Abstract—Digital watermarking technique often used in the field of digital data copyright protection. With the rapid growth of internet and technology, millions of digital data transmitted through various channels which may cause an unpredictable ability to access and distribution of original digital data. This paper is an approach to watermark all type of attribute in relational database. Most of the research done in the area of digital watermarking based on relational database focused only on numeric attributes and work well when there is one owner of relational data. Firstly it starts with overview, properties, challenges and analysis of some major watermarking techniques. Secondly, it contains the algorithms used for inserting and detecting watermarks. The proposed technique has provided reliability and protection of relation database between owner and authenticated users.

Index Terms—Digital watermarking techniques, Robustness, Detect ability, Usability, Blindness, security, Relational database, frequent updating, Watermark insertion, Watermark detection.

I. INTRODUCTION

Digital watermarking is the act of hiding information into a digital multimedia object such as image, text, video and audio which is difficult to remove.

It is a technology that provides a robust solution for copyright protection to the digital media, provide security, tamper detection, and ensure data authentication and traitor tracing purposes [1]. With the rapid growth and expansion of internet technology, unauthorized duplication and sharing of digital contain is becoming effortless requiring minimal capital expenditure and providing easy illegal revenue. It is easier to replicate, alter and transmit digital data over the internet.

In today’s internet-based world, copyright protection for a relational database is an important concern because unauthorized modification of data may cause serious consequences for an organization. Firstly watermarking techniques introduced only for image processing later it is extended to secure text and digital multimedia document. The techniques developed for multimedia object cannot directly use for database due to differences between multimedia objects and database. These differences include [4]:

- There are large numbers of bits in a multimedia object with extensive redundancy. Therefore, the watermark has more space to hide data, where as a database has relation consists of tuples and attributes which represents a separate object. So a separate object is required for an inserting watermark in database.
- The spatial/temporal position of a multimedia object remains unchanged during embedding information into a multimedia object. However, tuples may changes with updates in database.

If there are some missing portions in of multimedia data objects, it is easier to detect it. Whereas there are different tuples and attributes in a relational database, if tuples are replaced or deleted generating changes of database it is hard to identify it.

Digital watermarking provides copyright protection, tampers detection, and maintaining integrity to relational data. In general, the goal of digital watermarking is to insert a robust watermark into the digital content such that the mark does not destroy the value of the content, and the mark is hard to be removed by adversaries without destroying the utility of content.

In digital watermark embedding process, a private key K is used to embed the watermark W into the original database. After embedding a watermark in a database it is distributed across the network. For performing verification of a suspicious database, the verification process is used to compare an original database with the suspicious database. The private key K which is used at the time of embedding process is used to verify suspicious database. This process compares suspicious database with the original watermark information [5].
The watermarking technique for relational databases has the following properties [5, 13, 14]:

**Robustness:** A digital watermark must be able to resist malicious attacks. After the attack, it will not be destroyed easily, and the embedded digital watermark still be extracted.

**Detectability:** A digital watermark must be able to detect by simply examining the suspicious database.

**Usability:** It should not distort of data in which it embedded. A data must be important after embedding a watermark.

**Blindness:** The digital watermark must be retrieved without the original relational database or digital watermark information.

**Security:** Selection of the position used to embed the digital watermark is determined by some secret parameters, for example, a secret key. These secret parameters must keep secret, and they only can be known by certain people, e.g. database owner.

III. CHALLENGES
The challenging factor that affects watermarking relational database is as follows [12]-

**Few redundant data:** A relational database contains many tuples, each representing an independent object. Therefore, there is no place to hide watermarks in the database.

**Out-of-order relational data:** There is no fixed location for any tuples in a relational database, which makes building an equivalent relation very difficult in relational databases.

**Frequent updating:** In Relational database operation like insertion, dropping, updating is very frequent. Sometimes without any intention, users often casually drop some tuples or attributes. Sometimes the unauthenticated user also can add or drop the tuples and attributes.

