

Survey on Relational Database Watermarking Employing Evolutionary Methods

Mohammed Mazhar^{1*}, Rajesh Dhakad²

¹PG Student, Department of Computer Engineering, Shri Govindram Seksaria Institute of Technology and Science, Indore, Madhya Pradesh, India

²Associate Professor, Department of Computer Engineering, Shri Govindram Seksaria Institute of Technology and Science, Indore, Madhya Pradesh, India

*Corresponding Author: mazharm325@gmail.com

Received Date: December 30, 2022

Published Date: January 16, 2023

ABSTRACT

Ownership control, integrity, and tamper-proofing of relational data are essential concerns that must be addressed as the communication (network) system grows. Over the past several years, a plethora of potential solutions has been proposed to address these issues (including cryptography, steganography, digital signatures, watermarks, and many others). Digital watermarking (comparatively new) is a technology that aids in the resolution of issues such as data theft, authenticity, and copyright claims. This paper's primary contributions are as follows: 1) Examine evolutionary algorithm-based watermarking in relational databases that provide optimize data encoding space in search space which provides high robustness and imperceptibility. 2) Examine contemporary database watermarking strategies based on evolutionary algorithms for high embedding capacity and watermark bit insertion. 3) Malicious agent tracing through cluster-based (mainly fuzzy c-mean clustering algorithm or based on a hash function) or multiple watermarking approaches over shared or collaborative networks. This article focuses on watermarking numeric relational databases for authentication and integrity. Also provide a brief overview of the development of relational database watermarking and emergence, its characteristics and application, and the popular research methods currently in use. In addition, gist on various possible attacks. At last, this paper suggests directions for further research in these areas. Researchers can use the findings of this study to build secure watermarking methods for databases.

Keywords- Clustering, Evolutionary algorithm, Information, Ownership protection, Relational database watermarking

INTRODUCTION

The digital era has seen an increasing tendency toward the depletion of relational databases for data exchange across shared, distributed, or collaborative contexts (network, cloud, or centralized). Additionally, there have been advancements in data storage and analysis methods like data mining and data warehousing, which study different market trends and preferences for decision-making by researchers. These developments make data a valuable resource [1]. Relational databases are more vulnerable to security threats. Before using the data (source or content) for any application purpose, its dependability must be verified. As a result, watermarking is used on the data to verify its trustworthiness and ownership. Intruders are more likely to target certain databases. It is necessary to confirm the reliability of the data (source or content) before utilizing it for any intended purpose. As a consequence, watermarking is applied to the data to confirm its reliability and ownership by including a digital watermark. To establish the authenticity, integrity, and ownership of the data. Watermarking a relational database refers to the application of these techniques. The following case situation has a solution thanks to watermarking: Case 1, Ownership Claim: Relational data can be stolen or illegally copied and transferred via the internet. to assert data ownership. Case 2, Tamper proofing: Distributor difficulty, illicit data modification, dissemination, or piracy [2]. to overcome by

identifying a leak channel or fragility ratio. Case 3: Integrity: Attacks should not undermine the utility of data.

To help new researchers in the field, this paper aims to offer a thorough review of current relational database watermarking approaches, together with their advantages and disadvantages. The paper's classification approach is based on evolutionary algorithms, distortion-based methods, and tracing-based methods, taking into account where and how the watermark is introduced. We prioritized papers from renowned computer science conferences and publications to keep the comparison succinct and allow readers to see how watermarking techniques have changed over time.

BACKGROUND

The idea of watermarking is not a novel concept. Watermarks have long been used on paper to identify a publisher and discourage money counterfeiting. A pattern, logo, text, or other images may serve as a watermark. Authenticity and integrity are more and more important in the modern day since so much data and information are stored and transferred digitally. To implant specific information into a photograph, text, music, video, piece of software, or database, one can use digital watermarking [3]. Watermarks are classified in Fig. 1 according to their characteristics.

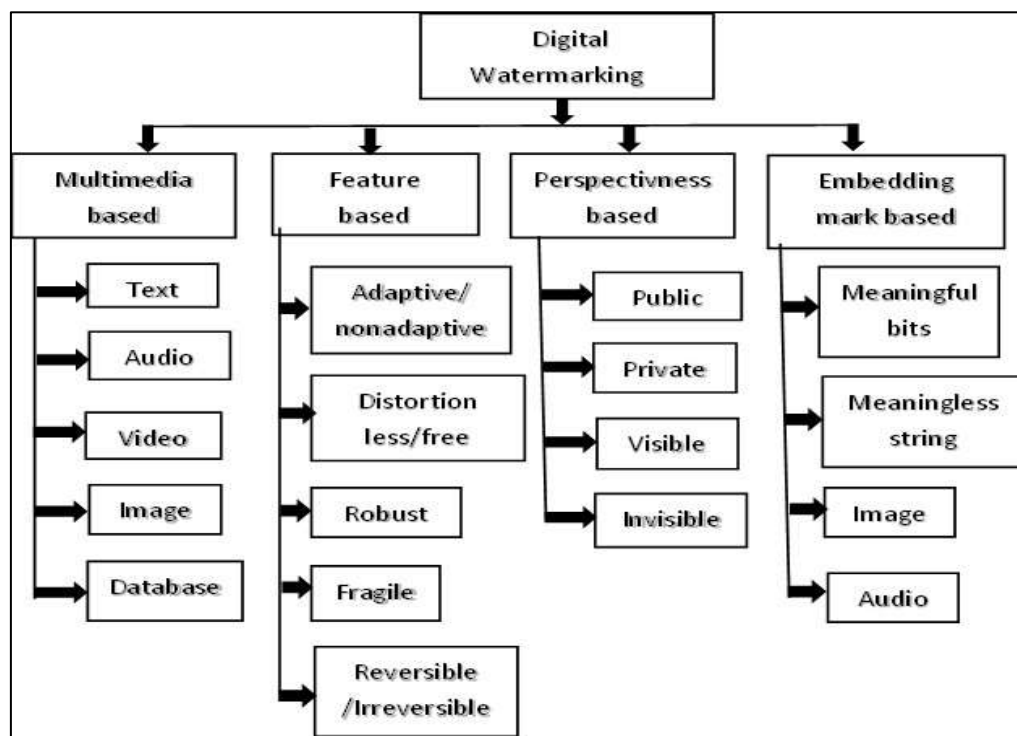


Figure 1: Types of watermarks based on properties.

Multi-Media/Nature of Cover Data-Based

Text Watermarking- Add a watermark to data that is alphabetic, numeric, or both. In this, grammar and ontology play a significant role. Audio is an audio watermarking format that may be used with mp3, mp4, wav, or WMA files. As the human hearing system's sensitivity increases, this becomes more difficult.

Video- Real-time extraction is necessary to embed the data in the video stream [4]. The process of WM's imperceptibility is laborious because of its three-dimensional properties.

Image- Integrating the primary information into the digital image. The image is huge, extremely sturdy, and undetectable to embed data, just like other sorts of data.

Database- Allows you to change the value of a few chosen data points or insert a watermark into the statistics.

Feature-Based

Adaptive vs. Non-Adaptive- A watermark is produced and will alter (adapt) as the data stream changes. Non-adaptive nature retains a

watermark that was previously created (one for all time).

Distortion Base vs. Distortion Free- Embedding the watermark in original data in a way to minimize value change is based on the idea of usability tolerance. More than a character or tuple level It offers ownership verification and copyright protection [5].

