and the LSB of candidate attribute. The k-means scheme assures the location of the embedded watermark irregular [19].

The watermark insertion phase in facilitates in three phases: (i) select a group of candidates in all attributes of the relation, and record it as the watermark schema; (ii) append the error correction code (ECC Code) to the watermark; (iii) executes the watermark insertion algorithm. The insertion algorithm generates pseudo random sequence using primary key and secret key. This pseudo random sequence is used to identify the

attribute to mark based on the significance of the attributes and the watermark bit to be embedded. At the time of marking, the local constraint and a bidirectional mapping (to reduce watermarking data of various types into numeric data) are used. The local constraints are defined as the upper bounds of "the distance" of attributes after/before watermarking. Finally, global constraints are evaluated to decide whether to commit the changes. Observe that watermark schema selection in watermark insertion phase and watermark schema detection in verification phase exploit the non-blindness property [20].
Watermarking scheme which follows same algorithmic steps as of but embeds other significant watermark information rather than image by initially converting it into a bit flow of certain length. The selection of the candidate attribute is based on the weights of all numeric attributes with a different hashing function. Observe that in watermark insertion stage of [22] the mark position is determined by using the mark bit [23].

B) DISTORTION-FREE WATERMARKING

Many of the distortion free watermarking techniques are fragile means in addition to the ownership claiming, they aim at maintaining the reliability of the information in the database. The watermark embedding phase does not based on any specific type of attribute and does not introduce any distortion in the basic data of the database

(A)EXTRACTING HASH VALUE AS WATERMARK INFORMATION

In order to attain the idea of fragile watermark, authors proposed watermarking schemes that are able to perceive any modifications made to a database relation. These systems are designed for categorical data that cannot accept distortion; therefore, the watermark embedding is distortion free. Partitioning of tuples is depend on the hash value parameterized with primary key and secret key [23]. whereas partitioning is depend on categorical attribute values. After partitioning, the tuple level and group level hash values for each group are calculated. A watermark length is equal to the number of tuple pairs in the group, is extracted from the group level hash value and for every tuple pair, the order of the two tuples are changed or unchanged according to their tuple hash values and the equivalent watermark bit [24]. In addition, Li suggests to perform the interchange of tuples' positions based on Myrvold and Ruskeys linear permutation unranking algorithm to enhance the embedding capacity. In these schemes, any modification of an attribute value will affect the watermarks in two groups as the modified tuple may be removed from one group and be added to the other group [24], [25],[26].
The system proposed by Tsai et al. aims at maintaining the integrity of the information in the database and is depend on public authentication mechanism. The design behind of this scheme is that, firstly, watermark W is created which is a $\sqrt{n} \times \sqrt{n}$ white image, whereas n is the no. of tuples in the relation, besides four corners having mark of the owner. It generates a value Ci ($0 \leq Ci \leq 255$) for each tuple ti in the database using hash function MD5 and XOR operation. If there are n tuples in the database, it creates a feature C of length n by combining all Ci in order. Finally, a certification code R is created by XOR-ing C and W. The encrypted form of R using private key is made available widely. At the time of verification phase the integrity of the relation T′, similarly, it creates feature C' from T'. After decryption by using public key the certification code R is XOR-ed with C' that yield the watermark W'. The honesty of this extracted watermark proves the integrity of the database [27].

(B) CONVERTING DATABASE RELATION INTO BINARY FORM USED AS WATERMARK INFORMATION

The public watermarking scheme by Li and Deng is relevant for marking any type of data including integer numeric, real numeric, character, and Boolean, without fear of any error constraints. The interesting features of this scheme are that it does not use any type of secret key and can be verified publicly as many times as necessary. The distinctive watermark key, used in both creation and verification phase, is public and obtained by one-way hashing from different information like the Identity of the owner(s) and characteristics of the database (e.g. DB Name, Version etc.). Observe that the public watermark key is different from the public-private key pair of asymmetric cryptography. This watermark key is used to create a watermark W from the relation R. The watermark W is a database relation whose schema is W(P,W0, ...,Wγ−1), where

$W0, \ldots, W\gamma-1 \in \{0, 1\}$. Compared to database relation R, the watermark W has the same number η of tuples and the same primary key attribute P.

