Watermarking of Relational Databases using Video

Anju Paul, Sunitha E V

Abstract— **Information security is a growing concern because of increase use of network and growing economical value of the data in the real world. There are a lot of security mechanisms for protecting the data while transferring through the network. Watermarking appears as an interesting security mechanism that can provide ownership verification of the data. Watermarking of relational database has just begun its maturity cycle towards full deployment in industry level applications. This paper proposes a new method for watermarking the relational databases using multimedia data that is by using video. For providing better security to the data this system offers a new watermarking method in relational framework.**

Index Terms— **Copyright protection, Information security, Watermarking, relational framework.**

I. INTRODUCTION

Nowadays the information technology is growing very fastly. The importance of security is a growing concern in this technological world. The developing technology will cause security breaches in many systems. One of the important securities in an information system is Data security. We can explain data security in different ways. Data security refers to protective digital privacy measures that are applied to prevent unauthorized access to computers, websites and databases. Data security protects data from corruption. Data security is also known as Computer security or information security. In simple terms, data security is the practice of keeping data protected from corruption and unauthorized access. The focus behind data security is to ensure privacy while protecting personal or corporate data. Data privacy refers to keep the data as private, that is unauthorized disclosure of the data will lead to privacy issues. Data is a raw form of information arranged as rows and columns in the databases. This kind of databases is known as relational databases. There has been a huge prominence on data security because of great invention of internet. There are number of security features for protecting the data like encryption, strong authentication, backup solutions, etc. Encryption is one of the most common security features for protecting the data from unauthorized access. Simply in Cryptographic terms encryption is the process of hiding data by the use of ciphers from unauthorized users. Encryption provides confidentiality, integrity and authenticity of information transferred from sender to receiver. All these are very important while transferring the data through internet. But when an attacker decrypts the data these properties are violated. It will not provide posteriori protection to the data. Encryption does not ensure copyright protection of content. The new technology known as watermarking ensure the copyright protection of data. Watermarks are increasingly used for copyright control, robustness against signal processing transformations and resistance to tampering attacks. The watermarking technology is primarily motivated for providing copyright protection to digital content such as audio, images and video. Watermarking and encryption are two complimentary techniques. Encryption protects data during transmission from sender to receiver. After receipt and decryption data is no longer protected. Watermarking is the process of embedding a signal directly into the data. Even after decrypting the data the goal of a watermark is to always remain present in the data. Watermarking of relational database is recently added technology in the world of electronic data exchange. There are many applications for watermarking such as owner identification, to identify the content owner, broadcast monitoring to determine royalty payments authentication and fingerprinting to identify the buyer of the digital content, etc.

The relational database consists of a collection of data items organized as a set of formally described tables which can provide better security to the data by means of confidentiality, integrity and availability. Unauthorized entry to a database is known as loss of confidentiality, unauthorized modification to the data is known as loss of integrity and lack of access to data in a database is known as loss of availability. Databases become more susceptible to security threats while transferring it through internet. We are providing an ownership verification and copyright protection of data in a database by embedding some information in it, which is by watermarking relational databases. Watermarking does not prevent copying but it deters illegal copying by providing a means for establishing the original ownership of a redistributed copy. This paper proposed an improved technique for database watermarking using video.

The rest of this paper organized as follows. Section II presents some of the related works for watermarking relational databases. Section III explains about common database watermarking scheme. Section IV presents the proposed scheme for embedding the watermark in the relational database. In that we explain new database watermarking scheme using video. Section V briefly describes the results. Section VI concludes this paper.

Manuscript received April, 2015.

***Anju Paul**, PG Student, Department of Computer Science, Toc H Institute of Science and Technology, Ernakulam, India,*

***Sunitha E V**, Assistant Professor, Department of Information Technology, Toc H Institute of Science and Technology, Ernakulam, India,*

ISSN: 2278 – 7798

II. Related Works

There are a lot of works related to watermarking of relational databases by giving copyright protection. This section deals with some of the old related works of watermarking the relational databases based on the values inferred from the database itself.