Robust- Watermark has to be able to withstand various attacks and noises.

Fragile- If an intrusive user manipulates WM data to erase a mark from a cover type referred to as WM Fragility, it is ideal with reliability and data integrity [6]. It is also categorized as semi-fragile and is mostly used for authentication [7].

Reversible vs. Irreversible- Often referred to as lossless or invertible watermarking. This phrase means that the watermarked records or data's original content is retrieved during extraction without affecting the data [8]. Data quality is compromised with irreversible watermarking because the original data is permanently altered. Only irreversible watermarking protects ownership rights and prevents piracy [9].

Human Perception Based

Public vs. Private- Without a secret key and marked information, public/blind watermarking extracts marked bits from the dataset. For private/non-blind watermarking systems, the original data is at the very least required.

Visible vs. Invisible- In visible schemes, the watermark (a word or logo) is included in the cover data. The eradication process is a pricy and time-consuming operation [10]. A general example is the logo used by electronic channels. The naked eye cannot see an invisible watermark. A unique extraction technique is the only way to detect it. WM applications are characterized as robust, fragile, or semi-fragile based on their nature [11].

Embedding Marked Based

Meaningful Bits- Incorporated as a mark is specific owner information (name, ID).

Meaningless String- Feature allows you to include a random bit's string that is created from the cover data itself and used as a watermark.

Image- Insert converted image bit string into the data set.

Audio- Embed speech as a watermark into cover data.

DIGITAL WATERMARK: FRAMEWORK AND CHARACTERISTICS

Fundamentally, a watermark is a distinct piece of data that is applied to digital resources as copyright data at a predetermined moment [12]. In addition to establishing ownership, preventing unauthorized copying, and preserving data integrity [13]. The phrase digital watermark was initially used by Komatsu and Tominaga in the late 1990s. After that, it receives increasing attention across a variety of fields (Text, Audio, Video, Database, and many more).

A watermark is a set of bits inserted into digital media that enables the identification of the creator or authorized users. Digital watermarks, in contrast to conventional printed watermarks, are intended to be imperceptible to customers. The data is scattered with encoded bits to prevent discovery or modification (image, text, audio, video, data set). A digital watermark must thus be resistant to detection and undetectable [3]. We may break down the watermarking process into four steps: creation, embedding, extraction, and recovery, for a better understanding. The fundamental paradigm of the watermarking process is shown in Fig. 2.

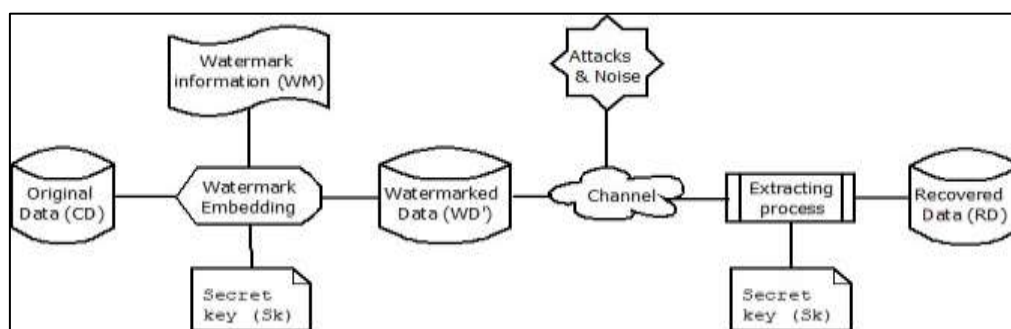


Figure 2: The basic model of watermarking approach.

Generation

The owner creates a secret code of information, which may or may not have any significance, as evidence of authenticity and ownership. Hash functions are frequently used to create secure watermark bits. The following are the main traits that might be followed:

Proper Watermark Selection- The watermark should be carefully selected to establish ownership. It shouldn't be overtly obvious.

Length of Watermark- With fewer mark bits, imperceptibility increased but robustness decreased. The bigger mark bit, on the other hand, causes greater data distortion.

Creating a Secret Key- The secret key's length (16 bits is suggested) must be carefully

determined to thwart a brute-force assault.

The Choice of Possible Locations- The property chosen should not provide the attacker with a clue as to where (attribute, bits) a possible watermark may be located.

Embedding

By utilizing a secret key Sk that is only known to the owner, the owner/mark information W is embedded into the original digital data or CD cover data (picture, text, audio, video, database). WD' generated watermarked data as the output. The typical procedure of embedding a watermark is shown in Fig. 3. How is the embedding procedure carried out? Below is a basic description of it:

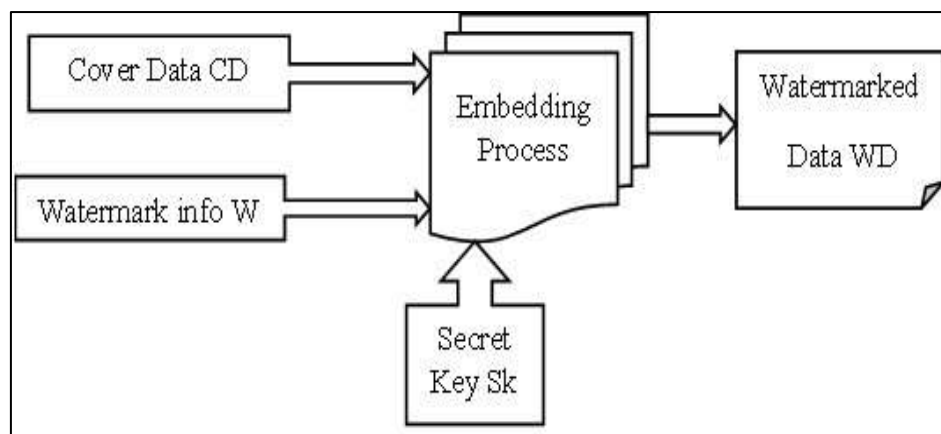


Figure 3: General watermark embedding process.

Pre-Processing- One of the pre-processing steps is converting watermark data (picture, owner identification, etc.) to bit strings. The cover dataset is furthermore typically divided into logical clusters by a predetermined threshold.

Encoding Policies- A watermark can be encoded using a variety of techniques, including attribute selection, tuple construction, watermark optimization, and embedding.

Usability Restrictions- To guarantee that the dataset's integrity and usability are maintained after the watermark. A watermark bandwidth example.

Watermark Embedding- By integrating the watermark into the dataset utilizing embedding regulations and usability limitations. Concerning factors include (a) distortion level (b) scheme robustness or fragility (or degree of fragility), (c) scheme loss lessens or irreversibility, and (d) Ought to be undetectable.

Decoding Parameters- Decryption also uses

embedding parameters.

Watermarked Dataset- The watermarked dataset is then sent to the intended recipients.

General Overview of Common Embedding Methods

For this categorization, we looked at how easily the watermark bits might be altered to incorporate grades.

Bit Resetting Techniques (BRT)- These methods carefully reset certain bits of a given data set, frequently LSBs. BRT can be divided into three smaller categories: fingerprinting-based, bit pattern-based, and image-based (which uses an image as a watermark).

Data Statistics-Modifying Methods- These methods use data statistics, such as mean, variance, distribution, etc. to include (embed) watermarks. Additionally, this module is split into bit-based and picture pattern-based groups.