The number γ of binary attributes in W is a control parameter that determines the number ω of bits in W , where $\omega = \eta \times \gamma$ and $\gamma <$ number of attributes in R. In the algorithm, a cryptographic pseudorandom sequence generator to randomize the order of the attributes and the MSBs of the attribute values are used for creating the watermark W. The use of MSBs is for thwarting potential attacks that alter the data. Where the watermark key K, the watermark W, and the algorithm are publicly known, anyone can locate those MSBs. Any modification to these MSBs introduces intolerable errors to the underlying data and can simply be captured at the time of verification phase. However, alteration of other bits in the data cannot be detected by this scheme [28].

The fundamental difference is that the former considers a private key instead of public, and thus, cannot be publicly verifiable. In addition, the former is partition based and considers the extracted binary watermark as an image which is used to ownership protection. This image is acted as the abstract counterpart of the concrete relation R, and the abstraction is sound in the sense that concretization of the abstract image must cover R. However, the disadvantage of this scheme is that the extracted image may not have any meaningful pattern [28].

The authors address the issue of persistency of watermarks, which serves as a method to recognize the integrity and ownership proof of the database while allowing the evaluation of the database by queries in a set of queries Q. The persistency of the watermark is preserved by exploiting the invariants (Static Part and Semantics-based Properties of the underlying data in the database w.r.t. Q) of the database states. The watermarking algorithms are designed as an enhancement of the proposal by Li and Deng in terms of fragileness and persistency [29], [30].

(C) PUBLIC VS. PRIVATE WATERMARKING TECHNIQUES

Most of the existing watermarking techniques in the literature are private, meaning that they are depending on some private parameters (e.g. a secret key). Only the authorized people (e.g. database owners) know these private parameters are able to verify the watermark and prove their ownership of the database in case of any illegal redistribution, false ownership claim, and theft etc. However, private watermarking techniques go through from disclosure of the private parameters to dishonest people once the watermark is verified in presence of the public. With access to the private parameters, attackers can easily invalidate watermark detection either by removing watermarks from the protected data or by adding a false watermark to the non-watermarked data. In contrast, in case of public watermarking techniques [27], [28], [30], any end-user can verify the embedded watermark as many times as necessary without having any prior knowledge about all of the private parameters to ensure that they are using correct (not tampered) data coming from the original source. For instance, when a customer uses sensitive information such as currency exchange rates or stock prices, it is very important for him to ensure that the data are correct and coming from the original source.

A watermarking-based color image authentication with both detection and recovery capability is proposed, in this a Halftone image is used as an approximated version of the luminance channel (Y) and a coded version of the coefficients of the Two Dimensional Discrete Cosine Transform of the chrominance channels (Cb, Cr) of the YCbCr color space are used as watermark signal [33].

A number of recent techniques such as those in [31]–[32] extend the work reported in [4] and embed a multiple bit watermark in selected least significant bits.

The focus of all of the abovementioned techniques is toward the watermarking of relational databases, and almost all of the techniques require a primary key for watermarking. However, often (if not always), there is no primary key or any other unique feature in social network data sets. Consequently, we consider our work to be novel.

PROPOSED TECHNIQUE

This section discusses the reversible watermarking of social network datasets. It has some modules such as preprocessing, watermark encoding, watermark decoding, and data recovery. The objective of this technique is to devise a reversible watermarking technique for the social network data to prove ownership rights and also provide a mechanism for data recovery**.**

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.

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.

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 "1's" and "0's". Finally, the watermark is extracted from the whole dataset to prove ownership.

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 embedded watermark.

CONCLUSIONS

In this paper we survey the current state of the different watermarking techniques for relational databases. In this paper, we studied the classification of watermarking techniques on the basis of various parameters. Also we studied different types of attacks. We studied the various types of watermarking techniques on the basis of various parameters like Human Perception, Robustness etc. This paper describes study of existing techniques of watermarking on relational database. Finally, we observe that the usability of the watermarked database and queries still remains an open issue for future research.

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) Gupta, G. and Pieprzyk, J., "*Database relation watermarking resilient against secondary watermarking attacks", In Proceedings of the 5th International Conference on Information Systems Security (ICISS '09), pp. 222–236, 2009.*

10) Xiao, X., Sun, X., and Chen, M., "*Second-lsb-dependent robust watermarking for relational database", In Proceedings of the 3rd International Symposium on Information Assurance and Security (IAS '07), pp. 292–300, 2007.*

11) Wang, C., Wang, J., Zhou, M., Chen, G., and Li, D., "*Atbam: An arnold transform based method on watermarking relational data", In Proceedings of the 2008 International Conference on Multimedia and Ubiquitous Engineering (MUE '08), pp. 263–270, 2008.*

