7th National Conference on
"Role of Engineers in Nation Building" organized by VIVA Institute of Technology, Mumbai
(1st and 2nd March 2019)
International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

Database Security: Optimization Approach

Akshata S. Raut[1], Bhavika P. Thakur[2]

[1]Department of Computer Engineering, VCDET, Virar, churiakshata333@gmail.com
[2]Department of Computer Engineering, VIT, Virar, bhavikapthakur@gmail.com

Abstract— *Nowadays there is increasing need of data should be available online as lifestyle becomes so digitize for betterment of lifestyle. Many technology exists today which offers the data should be available online on a single click but the security of that data is also the most important factor to consider. This paper discusses about the importance of data availability as well as importance of security to that database which contains the data which is helpful for persons to analyze market trends or make predictions about future trends and how this security measures are helpful for securing database. This paper also describes about evolution that can be made possible to improve security of that data by using optimization algorithms.*

Keywords— Database Security, Optimization, Watermarking, Genetic Algorithm, Reversible Watermarking.

I. INTRODUCTION

As in today's digitized world everything is available with a single click to make life easier. Availability of data needs to store that data into the structure known as database. Database contains the data which is helpful for users, which helps users to make decisions, business analysis, identify market trends, identify different research trends and therefore database are available online for free of cost. Relational data is shared by the owners with research communities and at virtual data storage locations in the Cloud. The purpose is to work in a collaborative environment and make data openly available so that it is useful for knowledge extraction and decision making [1], but as database is available online, it will be used by attackers, they can easily modify the available data and update the database which can create wrong assumptions. Now if updated database is use by legitimate users then there is a possibility that they analyze incorrect data and make incorrect predictions. This incorrect predictions can create major loss to the organization, therefore to avoid such scenario the security should be provided to that database which is openly available online. This paper discuss about the security measures provided by watermarking database and how we can improve that security measures.

II. IMPORTANCE OF DATA AVAILABILITY

As evolution takes place many organizations store their data online. As data is available online it is easy for the organizations to assign people for project from different geographical locations whose different skills are useful for project. Data will be shared easily among multiple locations with peoples at the same time which is time saving. Ease of access of data for organization helps them to make correct decision about future trends [2]. Business Intelligence makes use of online data availability. As the process become paperless it will not be tedious process for people who use it; also it will become cost effective. Online storage of data also make user not to worry about space utilization.

III. IMPORTANCE OF SECURITY TO DATABASE

Security for online data is also important factor to consider while we are discussing about features of online data availability. Data which is available online contains sensitive data such as health care data, personal communication, financial data, criminal records, personal information. If any unauthorized person gets access to these sensitive data he/she can create misuse of that sensitive data which in turn create problematic situation for person whose data is stolen or may create many financial losses to peoples. The attacker can makes data unavailable for some time of period or completely unavailable which may cause loss to the organization, therefore security of data must be consider as most important objective of database availability [3].

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

IV. SECURITY TECHNIQUES FOR DATABASE

While considering importance of security to database many researchers presents different techniques from which watermarking on relational database is emerging technique now a days. Watermarking on database is provided to prove owner's identity [3]. It is used to validate that the data is coming from a trusted source and is exact values which are stored in the database. The actual content of database is not modified by an attacker. Watermarking process on database consists of two phases as watermark embedding & watermark extraction [1]. During sending database to its intended receiver owner of the database embeds some information into actual database content by using secret key and sends database to receiver. At the other end receiver extract embedded watermark using key provided by sender, validate the identity of sender and get the actual database content, but while applying watermarking on database it is observed that due to bit embedding process some of the content of database has changed or quality of data gets compromised [1]. In critical applications of database this little bit change in actual content may affect the important decision of the organization to be incorrect as it analyze incorrect data values. This situation can cause a big loss to the organization.

To overcome this situation a solution is provided which is known as Reversible watermarking on database. Reversible watermarking embeds the watermark bit in such a manner that while verifying authority of database at receiver's end receiver extract watermark bits from received database and from extracted bits he/she identify authority of database. After completion of extraction process the remaining data is actual content of database [1]. Some of the Reversible Watermarking techniques with their phases are as follows:

4.1 Difference Expansion Watermarking (DEW)

Difference expansion watermarking techniques (DEW) proposed by [4] has following phases, a) performs some arithmetic operations such as addition, subtraction, multiplication and division on numeric features of selected attributes of database tables. b) Convert the original content of database into watermarked content. c) The watermark information is embedded in the Least Significant Bit of features of databases to minimize distortions.

Proposed reversible watermarking technique introduces distortions in the original content of database as a result of the watermark embedding process [4]. Amount of Change in the data is controlled by placing certain bounds on least significant bit which is decided by owner of database. On the other side, to limit the distortions, the data outside the limited bounds is not watermarked.

Difference expansion based watermarking (DEW) technique was first technique used to achieve reversibility in the database. DEW is able to restore the original database exactly but applicable to only small database. It also encourages the owner to distribute the trial version of the database, which can only be reverted by those users who have purchased the key. DEW technique of watermarking introduce distortion in the data which make watermarked attribute more visible to attacker [3]. Reducing distortion by using small value of distortion tolerance may result in limited amount of watermarked in database. DEW approach is not suitable to increase watermark capacity without increasing distortion tolerance of attributes.