Constrained Data Content-Modifying Methods- These methods, which retain the usage of watermarked data, are based on modifying the contents of the data, such as reordering the order of the tuples or adding extra spaces to attribute values. Put these methods into two categories: attribute-based and tuple-based.

Extracting

The watermark from the dubious dataset should be extracted. The brief mentions the following elements as being involved.

Pre-Processing- The partition is rebuilt using the same procedure and parameters. if the partitioning was carried out when the watermark was embedded.

Decoding Strategies- When removing the watermark, use check policy factors like the kind of blindness (public or private), and the method of removal (probabilistic or deterministic).

Decoding Process- Is done after the column(s), row(s), and bits that are embedded have been found.

Post Processing- The main goal is to, if necessary, use error correction techniques on the decoded watermark bits.

Extraction Process

The watermark is extracted using every decoding technique option that may be used. The following are typical extraction procedures (pictorial view shown in Fig. 4):

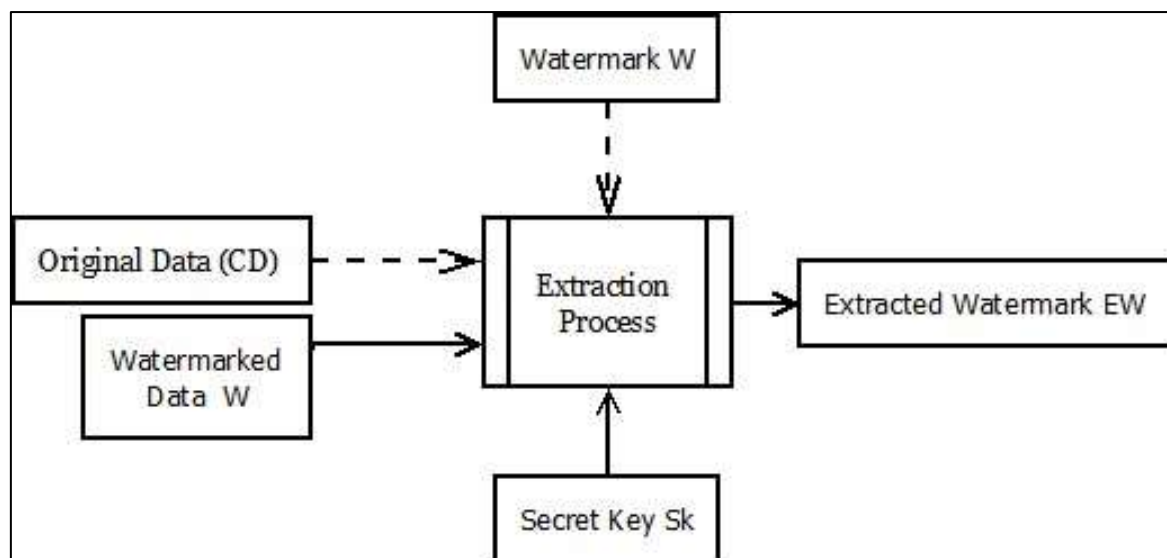


Figure 4: Public watermarking extraction process.

Blind/public watermarking eliminates the need for the original data to identify the watermark [13]. With the use of the secret key and watermarked data as input, it recognizes watermarked bits for validation. In the semi-blind watermarking procedure (Fig. 4), all that is required to extract the watermarked bit is the watermarked data, the watermarked information, and the secret key. Additionally, this procedure does not demand original data [14].

Using original data (OD), watermarked data (WD'), a secret key, and an embedded watermark W, an application of nonblind/private watermarking (Fig. 4) to detect the watermark for ownership verification purposes extracts the watermarked bits W' and compares it to the embedded watermark W.

Recovery

Removing the original data or mark from data with a suspicious watermark. If an (only in) reversible watermarking system is employed.

APPLICATIONS OF DIGITAL WATERMARKING

The characteristics of watermarking are crucial and significant when developing a system for watermarking diverse applications [15, 16]. Numerous aspects of watermarking are explored in [13, 17]. Table 1 lists the essential characteristics and related applications of digital watermarks [18, 19].

Table 1: Basic characteristics of watermarks with application domain and purpose.

Essential Features	Application Domain	Purpose Achieved
Robustness	Telemedicine	Copyright Protection
Imperceptibility	Digital Image	Authentication
Security	Digital Documents	Authentication
Effectiveness	Telemedicine	Identification
False Positive Rate	Telemedicine	Copy Control
Payload Size	Sensors	-
Capacity	Broadcast Monitoring	-
Complexity	Tele-Medicine	Copyright Protection
Verifiability	Tele-Medicine	Owner Identification

Applications for digital watermarking are numerous. Identification of Ownership is one of the core uses for embedding a distinct digital identity into digital stuff without changing how the user perceives it. [20]. Transaction or Traitor Tracing is another instance of a fragile and partition-based technique being utilized to mitigate the anger caused, a traitor who makes improper use of and a pirate who obtains watermark content [20, 21]. Tardos' anti-collusion code solution is employed to identify the owner of content that has been shared by many people [22]. Copy Right protection/Control helps in identifying the copyright owners by embedding the watermark info into the original content [23]. The Content Authentication System aims to ensure that a multimedia file has not been tampered with. Fragile watermarks or semi-fragile watermarks, a low-level robustness approach, can help with this [24]. Ex. National ID cards, which are widely used for business transactions, banking, and traveling, use this [25]. In Broadcast Monitoring, embedding a watermark into broadcasting content for monitoring and integrity purpose plays a vital role [26, 27]. It detects illegal re-broadcasting of IPR content by piracy networks, for example, Verance Confire Media [28]. Integrity Control is an important application that comes with integrity notations Legal (knowledge of data alteration) and Guarantee (delivery of original data). Fragile and

semi-fragile approaches are used depending on the applicative context [29-32]. The insertion of Meta-data extends the description and functionalities of watermark content as a metadata transmission with the signal, e.g., ARTUS project [33, 34]. In Forensics and Piracy Deterrence, retrieval of a forensic watermark is done as proof or evidence by embedding contextual metadata (recipient IP address, received format, transmission time) [35]. Device Control is used to regulate access to a resource using a verifying device ex. Roy Dolby [36].

ATTACKS AND CONSIDERED PARAMETERS

The act of erasing or neutralizing the watermark is the main component of the watermarking attack [37]. Data is gathered in relational databases that are shared and open environments in the case of healthcare facilities. Medical data's special characteristics pose significant security risks similar to the HIPAA Act and the Hippocratic Oath [38].

The removal or deactivation of the watermark is key to the watermarking attack. The purpose of the attack is to remove (or degrade) the functioning of the content owner's mark to exploit the watermarked data (watermark). Two categories of attack kinds are possible [21]. Fig. 5 depicted the categorical view of both direct and indirect attacks.