R. Agrawal and J. Kiernan [1] proposed a technique for watermarking the relational database based on specific bit values that are determined under the control of a private key. This private key is only known to the owner of the data. These bit patterns create a watermark. They described the properties of watermarking like detectability, robustness, incremental updatability, imperceptibility, blind system, key-based system, etc. This paper focused on benign updates and malicious attacks. The different types of malicious attacks are bit attacks, randomization attack, rounding attack, subset attack, mix and match attack, additive attack and invertibility attack. The watermarking technique which was proposed by R. Agrawal and J. Kiernan is based on the Message Authentication Code (MAC). This technique watermarks only numeric attributes in a relational database. All the numeric attributes do not need to be watermarked, the selection of the attributes is done by the owner of the data. The relational database consists of number of tuples and each tuples have number of attributes. There is a primary key attribute P. The properties of one way hash function are utilized in this technique. The hash function returns a fixed length hash value for an arbitrary length of message M. That is h= H (M). Let F be a MAC that randomizes the values of the primary key attribute r.P of tuple r and returns an integer value in a wide range.

F (r.P) = H (K o H (K o r.P)) Where o represents concatenation. The watermark insertion is a kind of bit setting. According to the watermark insertion algorithm described in this paper, the hash value is computed, and if this first hash value is even then the j*th* least significant bit is set to zero otherwise it is set to one. Do not apply the watermark to the null value attributes. In watermark detection algorithm the match count is determined, and this match count is compared with the minimum threshold value. Here it is assumed that the attacker could not change the value of the primary key. Because primary key consists of useful information. Further this approach would be extended to the relations without primary keys. In that case partition the bits of attribute into two groups. Half of the bit values are used as primary key substitutes and the remaining bits are used for watermarking. The disadvantage is that the duplicates should be less otherwise too many duplicates in primary key substitute will result in identical watermarks which an attacker can exploit. Another drawback is that if one of the attributes is omitted by an attacker, the owner will not be able to detect the watermark. The proposed technique could withstand above described attacks. This technique does not provide a mechanism for multi bit watermark. The Least Significant Bit (LSB) based data hiding technique does not prevent trivial attacks. It can watermark only numeric attributes only.

S. Bhattacharya and A. Cortesi [2] proposed a distortion free invisible watermarking technique for relational databases. The main idea is to build the watermark after partitioning tuples with actual attribute values and then build hash functions on top of these groupings. The watermark embedding process is described as follows. The tuples are partitioned into groups and each group is considered independently. A keyed group hash value is computed on the number of grouping tuples using HMAC function. That is H*qk*. Then a watermark W = extractBits (H*qk*, ln(qk)) of ln(qk) is derived from H*qk*, where qk is the number of tuples. The watermark W is embedded into this group by permuting the order of the tuples. The new order can be easily calculated from W using Myrvold and Ruskey's linear permutation unranking algorithm based on W.

The watermark detection is described in this paper as follows. A group of qk tuples are selected and their ordering *k* is identified, where *k* is a permutation of (0,…,qk-1). A watermark W can be derived from *k* using Myrvold and Ruskey's linear permutation ranking algorithm. Based on the tuple hash of the sorted tuples, a group hash value is computed (H*qk*) using HMAC function with same key value used during embedding. Then a watermark W can be extracted from the group hash. If W matches W' the tuples in this group are authentic; otherwise the data in this group have been modified or tampered with. This scheme can withstand the attacks like modification of an attribute value, insertion of a tuple and also deletion of a tuple. So it can detect and locate modifications. It is distortion free and invisible as it consists into a permutation of the table tuples. Therefore no space overhead is required. This watermarking technique can be tuned according to different security levels. The longer a watermark, the more secure is the scheme.

III. Common Database Watermarking

The common database watermarking can be explained as follows. Consider a relational database R which consists of n tuples and each of tuples have n attributes. The tuple grouping operation is performed by finding group index number n_u [3] [4] is found out by the hash function of watermarking secret key concatenated with the hash of watermarking secret key and primary key attribute as in equation (1).

$$n_u = H(W_s \mid H(W_s \mid t_u.PK)) \bmod N_g \qquad (1)$$

W_s is the watermarking secret key chosen by the owner, t_u.PK represents primary key attribute H is the Hash function, Ng non intersecting group of tuples and | represents concatenation operation. There are two important stages in common database watermarking scheme. Message embedding and message detection or extraction phase as shown in Fig.1. Message embedding consists of pretreatment and modulation. Tuple grouping operation is performed in pretreatment phase. After pretreatment phase the attribute values are modified. The detection phase is just reverse of the embedding process. By analyzing most of the related works and common database watermarking we can say that most of the works concentrated on numeric attributes only and also the capacity of watermarking is less. The proposed work eliminates the above disadvantages by introducing a new watermarking scheme using host system as video.