IV. LITERATURE REVIEW
Watermarking relational databases is a comparatively new area for the researcher, which provides copyright protection on relational databases. Therefore, literature in this sector is very limited and focused only on embedding watermark of binary bits in numerical databases.

In [1] R.Agrawal et al. proposed algorithm the watermark bits is embedded into the least significant bits (LSB) of selected attributes of a selected tuples in a relational database. In this technique, only a secret key is used and it does not provide any technique for multibit watermarks. A secure message authenticated code (MAC) is computed for each tuple with the use of secret key and the tuple’s primary key, which is used to select candidate tuples, attributes and the LSB position in the mark attributes. It is efficient to hide bits in LSB. However, this algorithm compromises by very trivial attacks.

In [2][3] Sion proposed algorithm watermark bits allocated within the properties of tuples rather than the data itself. In the proposed algorithm each set of tuples is partitioned into a maximal number of unique, nonintersecting subsets of tuples. Each selected subset of tuples, construct by a minor change in some of the data values of the tuples and then usability bounds is checked it. If usability bounds exceeded, retry different encoding parameter variations or, if still no success, try to mark the subset as invalid, or if still no success ignore the current set.

In [6] Ali Al-Haj and Ashraf Odeh proposed an efficient relational database watermarking algorithm bases on hiding binary image watermark in the non-numeric multi-word attribute of selected databases tuples. This watermark embedding process based on creating a double space at a localization determined by the decimal equivalent string. The extraction process is based on a counting of single spaces between two separated double space locations. A major advantage of this approach that there is a large bit space is available to hide image watermark.

In [7], Sukriti Bhattacharya, Agostino Cortesi proposed watermarking scheme is group based. Grouping is done on a basis of given categorical value. This grouping can be seen as a virtual grouping, which does not change the physical position of the tuples. After grouping, all tuples in each are sorted according to their primary key. Like grouping, the sorting operation does not change the physical position of tuples either. Each group considered independently. The length of watermark Wk for a k-th group is in (the number of tuples in a k-th group). It builds hash functions on top of this grouping and gets a watermark as a permutation of tuples in original table. This technique is distortion free because the ordering of tuples does not affect the original database.

In [8], G.H.Gamal, M.Z.Rashad and M.A.Mohamed Proposed technique are based on changing DB schema; which is the model of database contents, thus the structure of
data will be changed by adding a new calculated column (altering the table) based on the original columns of the relational database. The formula used in constructing the new column as well as the secret key known only by the data owner.

In [9], Ersin Uzun and Bryan Stephenson proposed two new techniques that can be used to watermark relational database. These methods can accommodate databases with no numerical attributes. In the proposed techniques they show that watermarking technique can be extended to fingerprinting algorithms that can accommodate collusion to secure binary key.

In [10], Hazem M.EI-bakry, Nikos Mastorakis proposed technique relies on changing database schema; which is the model of database contents, thus the structure of the data will be changed by adding a new record (altering the table) relies on the original data in each field of the relational database. The function used in constructing the new record as well as the secret key known only by the data owner. In general, the function used in protecting this relational database is locked via a predefined secret key. In general, this technique can be used for ownership protection of the relational database that contains only numeric values. This technique adds only one hidden record with a secret function which also locking this calculated row from an attack for changes such as deleting or updating.