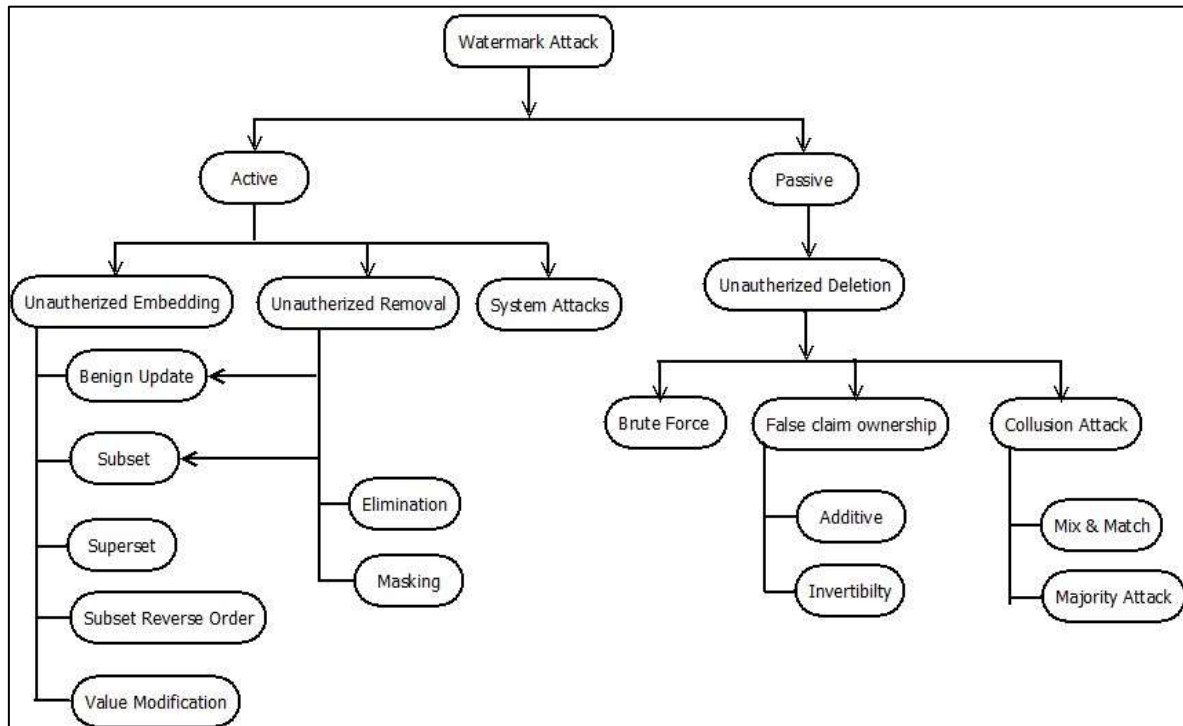


Figure 5: Various attacks on relational database watermarking.

In Active Attack, Unauthorized removal and embedding fall under this category. An attacker attempts to alter the material, conceal their efforts, or discover weaknesses in the watermarking process itself.

Unauthorized Embedding- The attacker does not create a new message to embed, but rather obtains a pre-existing legitimate message and secretly incorporates it into the task [21]. **Benign Update-** Relational database operations (such as inserting, deleting, or updating data) might interfere with the watermark extraction process [39]. **Subset-** To eliminate the watermark by erasing or changing a predetermined group of tuples or records [5]. **Superset-** Incorporating new fields, records, or tuples into the watermarked database. **Subset Reverse Order/Tuple Reshuffle Attack-** When the order of the tuples or attributes is changed; the watermark might be altered or destroyed [25]. **Value Modification-** This is referred to as a malicious attack. The following are the three main types of this attack: (a) *Bit Attack*- Attempts to decipher the watermark by manipulating the bits that make up the watermark information data [25]. Two common methods of accomplishing this are known as "randomization attacks" and "putting bits to zero at predetermined areas." (Also known as a *zero-out attack*) or by inverting particular bit

locations (also known as *bit flipping attack*) [40]. (b) *Rounding Attack*- via averaging scalar characteristics. (c) *Transformation Attack*- translates numerical data linearly, i.e., changes the unit of measurement (e.g., meter to an inch, second to minutes).

Unauthorized Removal- In this instance, a hacker who is not authorised to remove watermarks is successful. Either a masking assault or an elimination attack can do this [21].

Elimination- The attacker attempts to approximate and eliminate the watermark pattern that has been encoded. It is possible to evaluate an attack's seriousness based on the fidelity constraints.

Masking- When the watermark is there in the data but the authorize detector cannot such as image rotation.

System Attacks- The mechanism itself may frequently be thwarted by an attacker (hardware). Ex. removing the recording device's copy control chip [21].

Unauthorized detection comes into the *passive attack*, instead of obtaining information from the cover data without the owner's authorization; the attacker's goal in this instance is to prevent data modification.

Unauthorized Detection- when a malicious party successfully recognizes and decodes watermarks when they are not supposed to.

Brute Force Attack- Until the right answer is determined, the attacker tries to guess at every possible parameter (watermark information, a secret key, or even an algorithm) [39].

False Claim Ownership- This attack, when supported by proof or a claim, raises questions about the attacker's ownership of the data. Two sub-kinds exist.

Additive Attack- The intruder makes a fresh claim of ownership by watermarking already watermarked files.

Invertibility Attack- If attackers are successful in locating a watermark's friction, they can use it to claim bogus ownership.

Collusion Attack- Assuming the attacker has many copies of the database under their control. The two main collusion attacks are: [5].

Mix Match Attack- To create a fictitious relationship by applying several relations to records and tuples that are not connected by any common link.

Majority Attack- This method calculates every possible bit value of the fictional relation across all databases in a way that makes it impossible for anybody, not even the owner, to determine which database has which false relation. In the attacks mentioned above, database size and security measures are also crucial. Due to significant advancements in networking and computing tools, only a small number of attack types have been addressed yet. Researchers are thus concentrating on stronger methods to create watermarks to defend against new tools and attacks [41].

CHALLENGES OF RELATIONAL DATABASE WATERMARKING OVER MULTIMEDIA WATERMARKING

Watermarking has been used to protect many different types of data, including images, videos, documents written in natural language, programs, and semi-structured data formats including XML, map data, relational databases, electronic medical records, and geographic features. Among these, RDB watermarking is relatively new but extremely important since it is necessary to protect outsourced databases from illegal or unauthorized usage. Compared to other data formats, relational databases have a different watermarking method. The ordering of database tuples is not important, unlike with other data formats, and just a fraction of the database must be utilized when using database

watermarking [12].

Agrawal and Kiernan first raised the problem of database watermarking, which encrypts the numeric features of relational data, in a landmark study [42]. The main distinction between databases and other multimedia artifacts is highlighted by the author as follows:

- Multimedia items include a big number of bits (space), necessitating a substantial watermarking concealment bandwidth if a database lacks the room to alter or conceal watermark bits.
- Arbitrary modifications to a piece of a multimedia item have no impact on the object itself; rather, tiny modifications to a database or relational database result in massive modifications to the data.
- Many organizations and apps that deal with data do not accept permanent deformation (Military, Law enforcement, Medical, and Government data).
- In contrast to multimedia data, which is continuous, relational data is autonomous and distinct.
- Typically, the relative spatial and temporal positions of the different components of a multimedia entity remain constant. Instead, tuples are edited in an RDB.
- Multimedia entities often maintain their original form; parts of an item cannot be randomly discarded or changed without changing the object. In contrast, manipulation operations (insertion, deletion, and updating) are typical in a database.