4.2 Genetic Algorithm based Difference Expansion watermarking (GADEW)

GADEW proposed by [3] is a reversible watermarking technique which recovers watermark and original database exactly as it is. It is able to increase watermarking capacity of database which leads to improve security by setting up fixed distortion tolerance. Distortion tolerance specifies the percentage of change the original data can bear so that the actual data value may not lose its meaning during watermark insertion. Distortion occurs due to watermark bits insertion is reduced to minimum by using tuple wise and attribute wise distortion measures in GADEW technique of watermarking [3]. DEW only checks for selected tuples and applies tolerance for selected attributes, it will not check for combination of different attributes of same tuple therefore if distortion tolerance is not satisfied, watermark is not applied on tuples and less watermark bits are inserted into database. To overcome this problem the solution provided is to use optimization approach which checks for combinations of input and gives optimum value solution which is best for the problem to be solved [3].

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

Genetic Algorithm uses concept of optimization for generation of watermark bits to be embedded in database. An optimization technique is used for calculating an appropriate watermark string i.e., beta value. This beta value is embedded in original dataset to get watermarked dataset.

GADEW has following phases: a) Using Genetic Algorithm, different attributes are explored to meet the optimal criteria rather than randomly selecting less effective attributes among available attributes in database for watermark insertion. Checking only the distortion tolerance of two attributes for a selected tuple may not be useful to increase watermark capacity and reduced distortion tolerance therefore, distortion tolerance of different attributes are explored. Distortion caused by difference expansion [4] can help an attacker to predict watermarked attribute. Therefore GADEW incorporated tuple and attribute-wise distortion in the fitness function of Genetic Algorithm making it tough for an attacker to predict watermarked attribute [2]. b) In order to solve optimization and search problems, it uses a collection of data structures named as chromosomes. A chromosome can be considered as an array vector. It reserves two target values for each of the selected tuple in which the watermark is expected to be inserted. c) Number of chromosomes determines population size of genetic algorithm. Consecutive populations are known as generations. New generation is formed by applying genetic operators over current population. Selection, crossover and mutation are common genetic operators. d) Fitness value for each of the chromosomes is evaluated by using a fitness function. Genetic algorithm terminates if the number of generations are completed or the required criteria of fitness is achieved. e) Finally, the chromosome with best fitness value provides the near optimal solution.

The selection operator selects chromosomes from the population to forward them to the next generation. The emphasis is on extraction in the search space more than exploration. The crossover operator joins segments of two or more chromosomes to generate a new chromosome. The mutation operator creates new chromosome by changing values of one or more cell at random. Unlike the process of crossover, the role of mutation is to explore the search space. A fitness function can have more than one objectives (fitness parameter) named as multi objective fitness function; each of these fitness parameters can be measured and combined together in a single fitness function. Multi objective optimization problem deals with concurrent optimization of several objectives, such as cost and performance. An optimal solution cannot be dominated by any other solution in the search space. These objectives may be conflicting and cannot be optimized simultaneously, thus a suitable tradeoff can be found. Therefore, it is necessary to have a decision making process in which preference information is used in selecting an appropriate tradeoff [3]. This results in generation of optimal watermark string.

Genetic Algorithm (GA) is employed to increase watermark capacity and minimize introduced distortion. This is because the watermark capacity increases with the increase in number of features and the GA runs on more features to search the optimum one for watermarking. However watermark capacity decreases with the increase in watermarked tuples. GADEW used the attribute wise distortion (AWD) and tuple wise distortion (TWD) to control distortions in the resultant data. In this context, the robustness of GADEW can be compromised when AWD and TWD are given high values.

4.3 Robust and Reversible watermarking (RRW)

Paper [1] proposes a semi-blind reversible and robust watermarking technique which is applicable to numerical database only. In all previous watermarking techniques [4], [3] attributes or features are selected for watermarking without taking in to consideration their role in analysis of data or their importance in analyzed data. Quality of data is preserved in RRW [1] by taking into consideration relevance of the feature (attribute) in knowledge discovery while selecting a feature (attributes) for watermarking that is RRW takes into account mutual information (MI) measure for finding out relative importance of features [5]. After calculating MI value attributes are selected for creation of optimal watermark bits which are going to be embedded in data using principle of genetic algorithm. Attributes whose MI value is less than threshold value (distortion tolerance decided by owner of database) are selected for embedding watermark. Using RRW all or large portion of database can be watermarked. As large portion of database is watermarked it is difficult for an attacker to attack on database and update the data. So using RRW user is able to watermarked large portion of database or large database also.

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

V. CONCLUSION

Optimization Algorithm use for generation of watermark bits is a great research area as many optimization algorithms are available. Available techniques uses genetic algorithms to get optimal solution, but it is time consuming, and the techniques can be effectively applied on numerical database, therefore the researchers needs to be focused on other algorithms which can give better result in less amount of time, as well as the algorithms which can be applicable for multimedia database, because today most of the data is multimedia data. Now a days most of the operations are carried out on multimedia database, therefore to improve security of multimedia database different optimization algorithms is a great research area.

References

[1] S. Iftikhar, M. Kamran and A. Zahid, "RRW - A Robust and Reversible Watermarking," *IEEE*, vol. 27, no. 04, 2015.

[2] G. Ghogare and A. Junnarkar, "Genetic Algorithm Based Reversible Watermarking Approach for Numeric and Non-Numeric Relational Data," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 7, 2017.

[3] J. khurram and A. khan, "Genetic algorithm and difference expansion based reversiblewatermarking for relational databases," *Elsevier*, p. 12, 2013.

[4] G. Gupta and J. Piperzyk, "Database relation watermarking resilient against secondary watermarking attacks," *Springer*, p. 222–236, 2009.

[5] K. Patil and S. Chopra, "Enhanced and Resilient Watermarking Technique," *International Journal of Engineering Research in Computer Science and Engineering*, vol. 5, no. 8, 2018.

[6] s. Iftikhar, M. Kamran and A. Zahid, "A survey on reversible watermarking techniques for Relational Database," *Wiley*, 2015.