ISSN: 2278 – 7798

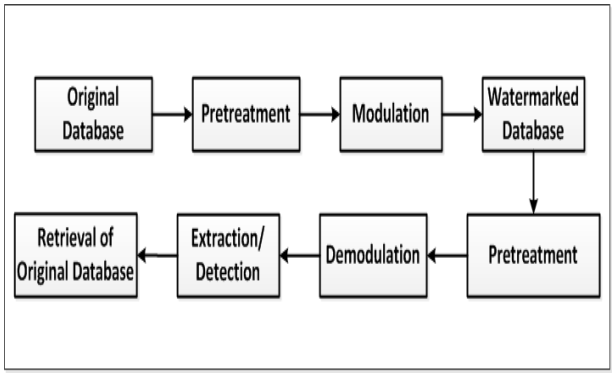

Fig.1.Common Database Watermarking

IV. PROPOSED SCHEME

The proposed work adapts some of the concepts proposed by De Vleeschouwer et al [5] and integrating it with the common database watermarking scheme. We select the host system for watermarking the relational database as multimedia data, that is video. The most common strategy used here is Circular histogram modulation. One of the main applications of the proposed system is on protecting the medical databases for the analysis of patient records and diagnosis of diseases. Here relational databases are watermarked using video to ensure the protection of data by verifying the ownership authentication or by giving copyright protection. Consider a relational database R consists of t_n tuples and each of the tuples consists of n attributes A1, A2… An. First we group these n tuples by finding group index number n_u as in equation (1) to make the watermark independent of the database. We are using the hash function to group the tuples. The selection of watermarking secret key W_s is purely depending on the sender's choice. The proposed scheme takes a video in Audio Video Interleave (AVI) format. The input video consists of number of frames. We are separating this video into number of frames. Then select one frame for watermarking the relational databases. Each of the single frames is treated as an image. Image is a collection of pixel values. Separate this image into two planes, suppose A plane and B plane. We are setting the B plane as target image for watermarking the database. Then histograms are created for each of the plane. The histograms are created using pixel values of the images. Histogram equalization is done for both of the planes. The values of the histograms will be mapped into circle then we get circular histogram. According to the basic strategy of circular histograms the modulation is performed by modulating the relative angle between two blocks of pixels. For the embedding process the frame should be encrypted. Embed the database into encrypted frame by XOR ing modulated values. The proposed scheme is shown in Fig.2 and detailed block diagram in Fig.3 gives a fine idea about the proposed scheme. The space availability for embedding database in a frame is calculated in terms of ASCII characters.

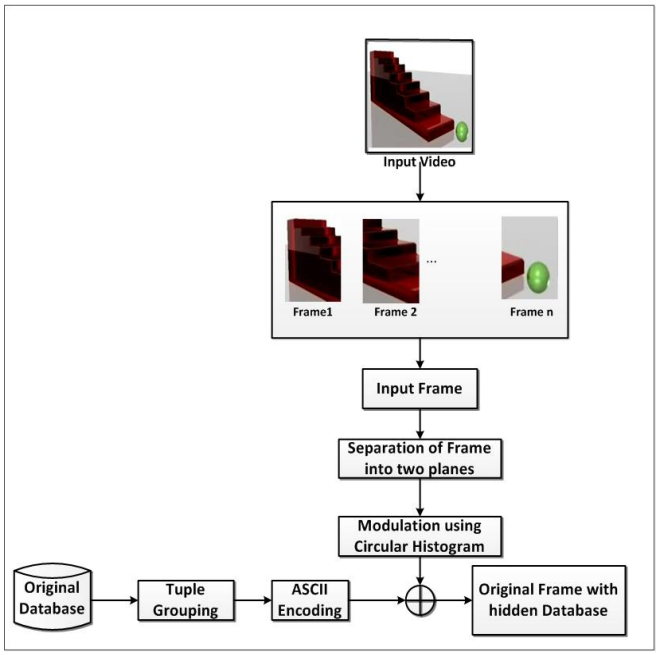

Fig.2.Proposed Scheme

We can explain above process in simple steps:

- Tuple grouping Operation
  Database tuples are grouped by finding group index number n_u as in equation (1). The grouping operation is performed for making the watermark independent of the database.
- Creation of two image planes
  The input video is in AVI format will be separated into number of frames. We take one frame for watermarking the relational database. Create histograms for two planes A and B. Histogram equalization is done for two planes.
- Modulation and Embedding
  After creating the histograms of two planes map the values of histograms into a circle. Then the relative angle between both circular histogram's center of mass is modulated. The target frame is encrypted for embedding the database. The database will be hidden inside the encrypted target image. The embedding is mainly done with the bit XOR operation. The output obtained as original frame with hidden database.
- Data extraction and Recreation of Original video
  Data extraction is performed using histogram shifting. The shifted planes are combined for getting the original video.