SUMMARY OF DATABASE WATERMARKING TECHNIQUES

The difficult requirement to safeguard the ownership, integrity check, traitor tracing, and copyrights protection of digital material in a new approach has arisen with the phenomenal expansion of network communication. can withstand all potential attacks, and a balanced ratio of the parameters affecting robustness, imperceptibility, and capacity. Researchers proposed many digital watermarking methods for relational data that take these characteristics into account, including:

The LSB approach was proposed by the first author to employ watermarking in a relational database (RDB), and it uses a unique

key to include a watermark [42]. After that, they developed a method that partitions and distributes the attribute values in the tuple to include the watermark using cryptic keys [43]. Tuple deletion attacks can easily defeat this method. Later several works outlined a technique for inserting relevant strings into data [44]. By ensuring that the selected attribute value of the selected tuple has an embedded matching relation, this technique verifies the value of the watermarked bits. Later on, it was established a genetic algorithm to lessen data distortion and created a main key attack-resistant approach using majority vote and Hemming codes to reinforce the program [45]. The flaws in this technique cause data misinterpretation.

By utilising the differences in attribute values, the first reversible RDB watermarking method creates histograms and broadens the histogram approach to achieve database watermarking reversibility [46]. By utilizing the reversible quality of exclusive OR operations, the same author develops another reversible technique. However, it is useless against deletion assaults [31].

The differential extended watermarking strategy was used to create a novel reversible method for relational data (DEW). The scheme's resilience, however, is low [47]. Later, a predictor is utilized to incorporate the watermark bit using the prediction error extended watermarking (PEEW) approach that was created [48]. Utilizing differential extended watermarking methods and genetic algorithms, one additional reversible watermarking scheme was created (GADEW).

The first-time genetic algorithms were applied to database watermarking was in this case. To avoid data tampering and increase watermarking capacity [49]. The watermarking issue was resolved by combining evolutionary algorithms with a data analysis technique from information theory. They used evolutionary algorithms to produce the finest watermarks and minimize data distortion. There are considerable computing expenses when working with a lot of

data [50].

The researchers recommended employing circular histograms to alter data in plain text domains as a trustworthy technique to watermark data [51]. Using the Firefly algorithm and integrated differential extension (DEW) techniques (FFA), created the FFADEW reversible watermarking technique for relational data [52]. The authors selected the ideal attribute and value pairings using the Firefly algorithm to obtain the least degree of distortion. The concept of a genetic algorithm-based histogram shifts watermarking system (GAHSW) divided the tuples using a genetic process, and then they concealed the watermark using a histogram shift approach. This method minimizes data distortion while boosting the watermarking's robustness [53]. It lessens data distortion without reducing the endurance of watermarking [54]. Histogram gaps (HGW), a further reversible database watermarking technique, are presented. Compared to GAHSW, it lessens data distortion without reducing the endurance of watermarking [55]. In a work done, to lessen the likelihood of privacy breaches without affecting the database's regular access, they developed a method that encrypted the data using order-preserving encryption (OPES) based on the circular histogram watermarking methodology [56]. In addition to the many approaches discussed above, this article focuses primarily on the use of watermarks to safeguard the validity and integrity of numerical relational databases. The main contribution is: (1) to investigate relational databases that can monitor data and employ watermarking based on evolutionary algorithms for study purposes. (2) Researches contemporary evolutionary algorithms-based high embedding density database watermarking techniques. (3) Employing a mix of watermarking methods or a cluster-based approach to find rogue agents. A comparison of current relational database watermarking systems is included in Table 2, having a literature study on papers, along with the cover paper criteria given in the abstract [57-70].

Table 2: Survey table to compare various watermarking techniques.

Method	Flaw in Work	Intend	Attack Analyses	Performance Measures
Prediction-error expansion [57]	Include non-numerical, object-oriented, and	Data recovery tamper detection	Tuple/value alteration, insertion, and	Mean and variance

	extensible mark-up language (XML) data in the solution		deletion attack.	
Based on clustering [61]	Use this method for network data sharing to reduce distortion and watermark redundantly.	Ownership rights and limited data trackability.	A data structure, subset insertion, deletion, and modification	The ratio of extracted sub-watermarks, mean, and variance
Modulation of attribute circular histogram's center of mass [58]	Avoids collisions on a shared database and ensures synchronization between the WM embedding and extraction phases, as well as the key length and record.	Correlation-based detector	Brute force to detect with a self-generated key identifier	Gaussian noise, correlation variance
Symmetric cryptosystem instead of the public key encryption [59]	Lacks a stipulation of WM's assumed source. the delivery mechanism for private keys. Public key encryption and anonymity concerns	Privacy protection	Coalition resistance, a man-in-the-middle attack	Unbinding problem, anonymity problem
Tardos fingerprinting code [60]	Include non-numerical data.	Tracing	Not define	Mean, standard deviation
Based on fcm clustering with single item hashed function [62]	Numerical attribute values can adjust to slight modifications.	Ownership	Bit reversal attacks, secondary watermark attacks,	All common attacks
Using bacterial foraging algorithm [63]	Robustness to attacks from a variety of angles and workable solutions for non-numerical qualities as well.	Identification and proof of ownership	zero out, bit flipping attack, subset deletion attack, synchronization error, linear transformation attack	Mean, variance
Integrating a dual watermarking technique with a cryptographic hash function. [65]	Concentrate on generating correct data for the essential attributes of each tuple when altering or removing data.	Data recover,	-	Robustness analysis
The adaptive technique (lexical inverse optimal, space, and pattern search, LSB,	The space-embedding technique, which typically protects against subset addition and modification attacks,	Copyright protection, owner identification	Attacks on subsets that include subset addition, subset modification, and subset deletion.	Robustness ability (means and variance)

and Symbol modification) [66]	has to be improved. There is a chance that all current approaches can be rendered 30% more resistant to subset deletion			
Random forest and genetic algorithm histogram shifting WM [67]	Most numerical databases utilize Schemes for security and copyright monitoring, however, this method only uses the tuple's main key, leaving the database vulnerable to attacks that seek to delete or modify data.	Copyright protection and traceability	Robustness analysis (insertion, deletion, modification), capacity and distortion analysis	Statistical analysis
Based on clustering, along with watermark embedding, detection, and data recovery algorithm [68]	High-intensity tuple alteration attacks can annihilate the watermark, significantly exhaust the data, and render the data inapplicable.	Manipulation by malicious attackers.	Tuple delete attack, tuple modification attack	WM hiding rate, standard deviation
Evolutionary techniques, maximum relevance, and minimum redundancy (mrmr) [69]	Approximately 70% of the data was retrieved from modification assaults, and 50% from deletion attempts.	Ownership documentation.	Insertion, deletion, alteration	Mean and variance
A generic watermarking model for object-relational databases [70]	strategy to tackle additional security issues including data provenance and fingerprinting.	tamper detection	Integrity analysis synchronization error	Mean and variance
Genetic algorithm and histogram shifting watermarking [53]	Proposed creation of reliable and reversible WM for non-numerical data in remote shared database systems.	Reduce distortion and boost robustness.	Insertion, deletion, alteration	Bit error rate, mean and standard deviation
Through the HOLPSOFA algorithm [70]	Implementing OLPSO and FA together can significantly enhance performance.	Optimal location	Insertion, deletion, and alteration	Mean square error

FUTURE DIRECTION FOR NEW SCHEME

In this part, we describe in detail how the researcher may make a well-informed

decision about the watermarking approach to implement. For that, we examine multiple watermarking procedures. Several observations are provided below:

One location selection strategy: rules to

determine the number of characteristics, tuples, and bit locations to be embedded, as these parameters are typically chosen based on the owner's prior experience or preferences in large-scale surveys.