12) Hu, Z., Cao, Z., and Sun, J., "*An image based algorithm for watermarking relational databases", In Proceedings of the 2009 International Conference on Measuring Technology and Mechatronics Automation (ICMTMA '09), pp. 425–428, 2009.*

13) Zhou, X., Huang, M., and Peng, Z., "*An additive-attack proof watermarking mechanism for databases copyrights protection using image", In Proceedings of the 2007 ACM symposium on Applied computing (SAC '07), pp. 254–258, 2007.*

14) Tsai, M., Hsu, F., Chang, J., and Wu, H., *"Fragile database watermarking for malicious tamper detection using support vector regression", In Proceedings of the 3rd International Conference on International Information Hiding and Multimedia Signal Processing (IIH-MSP '07), pp. 493–496, 2007.*

15) Wang, H., Cui, X., and Cao, Z., "*A speech based algorithm for watermarking relational databases", In Proceedings of the 2008 International Symposiums on Information Processing (ISIP '08), pp. 603–606, 2008.*

16) Meng, M., Cui, X., and Cui, H., *"The approach for optimization in watermark signal of relational databases by using genetic algorithms", In Proceedings of the 2008 International Conference on Computer Science and Information Technology (ICCSIT '08), pp. 448–452, 2008.*

17) Zhang, Y., Niu, X., Zhao, D., Li, J., and Liu, S., "*Relational databases watermark technique based on content characteristic", In Proceedings of the 1st International Conference on Innovative Computing, Information and Control (ICICIC '06), pages 677–680, 2006.*

18) Guo, H., Li, Y., Liua, A., and Jajodia, S., "*A fragile watermarking scheme for detecting malicious modifications of database relations", Information Sciences, pp. 1350–1378, 2006.*

19) Huang, K., Yue, M., Chen, P., He, Y., and Chen, X., "*A cluster-based watermarking technique for relational database", In Proceedings of the 1st International Workshop on Database Technology and Applications (DBTA '09),pages 107–110, 2009.*

20) Hu, T., Chen, G., Chen, K., and Dong, J., "*Garwm: Towards a generalized and adaptive watermark scheme for relational data", In Proceedings of the 6th International Conference in Advances in Web-Age Information Management (WAIM '05), pp. 380–391, 2005.*

21) Cui, X., Qin, X., Sheng, G., and Zheng, J., "*A robust algorithm for watermark numeric relational databases", In Proceedings of the 2010 International conference on Intelligent computing (ICIC '06), pp. 810–815, 2006.*

22) Guo, F.,Wang, J., Zhang, Z., Ye, X., and Li, D., *"An improved algorithm to watermark numeric relational data", In Proceedings of the 6th International Workshop on Information Security applications (WISA '05), pp. 138–149, 2005.*

23) Xinchun, C., Xiaolin, Q., and Gang, S., "*A weighted algorithm for watermarking relational databases" Wuhan University Journal of Natural Science, (1) pp. 79–82, 2007.*

24) Li, Y., Guo, H., and Jajodia, S., "*Tamper detection and localization for categorical data using in fragile watermarks", In Proceedings of the 4th ACM workshop on Digital rights management (DRM '04), pp. 73–82, 2004.*

25) Bhattacharya, S. and Cortesi, A., "*A distortion free watermark framework for relational databases", In Proceedings of the 4th International Conference on Software and Data Technologies (ICSOFT '09), pp 229–234, 2009.*

26) Li, Y., "*Database Watermarking: A Systematic View", Springer Verlag, 2007.*

27) Tsai, M., Tseng, H., and Lai, C., "*A database watermarking technique for temper detection", In Proceedings of the 2006 Joint Conference on Information Sciences (JCIS '06), 2006.*

28) Li, Y. and Deng, R. H., "*Publicly verifiable ownership protection for relational databases", In Proceedings of the 2006 ACM Symposium on Information, computer and communications security (ASIACCS '06), pp. 78–89, 2006.*

29) Halder, R. and Cortesi, A., "*A persistent public watermarking of relational databases", In Proceedings of the 6th International Conference on Information Systems Security (ICISS '10), pp. 216–230, 2010.*

30) Halder, R. and Cortesi, A., "*Persistent watermarking of relational databases", In Proceedings of the IEEE International Conference on Advances in Communication, Network, and Computing (CNC '10), pages 46–52, 2010.*

31) 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.*

32) 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.*

33) L. R. Roldan, M. C. Hernández, J. Chao, M. N. Miyatake and H. P. Mean, *"Watermarking-based Color Image Authentication With Detection And Recovery Capability", IEEE transaction , Vol 2, Issue No. 2,pp Feb. 2016.*