The proposed work gives a distortion less output. That is it can watermark the database without any distortion. The watermark is fully independent of database so that a variation in the frames does not lead to any loss of data which is embedded into the video. The capacity of watermarking is high compared to other watermarking schemes of relational databases. The authenticity and integrity of the data inside the database depend on watermarking secret key and properties of hash function.

ISSN: 2278 – 7798

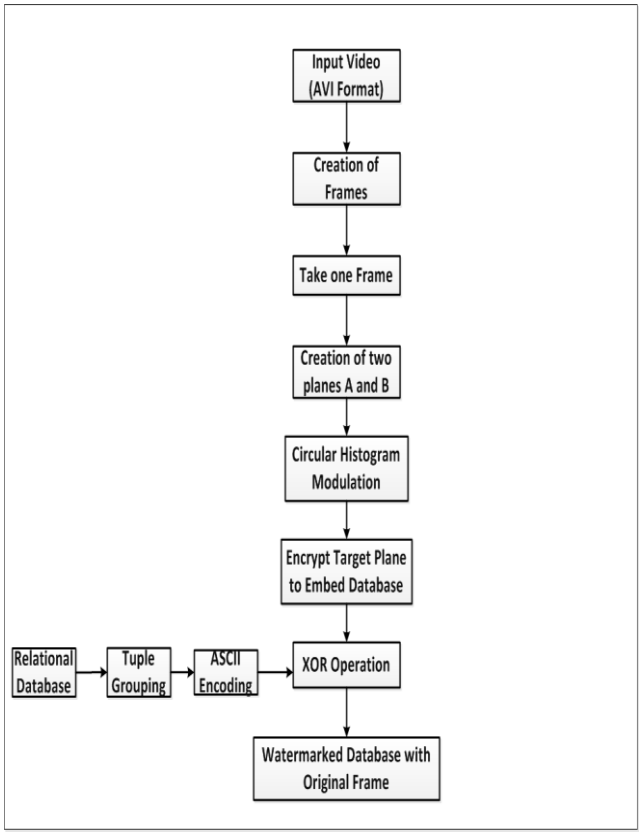

Fig.3.Process Flow Diagram of Proposed Scheme

V. RESULTS

This section gives a brief idea about the results while we implementing it. The program code is written in MATLAB. The various graphs and figures demonstrate the results of our proposed work. The input video is divided into number of frames. We are selecting one frame as input.