Second, Algorithms impact: The effectiveness of algorithms is contingent on their determinism and the degree to which they optimize for the following three criteria: computational cost (low), data usability (high), and robustness (high). Watermark production, embedding, and extraction all add time and storage space to the computation cost. Usability covers less distortion in the data values it measures with a statistical parameter like mean and variance.

Third, Database: The majority of the database schemes other than are intended for numerical schemes as per the covered paper [70]. Need to develop a generic scheme for any dataset, whether it's category, textual, XML, web-based, or query-based.

Fourth, Usability: After embedding the watermark, data usable for research or data mining easily. The ability to be put to practical use, requires a low distortion rate, a large embedding capacity, near-invisibility, and resistance to attack.

Fifth, Attacks: Relational database watermarking is a difficult challenge in many data distribution platforms running in today's web-based application settings. Through rigorous study, we suggested solution is secure against a wide range of database attacks, such as those that insert, delete, and modify data, brute force, or collision attacks.

CONCLUSION

Organizations and institutes (public and commercial) are gaining an advantage over large data resource providers because of significant improvements in cloud computing, machine learning, IoT, and big data. Adopting privacy-preserving strategies is essential to provide data ownership, copyright, and traitor tracing in light of the multiple requests for relational datasets through public or private networks, either centralized or decentralized. This work discusses several watermarking attacks as well as an overview of some cutting-edge watermarking methods. This review article examines approaches for watermarking numerical databases from 2013 to 2022. The methods of

classification that have been described are based on clustering, the capacity to track leak sources and traitors, and evolutionary strategies for adding watermarks. Maintaining a balance between the three key characteristics of watermarks robustness, imperceptibility, and distortion can be difficult.

In conclusion, we want a strong watermarking system that ensures data quality and is resistant to harmful attacks. To much-needed plan, development focus on recuperation and trustworthy authentication, mobile application, distributed real-time technology, NoSQL, and web-based databases (dynamic database web-based). Due to the rapid advancement of networking and computing tools, only a few different types of attacks have been dealt with thus far. As a result, scientists are concentrating on stronger methods to create watermarks that can withstand new tools and attacks.

REFERENCES

1. M Ru Xie, et al. (2016). A survey of data distortion watermarking relational databases, *International Journal of Network Security*, 18, 1022-1033, Available at: [https://doi.org/10.6633/IJNS.201611.18\(6\).03](https://doi.org/10.6633/IJNS.201611.18(6).03).
2. M. A. Malik (2004). Efficient data hiding techniques for digital rights management of multimedia archives. University of Illinois, Chicago, US, Available at: https://www.academia.edu/8798759/Efficient_Data_Hiding_Techniques_for_Digital_Rights_Management_of_Multimedia_Archives.
3. M. Kamran and M. Farooq (2018). A comprehensive survey of watermarking relational databases research, *arXiv*, Available at: <https://arxiv.org/abs/1801.08271>.
4. K. E. Drandaly, et al. (2020). A digital watermarking for relational database: state of art techniques, *International Journal of Advanced Science and Technology*, 29(3), 870-883, Available at: https://www.researchgate.net/publication/283773348_Applying_Website_Usability_Testing_Techniques_to_Promote_E-services.
5. F. Y. Shih (2017), Digital Watermarking and Steganography: Fundamentals and Techniques, 2nd Edition. CRC Press, Florida, US. ISBN-10: 1498738761,

- Available at: <https://www.amazon.com/Digital-Watermarking-Steganography-Fundamentals-Techniques/dp/1498738761>.
6. V. Khanduja (2017). Database watermarking, a technological protective measure: Perspective, security analysis and future directions, *Journal of Information Security and Applications*, 37, 38-49, Available at: <https://doi.org/10.1016/j.jisa.2017.10.001>.
 7. R. Goyal and N. Kumar (2014). LSB based digital watermarking technique, *International Journal of Application or Innovation in Engineering & Management*, 3(9), 15-18, Available at: <https://www.ijaiem.org/Volume3Issue9/IJA IEM-2014-09-06-5.pdf>.
 8. D. Rosiyadi, et al. (2012). A comparison between the hybrid using genetic algorithm and the pure hybrid watermarking scheme, *International Journal of Computer Theory and Engineering*, 4(3), 329-331, Available at: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=896deb40e527506884d5e5222129cd8ba9b517f5>.
 9. S. Mohammadi (2015). A semi-blind watermarking algorithm for color images using chaotic maps. *2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI)*. IEEE, Available at: <https://doi.org/10.1109/KBEI.2015.7436030>.
 10. Abd. S. Alfagi, et al. (2015). A Systematic Literature Review on necessity, challenges, applications, and attacks of watermarking relational database. *ICOCOE*. Springer, Available at: https://www.researchgate.net/publication/306376641_A_Systematic_Literature_Review_on_necessity_challenges_applications_and_attacks_of_watermarking_relational_database.
 11. Abd. S. Alfagi, et al. (2015). Survey on relational database watermarking techniques, *ARNP Journal of Engineering and Applied Sciences*, 11(1), 421-432, Available at: https://www.researchgate.net/publication/291830901_Survey_on_relational_database_watermarking_techniques.
 12. B. M. Sahoo, J. Behera and R. K. Rout (2015). A robust fragile watermarking technique for digital image, *International Journal of Engineering Research & Technology*, 3(25), 1-7, Available at: <https://www.ijert.org/research/a-robust-fragile-watermarking-technique-for-digital-image-IJERTCONV3IS25024.pdf>.
 13. A. Tefas, N. Nikolaidis and I. Pitas (2009), Chapter 22 - Image Watermarking: Techniques and Applications, *2nd Edition. Academic Press, Cambridge, Massachusetts, US*. Available at: <https://doi.org/10.1016/B978-0-12-374457-9.00022-6>.
 14. R. Caldelli, F. Filippini and R. Becarelli (2010). Reversible Watermarking Techniques: An Overview and a Classification, *EURASIP Journal on Information Security*, Available at: <https://link.springer.com/article/10.1155/2010/134546#citeas::~text=DOI-https%3A%2F%2Fdoi.org%2F10.1155/2010/134546,-Share%20this%20article>.
 15. S. Iftikhar, M. Kamran and Z. Anwar (2015). A survey on reversible watermarking techniques for relational databases, *Security and Communication Networks*, 8, 2580-2603, Available at: <https://onlinelibrary.wiley.com/doi/epdf/10.1002/sec.1196>.
 16. C. H. Huang and J. L. Wu (2004). Attacking visible watermarking schemes, *IEEE Transactions on Multimedia*, 6(1), 16-30, Available at: <https://doi.org/10.1109/TMM.2003.819579>.
 17. S. Craver, et al. (2006). Resolving rightful ownerships with invisible watermarking techniques: limitations, attacks, and implications, *IEEE Journal on Selected Areas in Communications*, 16(4), 573-586, Available at: <https://doi.org/10.1109/49.668979>.
 18. J. Kiernan, R. Agrawal and P. J. Haas (2003). Watermarking relational data: framework, algorithms and analysis, *The VLDB Journal*, 12, 157-169, Available at: <https://link.springer.com/article/10.1007/s00778-003-0097-x#citeas::~text=DOI-https%3A%2F%2Fdoi.org%2F10.1007/s00778%2D003%2D0097%2Dx,-Keywords%3A>.
 19. S. Kumar, B. K. Singh and M. Yadav (2020). A recent survey on multimedia and database watermarking, *Multimedia Tools and Applications*, 79, 20149-20197, Available at: <https://link.springer.com/article/10.1007/s11042-020-08881-y#citeas::~text=DOI-https%3A%2F%2Fdoi.org%2F10.1007/s11042%2D020%2D08881%2Dy,-Keywords>.