In [11], S. Suhail, M. Kamran proposed a mechanism based on bi-level security that provides solutions to resolve conflicting ownership claims over watermarked dataset. If the proposed bi-level security mechanism is not appropriate, then they provide several other mechanisms for resolving conflicting ownership claims. But proposed technique is restricted to numeric data.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>Original Database</td>
</tr>
<tr>
<td>DW</td>
<td>Watermarked Database</td>
</tr>
<tr>
<td>D1</td>
<td>Sub set of tuples which will be marked</td>
</tr>
<tr>
<td>D2</td>
<td>Subset of tuples which will not be marked</td>
</tr>
<tr>
<td>K1</td>
<td>First secret key</td>
</tr>
<tr>
<td>K2</td>
<td>Second secret key</td>
</tr>
<tr>
<td>r.a1</td>
<td>First attribute of tuple r</td>
</tr>
<tr>
<td>r.a2</td>
<td>Second attribute of tuple r</td>
</tr>
<tr>
<td>x.pk</td>
<td>Primary key of tuple x</td>
</tr>
<tr>
<td>∞</td>
<td>Infinite value</td>
</tr>
<tr>
<td>p</td>
<td>Marking intensity parameter</td>
</tr>
<tr>
<td>d(x,y)</td>
<td>Distance function</td>
</tr>
<tr>
<td>L</td>
<td>Temporary variable used for finding minimal distance of two attribute values</td>
</tr>
</tbody>
</table>

III. V. PROPOSED APPROACH
The proposed approach is based on the modification of R.Agrawal et al. [1] gave the idea of watermarking numerical

VI. ALGORITHMS
The main purpose of writing this algorithm is to provide robustness to database watermarking technique. In previous techniques watermark would insert at one place only so it is easier to find it. It can be removed by some of the database update operation. In the proposed approach two secret keys K1, K2 for watermarking technique.

- K1 will be used for mark any type of an attribute in Relational Database.
- K2 will be used for giving the permission to user for updating on Relational Database.

If users have one secret key K1 that means users can only see original database from watermarked database, they can’t modify a relational database. If users have both keys K1 & K2 that means they can see and update a database.

Table 1: The Notation used in algorithms and their description

VII. WATERMARKED INSERTION
Step1:
for each tuple r in D do
  if (hash (k1||r.a1) mod p =0) then
    insert r into D1
  else
    insert r into D2
end if
end for loop

Step2:
for each tuple r in D1 do
  L= ∞
  for each tuple x in D2 do
    if (hash (k1||x.pk) mod p = 0) then
      if (d (r.a1, x.a1) < L) then
        L = d (r.a1, x.a1)
        y = x
      end if
    end if
  end for loop
  Swap (r.a2, x.a2)
end for loop

Step3:
Combined D1 and D2 then produced Dw.
A Novel Digital Watermarking Approach for Relational Database

Step 4:
Add new calculated column. Column value based on the original Columns values of DW and Secret key k2.

Step 5:
Copy the values of a new calculated column to another new column without formula.

Step 6:
Hide the calculated column and export the table with new added copied column.

Step 7:
Now DW can be published or distributed.

VIII. WATERMARKED DETECTION

Step 1:
for each tuple r in Dw do
if (hash (kl||r.a1) mod p =0)
    insert r into D1
else
    insert r into D2
end if
end for loop

Step 2:
for each tuple r in D1 do
    L = ∞
for each tuple x in D2 do
    if (hash(k1||x.pk) mod p = 0)
        if (d(r.a1, x.a1) < L) then
            L = d(r.a1, x.a1)
y = x
        end if
    end if
end for loop
Swap (r.a2, x.a2)
copy r.calculated column, r.copy column)
copy (x.calculated column, x.copy column)
end for loop

Step 3:
Combined D1 and D2 into D then produced original database D to the right owner.

IX. CONCLUSION

In this paper presented a novel approach to watermark an all type of an attribute in the relational database. The proposed technique has provided reliability, robustness and copyright protection of relation database with the aid of the user predefined function; which inserts an additional hidden column to the available relational database. This technique has many advantages over existing techniques. First, it is available for any type of relational data. Second, it is not possible to delete the hidden column because it has been locked with a secret key known by owner. Third, it is maintaining existence multiple users with a different mode. This technique can be used effectively where a huge amount of confidential relational data is transmitted between owner and authenticated users.

ACKNOWLEDGMENT

I am heartily thankful to my guide, Prof. Dr. Saroj Hiranwal, and Prof. Mr. Vijay Sharma who encourage me and provide me necessary guidance and I also want to thank my family and friends for helping and supporting me.

REFERENCES