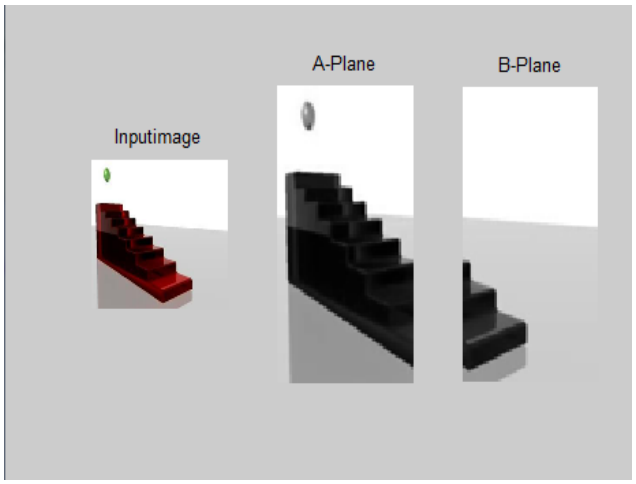

Fig.4.Creation of Image Planes

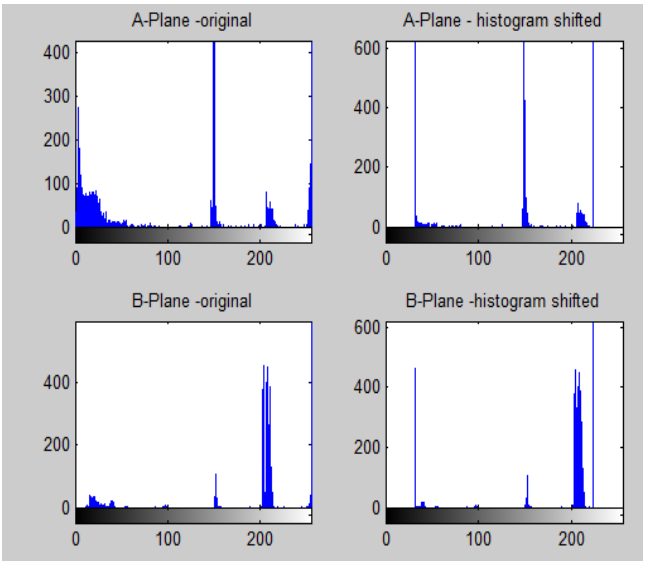

Fig.5.Histograms for two Image Planes

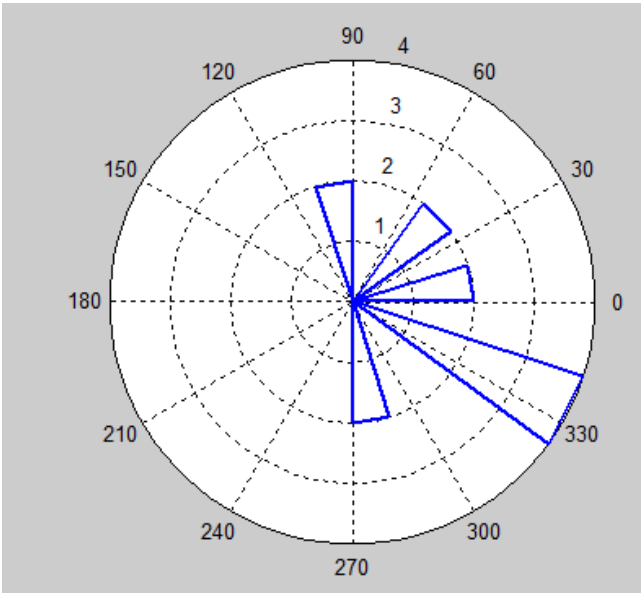

Fig.6.Circular Histogram

We can see the circular histogram shifted image planes.

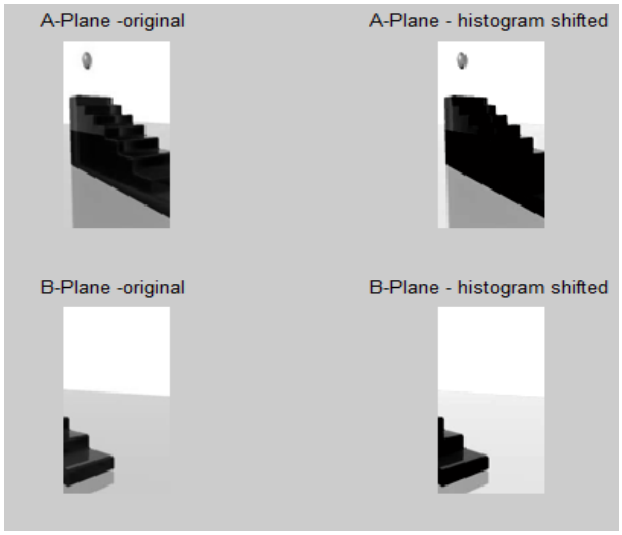

Fig.7.Circular Histogram Shifted Image Planes

ISSN: 2278 – 7798

Target frame or image is encrypted for embedding the database. Embed the database into the encrypted image.