20. A. Alqassab and M. Alanezi (2021). Relational database watermarking techniques: A survey. *Iraqi Academics Syndicate International Conference for Pure and Applied Sciences*, (pp. 1-10). IOP Publishing, Available at: <https://iopscience.iop.org/article/10.1088/1742-6596/1818/1/012185/pdf>.
21. Bloom J., Cox I. J. and Miller M. (2009), Digital Watermarking and Steganography, 2nd Edition. Elsevier Ebsco Publishing, New York, Ipswich. ISBN: 978-0-12-372585-1, Available at: <https://www.sciencedirect.com/book/9780123725851/digital-watermarking-and-steganography>.
22. G. Tardos (2008). Optimal probabilistic fingerprint codes, *Journal of the ACM*, 55(2), 1-24, Available at: <https://doi.org/10.1145/1346330.1346335>.
23. M. Alattar (2000). Smart Images using Digimarc's watermarking technology. *Proceedings, Security and Watermarking of Multimedia Contents II*. SPIE, Available at: <https://doi.org/10.1117/12.384980>.
24. S. Abdulmunem. M. Al-Juboori (2019). Imperceptibility and Robustness Improvement using Segmented DWT Watermarking Technique (Master's Thesis). Middle East University, Amman, Jordan, 1-124, Available at: https://meu.edu.jo/libraryTheses/5ca9b19c25871_1.pdf.
25. Md Ali Nematollahi, C. Vorakulpipat and H. G. Rosales (2016), Digital watermarking, 1st Edition. Springer, Berlin Heidelberg, Germany. Available at: http://pdf.lib.vntu.edu.ua/books/Springer/2021/2017_Book_DigitalWatermarking.pdf.
26. W. D. Moon, et al. (1975). U.S. Patent No. 3,919,479. Washington, DC: U.S. Patent and Trademark Office. Available at: <https://patentimages.storage.googleapis.com/9b/2e/b1/3c32e3cf50bad0/US3919479.pdf>.
27. T. Kalker, et al. (1999). Video watermarking system for broadcast monitoring. *Proceedings SPIE 3657, Security and Watermarking of Multimedia Contents*. SPIE, Available at: <https://doi.org/10.1117/12.344661>.
28. Confirmedia, "Welcome to ConfirMedia", [Online] Available at: <http://www.confirmedia.com/>.
29. Y. Li, H. Guo and S. Jajodia (2004) Tamper detection and localization for categorical data using fragile watermarks. *Proceedings of the 4th ACM workshop on Digital rights management*, (pp. 73-82). ACM, Available at: <https://doi.org/10.1145/1029146.1029159>.
30. I. Kamel and K. Kamel (2011). Toward protecting the integrity of relational databases. *2011 World Congress on Internet Security (WorldCIS-2011)*. IEEE, Available at: <https://doi.org/10.1109/WorldCIS17046.2011.5749863>.
31. Y. Zhang, B. Yang and Xia-Mu Niu (2006). Reversible watermarking for relational database authentication, *Journal of Computers*, 17, Available at: https://www.researchgate.net/publication/228753085_Reversible_watermarking_for_relational_database_authentication.
32. G. Coatrieux, et al. (2011). Lossless watermarking of categorical attributes for verifying medical data base integrity. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (pp. 8195-8198). IEEE, Available at: <https://doi.org/10.1109/iembs.2011.6092021>.
33. J. F. Contreras (2014). Watermarking services for medical database content security (Doctorate Thesis). TELECOM Bretagne, University of Rennes, France, 1-206, Available at: <https://hal.archives-ouvertes.fr/tel-01206279/document>.
34. G. Bailly, et al. (2006). ARTUS: Synthesis and audiovisual watermarking of the movements of a virtual agent interpreting subtitling using cued speech for deaf viewers, *Modelling, Measurement and Control*, 177-187, Available at: https://www.researchgate.net/publication/29638292_ARTUS_Synthesis_and_audiovisual_watermarking_of_the_movements_of_a_virtual_agent_interpreting_subtitling_using_cued_speech_for_deaf_viewers.
35. S. E. Balci (2003). Robust Watermarking of Images (Master's Thesis). The Graduate School of Natural and Applied Sciences of The Middle East Technical University, Ankara, Turkey, 1-134, Available at: <http://etd.lib.metu.edu.tr/upload/1091917/index.pdf>.

36. R. M. Dolby (1981). US Patent No. 4,281,217. Washington, DC: U.S. Patent and Trademark Office. Available at: <https://patentimages.storage.googleapis.com/93/6b/30/7c25732e699c80/US4281217.pdf>.
37. P. Singh and R. S. Chadha (2013), "A Survey of Digital Watermarking Techniques, Applications and Attacks", [Online] Available at: <https://www.semanticscholar.org/paper/A-Survey-of-Digital-Watermarking-Techniques%2C-and-Singh-Chadha/3ae52abd8deb32189735613e9bb25252ce02a838>.
38. US Statut Large (1996), "Health Insurance Portability and Accountability Act of 1996", [Online] Available at: <http://www.eolusinc.com/pdf/hipaa.pdf>.
39. A. Nikolaidis, et al. (2001). A survey on watermarking application scenarios and related attacks. *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*. IEEE, Available at: <https://doi.org/10.1109/ICIP.2001.958292>.
40. R Murugan, J. T. Abraham and M J Aravind (2018). A study of digital watermarking on relational databases for ownership proofing and tamper detection, *International Journal of Applied Engineering Research*, 13(3), 160-165, Available at: http://mail.ripublication.com/ij_aerspl2018/ijaerv13n3spl_30.pdf.
41. Y. Song (2004). Watermarking of relational databases (Master's Thesis). National University of Singapore, Queenstown, Singapore, Available at: <https://scholarbank.nus.edu.sg/handle/10635/13800>.
42. R. Manjula and N Settipalli (2010). A new relational watermarking scheme resilient to additive attacks, *International Journal of Computer Applications*, 10(5), 1-7, Available at: <https://www.ijcaonline.org/volume10/number5/pxc3871998.pdf>.
43. R. Sion, M. Atallah and S. Prabhakar (2005). Rights protection for categorical data, *IEEE Transactions on Knowledge and Data Engineering*, 17(7), 912-926, Available at: <https://doi.org/10.1109/TKDE.2005.116>.
44. Z Hui (2003), "Watermarking Relational Databases for Ownership Protection", [Online] Available at: <https://www.semanticscholar.org/paper/Watermarking-Relational-Databases-for-Ownership-Hui/b2f69d9ca7d50f30a8b1487d3f4d43af22a09525>.
45. M. Shehab, E. Bertino and A. Ghafoor (2008). Watermarking relational databases using optimization-based techniques, *IEEE Transactions on Knowledge and Data Engineering*, 20(1), 116-129, Available at: <https://doi.org/10.1109/TKDE.2007.190668>.
46. Z. Yong and N. Xiamu (2006). Reversible watermark technique for relational databases, *Acta Electronica Sinica*, 34(S1), Available at: <https://www.ejournal.org.cn/EN/Y2006/V34/IS1/2425>.
47. G. Gupta and J. Pieprzyk (2010). Reversible And Blind Database Watermarking Using Difference Expansion. *1st International ICST Conference on Forensic Applications and Techniques in Telecommunications, Information and Multimedia*. ACM, Available at: <http://dx.doi.org/10.4108/e-forensics.2008.2691>.
48. M E. Farfoura and S Jinn Horng (2010). A novel blind reversible method for watermarking relational databases. *International Symposium on Parallel and Distributed Processing with Applications*. IEEE, Available at: <https://doi.org/10.1109/ISPA.2010.63>.
49. K Jawad and A Khan (2013). Genetic algorithm and difference expansion based reversible watermarking for relational databases, *Journal of Systems and Software*, 86(11), 2742-2753, Available at: <https://doi.org/10.1016/j.jss.2013.06.023>.
50. S. Iftikhar, M. Kamran and Z. Anwar (2015). RRW-A robust and reversible watermarking technique for relational data, *IEEE Transactions on Knowledge and Data Engineering*, 27(4), 1132-1145, Available at: <https://doi.org/10.1109/TKDE.2014.2349911>.
51. J Franco Contreras and G. Coatrieux (2015). Robust watermarking of relational databases with ontology-guided distortion control, *IEEE Transactions on Information Forensics and Security*, 10(9), 1939-1952, Available at: <https://doi.org/10.1109/TIFS.2015.2439962>.
52. B. Imamoglu, M. Ulutas and G. Ulutas (2017). A new reversible database