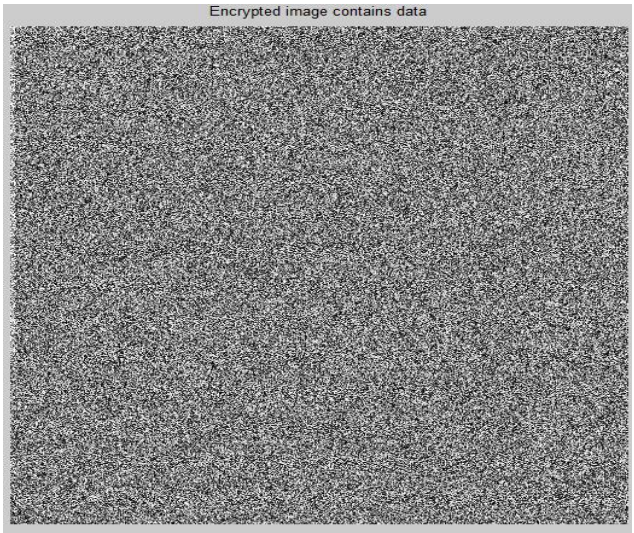

Fig.8.Encrypted Image Contains Data

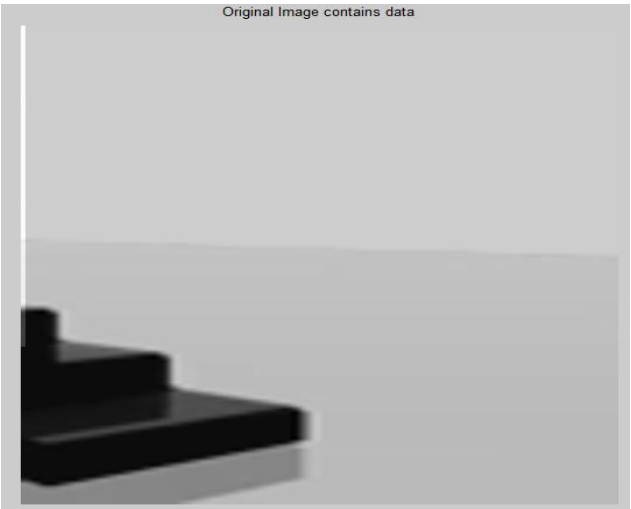

Fig.9.Output with Watermarked Database

ID	Name	Blood_Group	Height	Weight	Age
H120	Ann	A+	5.8	58	35
H121	Sam	B+	5.6	56	28
H122	Bob	O+	4.9	49	20
H123	Alice	AB+	5.1	51	38
H124	Mary	O+	3.2	32	15

Fig.10.Sample Medical Database

This is our simple medical database that was embedded into the above frame. After extraction we get this original database without any distortion. We calculated Mean Squared Error (MSE) value and Peak Signal to Noise Ratio (PSNR) value. We got MSE = 0.2413 and PSNR = 54.3051. Some of the other results are:

Number of ASCII characters Hidden: 277
Number of Bits Hidden: 2216
Number of Bits Space Available for Hiding: 49152
Utility: 4.50846 %

VI. CONCLUSION

This paper proposed a better approach for watermarking relational databases using video. The watermarking is done on both numerical and categorical attributes. There are a lot of existing methods for watermarking the relational databases based on grayscale images. The proposed work provides good results and also a new method. It can watermark potential large attributes. This system can withstand data manipulation attacks. The main applications of our system are on protection of medical database, military database, etc. We can extend our proposal for watermarking multiple databases in multiple frames at a time.

REFERENCES

[1] R. Agrawal and J. Kiernan, "Watermarking relational databases," in *Proc. 28th Int. Conf. VLDB*, Jul. 2002, pp. 155–166.

[2] S. Bhattacharya and A. Cortesi, "A distortion free watermark framework for relational databases," in *Proc. 4th Int. Conf. Softw. Data Technol.*, vol. 2. Dec. 2009, pp. 229–234.

[3] M. Shehab, E. Bertino, and A. Ghafoor, "Watermarking relational databases using optimization-based techniques," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 1, pp. 116–129, Jan. 2008.

[4] R. Sion, M. Atallah, and S. Prabhakar, "Rights protection for relational data," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 12, pp. 1509–1525, Dec. 2004.

[5] C. De Vleeschouwer, J.-F. Delaigle, and B. Macq, "Circular interpretation of bijective transformations in lossless watermarking for media asset management," *IEEE Trans. Multimedia*, vol. 5, no. 1, pp. 97–105, Mar. 2003.

[6] William Stallings, "*Message Authentication and Hash functions"* in Cryptography and Network Security, 4th edition, Pearson Education, Inc.2006, ISBN 978-81-7758-774-6.

[7] Zhi-Hao Zhang, Xiao-Ming Jin, Jian-Min Wan, De-Yi LP,"Watermarking relational database using image", In Proc of the Third International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004.

[8] Zhongyan Hu,Zaihui Cao, Jianhua Sun, "An image based algorithm for watermarking relational databases", In Proc of IEEE International Conference on Measuring Technology and Mechatronics Automation, 2009.

Anju Paul is a PG scholar in Toc H Institute of Science and Technology under CUSAT. She completed B Tech in Information Technology with distinction from ASIET, Kalady. She has presented a paper in IEEE International conference sponsored by Noorul Islam University Nagarcoil.

Sunitha E V is a Research scholar in Cochin University of Science and Technology (CUSAT). Now working as Asst. Professor in Information Technology in TIST. She is a B Tech degree holder in Information Technology from College of Engineering Poonjar and M Tech from CUSAT in Software Engineering. She has presented technical papers in many national and international seminars and conferences.

ISSN: 2278 – 7798