- watermarking approach with firefly optimization algorithm, *Mathematical Problems in Engineering*, 2017, 1-14, Available at: <https://downloads.hindawi.com/journals/mpe/2017/1387375.pdf>.
53. D. Hu, D. Zhao and S. Zheng (2019). A new robust approach for reversible database watermarking with distortion control, *IEEE Transactions on Knowledge and Data Engineering*, 31(6), 1024-1037, Available at: <https://doi.org/10.1109/TKDE.2018.2851517>.
54. Y. Li, et al. (2019). A reversible database watermarking method with low distortion, *Mathematical Biosciences and Engineering*, 16(5), 4053-4068, Available at: <https://doi.org/10.3934/mbe.2019200>.
55. Y. Li, J. Wang and X. Luo (2020). A reversible database watermarking method non-redundancy shifting-based histogram gaps, *International Journal of Distributed Sensor Networks*, 16(5), 1-11, Available at: <https://journals.sagepub.com/doi/pdf/10.1177/1550147720921769>.
56. S. Xiang, et al. (2022). Robust watermarking of databases in order-preserving encrypted domain, *Frontiers of Computer Science*, 16, Available at: <https://link.springer.com/article/10.1007/s11704-020-0112-z#citeas~:text=DOI-,https%3A//doi.org/10.1007/s11704%2D020%2D0112%2Dz,-Keywords>.
57. A. Hamadou, et al. (2020). Reversible fragile watermarking scheme for relational database based on prediction-error expansion, *Mathematical Problems in Engineering*, 2020, Available at: <https://doi.org/10.1155/2020/1740205>.
58. J. F. Contreras and G. Coatrieux (2017). Databases Traceability by Means of Watermarking with Optimized Detection, In: Shi, Y., Kim, H., Perez-Gonzalez, F., Liu, F. Editor. *Digital Forensics and Watermarking*, Springer, Cham; New York, US. 343-357, Available at: https://link.springer.com/chapter/10.1007/978-3-319-53465-7_25#citeas~:text=DOI-,https%3A//doi.org/10.1007/978%2D3%2D319%2D53465%2D7_25,-Published.
59. F. G. Jeng, et al. (2016). A multi-watermarking protocol for health information management, *Multimedia Tools and Applications*, 75, 8123-8135, Available at: <https://link.springer.com/article/10.1007/s11042-015-2728-9#citeas~:text=DOI-,https%3A//doi.org/10.1007/s11042%2D015%2D2728%2D9,-Keywords>.
60. A. Mohanpurkar and M. Joshi (2015). A fingerprinting technique for numeric relational databases with distortion minimization. 2015 *International Conference on Computing Communication Control and Automation*. IEEE, Available at: <https://doi.org/10.1109/ICCUBEA.2015.134>.
61. H. Chai, et al. (2019). A robust and reversible watermarking technique for relational dataset based on clustering. 2019 *18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE)*. IEEE, Available at: <https://doi.org/10.1109/TrustCom/BigDataSE.2019.00062>.
62. G. Di, X. Wang and H. Xian (2019). Watermark scheme of database based on fuzzy C-means. *Proceedings of the 2nd International Conference on Big Data Technologies*, (pp. 67-71). ACM, Available at: <https://doi.org/10.1145/3358528.3358598>.
63. V. Khanduja, O. P. Verma and S. Chakraverty (2015). Watermarking relational databases using bacterial foraging algorithm, *Multimedia Tools and Applications*, 74, 813-839, Available at: <https://link.springer.com/article/10.1007/s11042-013-1700-9#citeas~:text=DOI-,https%3A//doi.org/10.1007/s11042%2D013%2D1700%2D9,-Keywords>.
64. M. Kamran, S. Suhail and M. Farooq (2013). A robust, distortion minimizing technique for watermarking relational databases using once-for-all usability constraints, *IEEE Transactions on Knowledge and Data Engineering*, 25(12), 2694-2707, Available at: <https://doi.org/10.1109/TKDE.2012.227>.
65. Y. Zhang, et al. (2021). A robust and adaptive watermarking technique for relational database, In: Lu, W., Zhang, Y., Wen, W., Yan, H., Li, C. Editor. *Cyber Security*. Springer; Singapore, Available at: https://link.springer.com/chapter/10.1007/978-981-16-9229-1_1#citeas~:text=DOI-

- ,[https%3A//doi.org/10.1007/978%2D981%2D16%2D9229%2D1_1](https://doi.org/10.1007/978%2D981%2D16%2D9229%2D1_1), -Published.
66. C Ge, et al. (2020). Reversible database watermarking based on random forest and genetic algorithm. *2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*. IEEE, Available at: <https://doi.org/10.1109/CyberC49757.2020.00045>.
67. X. Shen, et al. (2020). Relational database watermarking for data tracing. *2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*. IEEE, Available at: <https://doi.org/10.1109/CyberC49757.2020.00043>.
68. H. Tufail, K. Zafar and A. R. Baig (2019). Relational database security using digital watermarking and evolutionary techniques, *Computational Intelligence*, 35(4), 693-716, Available at: <https://doi.org/10.1111/coin.12209>.
69. V. Khanduja and S. Chakraverty (2019). A generic watermarking model for object relational databases, *Multimedia Tools and Applications*, 78(19), 28111-28135, Available at: <https://doi.org/10.1007/s11042-019-07932-3>.
70. K Unnikrishnan and K V Pramod (2017). Robust optimal position detection scheme for relational database watermarking through HOLPSOFA algorithm, *Journal of Information Security and Applications*, 35, 1-12, Available at: <https://doi.org/10.1016/j.jisa.2017.04.005>.

CITE THIS ARTICLE

Mohammed Mazhar and Rajesh Dhakad (2023). Survey on Relational Database Watermarking Employing Evolutionary Methods, *Journal of Information Technology and Sciences*, 9(1), 13-29.